The Impact of No Child Left Behind on Students, Teachers and Schools*

Thomas S. Dee

University of Virginia and NBER

dee@virginia.edu

Brian Jacob

University of Michigan and NBER

bajacob@umich.edu

September 2010

* We would like to thank Rob Garlick, Elias Walsh, Nathaniel Schwartz and Erica Johnson for their research assistance. We would also like to thank Kerwin Charles, Robert Kaestner, Ioana Marinescu and seminar participants at the Harris School of Public Policy and at the NCLB: Emerging Findings Research Conference for helpful comments. An earlier version of this work was also presented by Jacob as the David N. Kershaw Lecture at the Annual meeting of the Association of Public Policy and Management (November 2008). All errors are our own.

1. Introduction

The No Child Left Behind (NCLB) Act is arguably the most far-reaching education-policy initiative in the United States over the last four decades. The hallmark features of this legislation compelled states to conduct annual student assessments linked to state standards, to identify schools that are failing to make "adequate yearly progress" (AYP) and to institute sanctions and rewards based on each school's AYP status. 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 can improve the focus and productivity of public schools.

NCLB has been extremely controversial from its inception. Critics charge that NCLB has led educators to shift resources away from important but non-tested subjects (e.g., social studies, art, music) and to focus instruction in math and reading on the relatively narrow set of topics that are most heavily represented on the high-stakes tests (Rothstein et al. 2008, Koretz 2008). In the extreme, some suggest that high-stakes testing may lead school personnel to intentionally manipulate student test scores (Jacob and Levitt 2003). While there have hundreds of studies of test-based accountability policies in the U.S. over the past two decades, the evidence on NCLB is more limited, both because it is a newer policy and because the national scope of the policy makes it extremely difficult to assess.

The goal of this paper is to examine the impact NCLB has had on students, teachers and schools across the country. We not only investigate how NCLB influenced student achievement, but also how it affected education spending, instructional practice and school organization. Given the complexity of the policy and the nature of its implementation, we are skeptical that

any single analysis can be definitive. For this reason, we present a broad collage of evidence, and look for consistent patterns in the results.

Several findings emerge. First, the weight of the evidence suggests that NCLB has had a positive effect on elementary student performance in mathematics, particularly at the lower grades. The benefits appear to be concentrated among traditionally disadvantaged populations, with particularly large effects among Hispanic students. We do not find evidence that the policy has adversely impacted achievement at either the top or bottom end of the test-score distribution. Instead, the policy-induced gains in math performance appear similar across the test-score distribution. However, the available evidence suggests that NCLB did not have comparable effect on reading performance.

A closer look at potential mechanisms provides some additional insight. For example, we find evidence that NCLB increased per-pupil district expenditures by roughly $700. These increased expenditures were allocated both to direct student instruction as well as to educational support services. We also find that these increased expenditures were funded largely by state and local, not Federal, sources. The test-score gains associated with these expenditure increases fall short of the ambitious goals enshrined in NCLB. However, we present some qualified evidence suggesting that they were cost-effective.

We also discuss evidence on how NCLB may have influenced alternative measures of educational practice and student outcomes. This evidence suggests that NCLB led to an increase in the share of teachers with masters degrees. We also find evidence that teachers responded to NCLB by reallocating instructional from social studies and science towards key tested subjects, particularly reading. We also present evidence that NCLB led to distinct increases in teacher-reported measures of student engagement (e.g., attendance, timeliness, and intellectual interest).

The remainder of the paper proceeds as follows. Section 2 outlines the theoretical underpinnings of school accountability and provides background on the NCLB legislation. Section 3 examines the impact of NCLB on student achieving, providing evidence from a variety of different sources. Section 4 investigates potential mediating mechanisms, discussing how the policy impacted educational expenditures, classroom instruction and school organization among other things. Section 5 concludes with recommendations for future policy and research.

2. Background on School Accountability and NCLB

2.1 Theoretical Underpinnings of School Accountability

A basic perception that has motivated the widespread adoption of school-accountability policies like NCLB is that the system of public elementary and secondary schooling in the United States is "fragmented and incoherent" (e.g., Ladd 2007). In particular, proponents of school-accountability reforms argue that too many schools, particularly those serving the most at-risk students, have been insufficiently focused on their core performance objectives and that this organizational slack reflected the weak incentives and lack of accountability that existed among teachers and school administrators. For example, Hanushek and Raymond (2001) write that accountability policies are "premised on an assumption that a focus on student outcomes will lead to behavioral changes by students, teachers, and schools to align with the performance goals of the system" and that "explicit incentives… will lead to innovation, efficiency, and fixes to any observed performance problems."

The theoretical framework implicitly suggested by this characterization of public schools is a principal-agent model. The interests of parents and voters are viewed as imperfectly aligned with those of teachers and school administrators. Furthermore, parents and voters cannot easily

monitor or evaluate the input decisions made by these agents. The performance-based sanctions and rewards that characterize accountability policies are effectively output-based incentives that can be understood as a potential policy response to this agency problem. Similarly, some of the provisions in NCLB with regard to teacher qualifications can be construed as an "agent selection" approach to a principal-agent problem.

The principal-agent lens is also useful for understanding criticisms of accountability-based reforms. The assumption that teachers and school administrators have misaligned self-interest implies that they may respond to accountability policies in unintentionally narrow or even counterproductive ways. For example, in the presence of a high-stakes performance threshold, schools may reallocate instructional effort away from high and low-performing students and towards the "bubble kids" who are most likely, with additional attention, to meet the proficiency standard (e.g., Neal and Schanzenbach 2010). Similarly, concerns about "teaching to the test" reflect the view that schools will refocus their instructional effort on the potentially narrow cognitive skills targeted by their high-stakes state assessment at the expense of broader and more genuine improvements in cognitive achievement. Schools may also reallocate instructional effort away from academic subjects that are not tested or even attempt to shape the test-taking population in advantageous ways.

2.2 Research on Pre-NCLB Accountability Reforms Adopted by States

School-accountability reforms similar to those brought about by NCLB were adopted in a number of states during the 1990s. Several research studies have evaluated the achievement consequences of these reforms. Because of the similarities between the NCLB and aspects of these pre-NCLB accountability systems, this body of research provides a useful backdrop against which to consider the potential achievement impacts of NCLB. In a recent review of this diverse

evaluation literature, Figlio and Ladd (2008) suggest that three studies (Carnoy and Loeb 2002, Jacob 2005, and Hanushek and Raymond 2005) are the "most methodologically sound" (Ladd 2007).

The study by Carnoy and Loeb (2002), which was based on state-level achievement data from the National Assessment of Educational Progress (NAEP), found that the within-state growth in math performance between 1996 and 2000 was larger in states with higher values on an accountability index, particularly for Black and Hispanic students in 8th grade.[1] Similarly, Jacob (2005) found that, following the introduction of an accountability policy, math and reading achievement increased in Chicago Public Schools, relative both to the prior trends and relative to the contemporaneous changes in other large urban districts in the region. However, Jacob (2005) also found that, for younger students, there were not similar gains on a state-administered, low-stakes exam and that teachers responded strategically to accountability pressures (e.g., increasing special-education placements).

Hanushek and Raymond (2005) evaluated the impact of 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 publicized school-level performance and could attach consequences to that performance. The types of

[1] 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.

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., identifying failing schools, 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 introduction of consequential accountability within a state was associated with statistically significant increases in the gain-score measures. The achievement gains implied by consequential accountability were particularly large 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. The authors argue that these achievement results provide support for the controversial school-accountability provisions in NCLB because those provisions were so similar to the consequential-accountability policies that had been adopted in some states.

2.3 Features of the No Child Left Behind (NCLB) Legislation

The NCLB legislation was actually a reauthorization of the historic Elementary and Secondary Education Act (ESEA), the central Federal legislation relevant to K-12 schooling. The ESEA, which was first enacted in 1965 along with other "Great Society" initiatives and previously reauthorized in 1994, introduced Title I, the Federal government's signature program for targeting financial assistance to schools and districts serving high concentrations of economically disadvantaged students. NCLB dramatically expanded the scope and scale of this Federal legislation by requiring that states introduce school-accountability systems that applied

to *all* public schools and students in the state. In particular, 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.

NCLB requires that states introduce "sanctions and rewards" relevant to every school and based on their AYP status. However, NCLB also mandates explicit and increasingly severe sanctions for persistently low-performing schools that receive Title I aid (e.g., public school choice, staff replacement, and school restructuring). According to data from the Schools and Staffing Survey, 54.4 percent of public schools participated in Title I services during the 2003-04 school year. However, it should be noted that some states applied these explicit sanctions to schools not receiving Title I assistance as well. For example, 24 states introduced accountability systems that threatened all low-performing schools with reconstitution, regardless of whether they received Title I assistance (Olson 2004).

3. The Impact of NCLB on Student Achievement

The overarching goal of NCLB has been to drive broad and substantive improvements in student achievement. This section discusses the available empirical evidence on the achievement effects of NCLB, drawing on a variety of research designs and data sources including national time trends, private-public comparisons as well as comparisons across schools and states within the U.S.

3.1 National Time Trends in Student Achievement

Because NCLB was introduced simultaneously throughout the United States, many observers have turned to state and national time-series trends in student achievement to assess the impact of these reforms. For example, several studies have noted that student achievement, particularly as measured by state assessment systems, appears to have improved both "overall and for key subgroups" since the implementation of NCLB (Stullich, Eisner, McCrary and Roney 2006, Center on Education Policy 2008). Others, however, argue that changes in student performance on high-stakes state tests can be highly misleading when states strategically adjust their assessment systems and teachers narrow their instructional focus to state-tested content (Fuller, Wright, Gesicki, and Kang 2007)

In Figure 1, we present data on the national trends in student achievement from 1990 to 2009. These data are from the Main NAEP[2] and provide separate trends by grade (i.e., 4th and 8th), by subject (i.e., math and reading) and by race and ethnicity (i.e., white, black, and Hispanic). The dashed horizontal line in 2002 visually identifies the year in which NCLB began. These trends suggest that NCLB may have increased the math performance of 4th graders. That is, these NAEP data suggest that grade-4 math achievement shifted noticeably higher during the NCLB era and may have also begun trending upwards more aggressively. The trend data suggest similar gains in the math performance of black 8th graders. However, apart from that, these trends provide no clear suggestion that the onset of NCLB improved performance in the other three grade-subject combinations. Figure 2 shows achievement growth for 9- and 13-year olds in math and reading using data from the Long-Term Trend (LTT) NAEP, which has tracked students from

[2] There are several different versions of the NAEP. The original NAEP, first administered in the early 1970s, is now called the Long-Term Trend (LTT) NAEP because the Department of Education has made an effort to keep this examine as consistent as possible over time in order to accurately gauge national trends. The LTT NAEP is based on a small random sample of 9, 13 and 17 year olds across the country, and generally focuses on what many educators now think of as "basic" skills. The "Main NAEP" was initiated in the early 1990s in an effort to both update the content and format of the national assessment in order to test a broader domain of knowledge and skills, and also to allow individual states to obtain their own state-representative estimates. This exam is administered to 4th and 8th graders.

the early 1970s. These data similarly suggest that NCLB had at most targeted effects on student achievement.

3.2 Evidence from International Comparisons

While these national achievement trends are suggestive, they do not necessarily provide the basis for reliable inferences about the impact of NCLB. Simple time-series comparisons may be biased by the achievement consequences of other time-varying determinants such as the economic recession that just preceded the introduction of NCLB. One straightforward way to benchmark the achievement trends observed in the U.S. is to compare them to the contemporaneous achievement trends in other counties.

In particular, because the time-series evidence from Figure 1 suggests that any positive achievement effects from NCLB were likely to have been concentrated in grade 4 math achievement, the comparative international achievement data from the Trends in International Mathematics and Science Study (TIMSS) provide a particularly relevant source of data. TIMSS collected trend data on grade 4 math achievement for participating countries in 1995, 2003, and 2007. Panel (a) in Figure 3 presents the grade-4 scale scores on math from the TIMSS for the United States, for the 12 other countries that collected these performance data in each of the these three study years (i.e., Australia, England, Hungary, Iran, Japan, Latvia, Netherlands, New Zealand, Norway, Scotland, Singapore, and Slovenia), and for the subset of these comparison countries that are OECD (Organization for Economic Co-operation and Development) members.

These trend data indicate that average math achievement on the TIMSS fell for all sets of countries by roughly equal amounts between the only available pre-NCLB year (1995) and the first year in which the U.S. implemented NCLB (2002-03). Without additional years of data, we cannot assess the extent to which these comparative changes deviate from pre-NCLB trends.

However, the available TIMSS data indicate that, by 2007, math achievement had comparatively improved in the United States, particularly with respect to the other OECD countries (i.e., 11 scale points versus 4). These cross-country trends provide suggestive evidence consistent with the hypothesis that NCLB led to targeted improvements in the math performance of younger students in the United States. However, the comparative test-score gain for the United States (i.e., 7 scale points) is relatively modest. For example, with respect to U.S. test scores prior to NCLB, this gain implies a 1.3 percent increase in average performance and an 8 percent increase relative to the standard deviation in test scores.

However, like the national time-series evidence, international comparisons provide no indication that NCLB improved the reading achievement of young students. Specifically, the Progress in International Reading Literacy Study (PIRLS) provides data on the reading achievement of 4th graders across several countries both in 2001 and in 2006. Panel (b) in Figure 3 presents the overall reading scores by year for the United States, the group of 26 other countries that participated in both surveys and the OECD members of this comparison group. On average, the United States outperformed these comparison countries. However, over the period NCLB was implemented in the U.S., these groups experienced quite similar and modest changes in PIRLS reading achievement. Overall, the international evidence is at best suggestive. The lack of multiple years of data make it difficult to distinguish possible policy effects from other trends or to identify any comparative differences with statistical precision. A more subtle shortcoming of national and international time-series comparisons is that the presumption of a common, national effect elides the heterogeneous effects of NCLB across particular types of states and schools.

3.3 Evidence from Accountability-Risk Studies

However, several recent econometric studies have creatively leveraged this heterogeneity to identify the effects of NCLB. In particular, a widely used approach involves structuring comparisons across schools or students that face a different risk of sanctions under NCLB. Neal and Schanzenbach (2010) 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. Using data from the state of Washington, Krieg (2008) finds that the performance of students in the tails of the distribution is lower when their school faces the possibility of NCLB sanctions.

Ballou and Springer (2008), using data from a low-stakes exam fielded in seven states over a four-year period, identify the achievement consequences of NCLB by constructing comparisons across grade-year cells that were included in AYP calculations and those that were not. Their approach takes advantage of the fact that between 2002-03 and 2005-06, states differed with respect to whether particular grades mattered for a school's accountability rating. Hence, their identification leverages the fact that if the math scores of 4th graders counted toward the a school's accountability rating in one year but the math scores of 5th graders in the school did not count until the following year, one would expect student achievement to grow more quickly among 4th graders relative to 5th graders in the current year. They find that the presence of AYP accountability modestly increased the mathematics achievement of elementary-school students, particularly lower-performing students.

A recent study by Reback, Rockoff, and Schwartz (2010) adopts a similar approach, comparing student performance across elementary schools on the margin of making AYP using nationally representative data from the Early Childhood Longitudinal Study (ECLS). They find

that low-stakes reading and science scores improve by as much as 0.07 standard deviations when a school is on the margin for making AYP, though the effects of mathematics scores are smaller and statistically insignificant.

These "accountability risk" studies provide credible evidence on how NCLB-induced pressure influences the level and distribution of student achievement. However, they have at least three potential limitations with respect to understanding the broad achievement consequences of NCLB. First, most of these studies have limited external validity because they do not rely on national data. Second, some of these studies rely on high-stakes assessments, which may have attenuated construct validity as a measure of student achievement in the presence of strategic responses to NCLB (e.g., teaching to the test). Third, and perhaps most importantly, the treatment contrast in these studies may not approximate the full impact of NCLB because they rely on comparisons across schools or students, all of whom were observed in the post-NCLB policy regime. To the extent that NCLB had broad effects on public schools (i.e., even on students and schools not under the direct threat of sanctions), these comparisons could understate the effects of interest.

3.4 Evidence from a Comparison of U.S. States over Time

In order to address some of the limitations described above and estimate what one might consider the "full" impact of NCLB, we utilize a strategy that compares changes in student performance within U.S. states over time (see Dee and Jacob, forthcoming). We leverage the fact that NCLB was explicitly modeled on an earlier generation of state-level school accountability systems. In the decade prior to NCLB, thirty states implemented "consequential" school-accountability policies that were fundamentally similar to NCLB in that they mandated systematic testing of students in reading and math, public reporting of school performance on

these exams and the possibility of meaningful sanctions (e.g., negative ratings, takeover, closure, reconstitution, replacing the principal and/or allowing student mobility) based on test-based school performance. In fact, some state officials argued that NCLB "needlessly duplicates" pre-existing state accountability systems (Dobbs 2005).

The existence of these prior NCLB-like accountability systems establishes natural treatment and comparison states. In our framework, states that adopted "NCLB-like" accountability to prior to NCLB form our comparison group. Other states, for which NCLB catalyzed an entirely new experience with consequential school accountability, form our treatment group.[3] Of course, states that adopted early accountability programs were not randomly distributed. For this reason, our estimation strategy (described in more detail below) relies on within-state variation over time, allowing not only for different levels of achievement across states prior to NCLB but also different *trends* in achievement across states prior to NCLB.

3.4.1 Graphical Evidence

In this subsection, we illustrate the logic of our identification strategy through a series of figures. This graphical evidence has the advantage of transparency and simplicity. In the following subsection, we present regression estimates that more clearly show the magnitude and statistical precision of our findings, and allow us to demonstrate that the results are robust to a variety of alternative specifications and several falsification exercises.

Figure 4 shows the trends in NAEP scores for two groups: (i) states that adopted school accountability between 1994 and 1998 (comparison states) and (ii) states that did not adopt

[3] We relied on a number of different sources to categorize pre-NCLB accountability policies across states, including prior studies of such policies (e.g., Carnoy and Loeb 2002, Lee and Wong 2004, and Hanushek and Raymond 2005) as well as primary sources such as 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 (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.

school accountability prior to NCLB (treatment states).[4] The NAEP data are particularly well suited to this evaluation for several reasons. The NAEP is a technically well-developed assessment that covers a broad domain of knowledge and schools. It provides consistent, state-representative measures of student performance for most states over the last two decades. Finally, the exam is "low-stakes" for students, teachers and schools.[5] Because teachers have no incentive to "teach to the NAEP," it is likely to provide the most accurate measure of student achievement (Fuller, Wright, Gesicki, and Kang 2007).

The figures plot the simple (unweighted) average scale score of each group of states in all years in which the exam was administered. Year refers to the spring of the relevant academic year (that is, year 1992 refers to the 1991-92 school year). Note that the sample of states is consistent across years (i.e., a balanced panel) and that the state classification is a time-invariant characteristic of the state. The horizontal line at 2002 denotes the year in which NCLB was signed into law, so data points to the left are considered "pre-policy" and data points to the right are "post-policy."[6] To illustrate the pre- and post-NCLB achievement trends within each group, we also plot the fitted regression line from a simple linear regression that is conducted separately for each group x period (i.e., pre- or post-NCLB).

---

[4] These figures exclude a small number of states that adopted state accountability programs between 1999 and 2001. We do so to provide the clearest distinction between our treatment and comparison groups. However, the regression analysis described in the following section includes these "late adopter" states. Dee and Jacob (forthcoming) show that the inclusion of these late adopters does not change the findings in any substantive way.

[5] That is, the NAEP is not used as the basis for student promotion/retention, teacher evaluation or school accountability. Indeed, the NAEP is only administered to a small, random sample of 4th and 8th grade students in each state.

[6] When one dates the start of NCLB is a potentially important issue. NCLB secured final Congressional approval (December 18, 2001) and was signed by President Bush (January 8, 2002) in the middle of the 2001-02 academic year. NCLB is often characterized as having been implemented during 2002-03 because states were required to use testing outcomes from the prior 2001-02 year as the starting point for determining whether a school was making adequate yearly progress (Palmer and Coleman 2003, Olson 2002). However, one could reasonably conjecture that the discussion and anticipation surrounding the adoption of NCLB would have influenced school performance during the 2001-02 school year. Alternatively, it could also be argued that NCLB should not be viewed as in effect until the 2003-04 academic year when new state accountability systems were more fully implemented as well as more informed by guidance from and negotiations with the U.S. Department of Education (Olson 2002, 2003). For a more detailed discussion of this issue, see Dee and Jacob (forthcoming).

Figure 4a shows trends in 4th grade math achievement. We see that in 1992, states that never adopted accountability scored roughly 5 scale points (0.18 standard deviations) higher on average than states that adopted school accountability policies by 1998. While all states made modest gains between 1992 and 2000, the states that adopted accountability policies prior to 1998 experienced more rapid improvement during this period.[7]

If the NCLB accountability provisions had a causal impact on student performance, one would expect achievement to increase more after 2002 in states with no prior accountability relative to states with prior accountability. It is possible that NCLB led to a level shift in student achievement, which would be manifest as a shift in the intercept post-NCLB. It is also possible that NCLB changed the *rate* of achievement growth, which would be manifest as a change in the *slope* of the achievement trend post-NCLB.[8] Whether one considers a shift in the intercept or the slope, our identification strategy relies on a comparison of treatment vs. control states that accounts not only for the pre-NCLB levels of achievement in those states but also the pre-NCLB achievement trends in those states.

In Figure 4a, the mean level of math achievement jumped noticeably in 2003 for both groups of states. However, relative to prior trends, this shift was largest among the "no prior accountability" group (i.e., the NCLB "treatment" states). Interestingly, there was little noticeable change in the growth *rate* across periods for the prior-accountability states (i.e., the "control" states). That is, the slope of the achievement trend before and after 2002 is roughly equivalent for this group. In contrast, states with no prior accountability grew at a faster rate

---

[7] This visual evidence is consistent with the prior evaluation literature that has studied pre-NCLB state accountability reforms (e.g., Carnoy and Loeb 2002, Jacob 2005, and Hanushek and Raymond 2005).

[8] The rate of achievement growth might increase post-NCLB for several reasons. First, it may take states time implement new curriculum, instructional strategies or other support services for students. Second, later cohorts of students will have been "exposed" to NCLB for a larger fraction of their school careers than earlier cohorts. Without imposing additional assumptions, it is not possible to cleanly distinguish between these effects. For this reason, we focus on the "net" impact of NCLB in different years after the legislation was passed.

from 2003 to 2007 than from 1992 through 2000, such that the growth rates after 2002 were roughly equivalent across both groups of states. These trends suggest that NCLB had a positive impact on 4th grade math achievement.

The trends for 8th grade math (Figure 4c) are similar, though somewhat smaller, than those for 4th grade math. The pattern for 4th grade reading in Figure 4b is much less clear. The pre-NCLB reading trends for both groups are much noisier than the math trends. In particular, both groups experienced a decline in achievement in 1994, little change in 1998 (relative to 1992) and then very large gains in 2002.[9] The prior accountability group experienced a drop in achievement from 2002 to 2003, both in absolute terms and relative to trend. The other group experienced very little increase following NCLB. Perhaps most importantly, however, a visual inspection of the data in these plots indicates that the prior achievement trend was not linear, which is a central assumption of the linear CITS model. Similarly, Figure 4d provides no evidence of an NCLB effect on 8th grade reading achievement.

3.4.2 Estimation Strategy

Perhaps the most straightforward approach to estimating the impact of NCLB in the framework described above is a simple difference-in-difference framework in which one compares the achievement levels of treatment vs. comparison states before and after the introduction of NCLB. However, a fundamental assumption of this model is that any pre-existing trends in the outcome variables are equivalent across treatment and control groups. The figures above clearly show that the control states (i.e., states that implemented consequential accountability prior to NCLB) realized more rapid improvements during the pre-NCLB period.

[9]" Note that the graph is scaled to accentuate what are really quite small absolute changes from year to year.

For this reason, we estimate a more flexible specification that allows for pre-existing trends to differ across groups. Specifically, we estimate the following model:

$$Y_{st} = \beta_0 + \beta_1 YEAR_t + \beta_2 NCLB_t + \beta_3 (YR\_SINCE\_NCLB_t) + \beta_4 (T_s \text{ x } YEAR_t)$$

$$\beta_5 (T_s \text{ x } NCLB_t) + \beta_6 (T_s \text{ x } YR\_SINCE\_NCLB_t) + \beta_7 X_{st} + \mu_s + \varepsilon_{st} \quad (1)$$

where Y_{st} is a 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 dummy variable equal to one for observations starting in the academic year 2002-03. $YR\_SINCE\_NCLB_t$ 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. X_{st} represents state x year covariates. In the main specification, the only state-year covariate included is the fraction (and fraction squared) of students who were tested but excluded from official reporting because of limited English proficiency or some type of learning disability. The variables, μ_s and ε_{st} represent state fixed effects and a mean-zero random error respectively.

T_s is a time-invariant variable that measures the treatment imposed by NCLB. In the most basic setup, T_s could be thought of as a dummy variable indicating whether a given state had *not* instituted consequential accountability prior to NCLB. This is the approach implicitly taken in Figure 4. However, it is more accurate to view the "treatment" provided by the introduction of NCLB in the framework of a dosage model. Slightly more than half of the states that introduced consequential school accountability prior to NCLB did so just within four years prior to NCLB's implementation. The simple binary definition of T_s defined above could lead to attenuated estimates of the NCLB effect because the "control" group would include some states for which the effects of prior state policies and NCLB are closely intertwined.

For this reason, we define T_s as the number of years during our panel period that a state did *not* have school accountability. Specifically, we define the treatment as the number of years *without* prior school accountability between the 1991-92 academic year and the onset of NCLB. Hence, states with no prior accountability have a value of 11. Illinois, which adopted its policy in the 1992-93 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. Our identification strategy implies that the larger the value of this treatment variable, the greater potential impact of NCLB.

This regression specification 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).[10] For the sake of parsimony, the results below report 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).[11]

The primary threat to causal inference in this approach is the existence of time-varying unobservable factors that are (a) coincident with the introduction of NCLB, (b) differentially effect treatment versus comparison states, and (c) independently affect student performance. One example is endogenous student mobility, as might occur if NCLB caused families to leave or return to the public schools. Another problematic scenario would be if treatment or comparison states recovered from the 2001 recession more quickly. As discussed below, we take

---

[10] We get similar results when we allow for a separate NCLB "effect" for each post-NCLB year. However, given the limited number of data points available, a shift in a linear achievement trend seems to capture the relevant variation.

[11] Specifically, this effect is calculated as $\beta_5 + \beta_6(5)$ in the simple case where T_s is binary but as $\beta_5(6)+ \beta_6(6 \times 5)$ when T_s is allowed to vary across states and the NCLB effect is identified relative to a state that implemented school accountability in 1997. As a practical matter, both approaches generate similar results (Dee and Jacob 2009, Table 3).

particular care to examine a variety of such potential concerns, and find no evidence that our findings are biased.

Finally, it is worth considering exactly how one should interpret the resulting estimates. Our estimates capture the impact of the accountability provisions of NCLB, but will not reflect the impact of other NCLB provisions such as Reading First or the highly qualified teacher provision. In addition, our estimates will identify the impact of NCLB-induced school accountability provisions on states without prior accountability policies. To the extent that one believes that states that expected to gain the most from accountability policies adopted them prior to NCLB, one might view the results we present as an underestimate of the average treatment effect of school accountability.

3.4.3. Results

Table 1 presents estimates of the impact of NCLB on student performance derived from equation (1) with no time-varying state-level control variables. The results suggest NCLB had a positive effect on elementary student math performance, but no impact on reading performance. The mean impact of 7.2 points for 4th grade math translates to an effect size of 0.23 standard deviations. The effects are even larger at the left tail of the ability distribution. The estimates suggest that NCLB increased the proportion of 4th graders reaching the basic level on NAEP by 10 percentage points, or a 16 percent increase relative to the control mean of 64 percent. While the mean effects for 8th graders are not statistically significant at conventional levels (a 0.10 standard deviation effect with a p-value of 0.12), the effects at the bottom tail are stronger. NCLB increased the fraction of 8th graders reaching the basic level in math by 5.8 percentage points (9 percent).

While we find that NCLB had larger impacts among lower-achieving students, we do not find any evidence that the introduction of NCLB harmed students at higher points on the achievement distribution. In contrast to some prior work within individual districts and states, we find that NCLB seemed to increase achievement at higher points on the achievement distribution more than one might have expected. For example, in 4th grade math, the impacts at the 75th percentile were only 3 scale points lower than at the 10th percentile.

In order to test the sensitivity of our results to some of the potential time-varying unobservable factors described above, we conduct a series of "falsification exercises" in which we re-estimate equation (1) with a variety of alternative outcome measures, including state-year poverty rates, median household income, employment-population ratios and fraction of students in the public schools. Across the 40 regressions we run (i.e., 10 models for each of the 4 grade x subject combinations), we find only 1 estimate significant at the 5 percent level and 3 estimates significant at the 10 percent level. These largely null findings suggest that the assumptions required for identification are indeed met. In Dee and Jacob (forthcoming), we also show that the results presented in Table 1 are robust to a host of alternative specifications, including the inclusions of a variety of state-year covariates, the inclusion of state-specific time trends, the inclusion of a full set of year fixed effects, and weighting the data based on the number of students enrolled in that state-year.[12]

Table 2 shows regression estimates separately by subgroup, both unweighted and weighted by student enrollment. Interestingly, the positive effects are particularly large among among lower-income and minority students. For example, in 4th grade, NCLB increased math achievement among Black and Hispanic students by 14.6 points (0.47 standard deviations) and

[12] Dee and Jacob (forthcoming) show that these results are also robust to measuring the intensity of the "treatment" imposed by NCLB in terms of the stringency of the proficiency standards imposed by the state. Cook et al. (2009) find this as well.

9.8 points (0.32 standard deviations), respectively. The impact on students eligible for subsidized lunch was 8 points (0.26 standard deviations). Interestingly, the enrollment-weighted estimates are systematically larger than the unweighted estimates for low-income and minority subgroups, particularly for Black students. Taken at face value, this suggests an important source of treatment-effect heterogeneity. Specifically, it implies that NCLB had a more positive effect on disadvantaged students in states with a greater number of such children (e.g., NCLB was more effective for Black students in Alabama than for Black students in South Dakota). However, given the relatively small number of "treatment" states with large populations of Black students, the possibility that this heterogeneity reflects other state-specific traits cannot be discounted.

3.5 Evidence from Public and Private-School Comparisons

The comparison of trends in student performance within states over time presented above suggests that NCLB had a substantial impact on math achievement, particularly among disadvantaged students in 4th grade. As with any non-experimental design, however, the findings rest on assumptions that cannot be fully tested. For this reason, we present results from a complementary analysis that makes use of an alternative comparison group.

In this approach, we assess the impact of NCLB by comparing trends over time in student performance in public versus Catholic schools.[13] While students in private schools are eligible to participate in a number of major programs under the Elementary and Secondary Education Act (ESEA), the NCLB reauthorization of ESEA left these prior provisions "largely intact" (U.S. Department of Education 2007), implying that the NCLB reforms were comparatively irrelevant for private schools. The use of Catholic schools as a comparison group improves upon international comparisons by providing a within-nation control group. However, as with the

[13] In earlier work, we identify several potential concerns with using Catholic schools to identify the impact of NCLB. See Dee and Jacob (forthcoming) for details.

national and international time-series evidence, this approach also conflates the effects of NCLB across states and schools where its impact was heterogeneous.

Figure 5 shows achievement trends across public and Catholic schools, following the same structure as the earlier figure comparing treatment and control states. While the performance of both public and Catholic students trended up during the sample period, Catholic-school students consistently outperformed their public-school counterparts. However, following the implementation of NCLB, the mathematics performance of public-school students converged somewhat towards the achievement levels in Catholic schools and began a somewhat stronger trend growth. This comparative convergence is particularly pronounced for 4th grade students and is consistent with the other time-series evidence suggesting that NCLB improved math achievement, particularly among younger students. The reading achievement trends of 8th graders are quite similar across public and Catholic schools, suggesting the absence of a meaningful NCLB impact. However, the reading achievement of public-school 4th graders trended upwards during the NCLB era, particularly relative to the reading achievement of Catholic-school 4th graders, which began a distinctive downward trend during the NCLB era.

These public-Catholic comparisons are broadly consistent with the state-based comparisons, suggesting that NCLB led to substantial gains in the mathematics achievement of 4th graders and, possibly, 8th graders. These particular cross-sector comparisons also suggest that NCLB increased the reading achievement of 4th graders. A recent study by Cook et al. (2009) includes regression estimates based on public-Catholic comparisons of this sort and draws similar conclusions. They also find similar, though less precisely estimated, results in comparisons based on non-Catholic private schools.

3.6 Summary of Achievement Effects

Given the national scope of the policy, assessing the causal impact of NCLB on student performance is not straightforward. However, the body of evidence presented above seems to suggest that the federal school accountability policy did improve elementary student math achievement, particularly among socioeconomically disadvantaged groups. There is not comparable evidence that NCLB generated meaningful improvements in reading achievement. Moreover, it is important to recognize that the analysis presented above focuses exclusively on elementary schools. NCLB also requires AYP determinations for high schools but relatively little is known about NCLB's effects on high schools due in part to data limitations (e.g., the main NAEP has no state-level data for secondary-school math achievement after 2000).

How policy-relevant are the overall gains in math achievement that appear to be due to NCLB? One way to benchmark a 7.2 point (0.23 SD) gain in grade 4 math achievement is to compare this effect to achievement gaps that are of interest. For example, a test-score gain of this size is equivalent to approximately 24 percent of the black-white test score gap observed in the 2000 NAEP data. Furthermore, because NCLB appears to have been more effective among disadvantaged subgroups, it may have contributed to closing some achievement gaps. For example, the effect of NCLB on the grade 4 math achievement of Hispanic students was roughly 5 points larger than the corresponding effects on white students, implying that NCLB closed the white-Hispanic achievement gap by 19 percent.

4. Impact of NCLB on the Organization and Practice of Education

Given the encouraging effects on math achievement and the somewhat puzzling lack of effects for reading, it is natural to ask how NCLB impacted the organization and practice of

elementary education across the country. Such evidence on potential mediating mechanisms could not only guide revisions to the NCLB legislation, but also shed light on the education production function in ways that inform other school reforms. To provide some coherence to the subsequent discussion, we group non-achievement outcomes from a variety of sources into several broad categories: (1) changes in educational resources; (2) changes in instructional focus and/or methods; and (3) changes in school organization, climate or culture.

4.1 Impact of NCLB on Education Expenditures

Standards-based reforms have often been presented to the public as a trade – greater resources and flexibility for educators in exchange for greater accountability. One of the most strident criticisms of NCLB is that it failed to deliver on this bargain. However, there is surprisingly little research on the relationship between school accountability and spending, despite an extensive literature on education finance more generally.

One notable exception is an analysis of district-level expenditure data from 1991-92 to 1996-97 by Hannaway, McKay and Nakib (2002). Examining four states that implemented comprehensive accountability programs in the 1990s – Kentucky, Maryland, North Carolina and Texas – they find that only two (Texas and Kentucky) increased educational expenditures substantially more than the national average. Hannaway and Stanislawski (2005) present evidence that the major pre-NCLB accountability reforms in Florida were associated with increased expenditures for instructional staff support and professional development, particularly in low-performing schools. Of course, it is difficult to determine whether the accountability policy caused the increased expenditures, or were merely part of a broader reform agenda. Overall, the extant literature provides at best suggestive evidence on how accountability reforms may have influenced school spending.

To provide new evidence on how NCLB influenced local school finances, we pooled annual, district-level data on revenues and expenditures from U.S. Census surveys of school district finances (i.e., the F-33 Survey of Local Government Finances) over the period from 1994 to 2008 (Dee, Jacob and Schwartz 2010). Our analytical sample consists of all operational, unified school districts (roughly 10,000) for each survey year. To identify the effects of NCLB accountability on district finances, we utilize the same cross-state trend analysis described above, comparing changes in school finance measures in states with and without pre-NCLB accountability programs.

Figure 6 shows trends in district expenditures over time, separately for states that adopted consequential accountability prior to NCLB versus those that did not. All results are reported in $2009 and are weighted by district enrollment. Like the earlier figures, the trend lines are fitted linear regression lines.[14] In Figure 7a, we see that total per-pupil expenditures rose more quickly from 1994-2002 in states that adopted pre-NCLB accountability policies. But following the introduction of NCLB, the spending grew more slowly in these early-adopting states, suggesting that NCLB increased expenditures. Figures 7b and 7c show comparable results for the two largest categories of total expenditures, instructional and support-service spending.

Table 3 presents regression estimates based on the model shown in equation (1), with the inclusion of the following district-year controls: enrollment, enrollment squared, fraction black or Hispanic, poverty rate (based on 2000 census data), poverty rate squared and the interaction between poverty rate and fraction black or Hispanic. As in earlier models, we present standard errors clustered by state. Consistent with the presentation of the achievement effects, we report

[14] The figures omit states that adopted school accountability programs between 1999-2001 because the impacts of these state programs might be confounded with the introduction of NCLB in 2002. In the regression estimates discussed below, however, we incorporate all states.

the impact of NCLB in 2007 for states that did not have consequential accountability prior to NCLB relative to states that adopted consequential accountability in 1997.

The results indicate that NCLB increased total current expenditures by $733 per pupil, reflecting a 9-percent increase from 1999-2000 mean of $8,357. The increased expenditures were allocated to direct instruction and support services in proportions roughly equivalent to average spending patterns, with effects of $504 (10 percent) and $256 (9 percent) respectively. Results presented in the bottom two rows reveal that the increased expenditures were funded entirely by state and local revenue, an empirical result aligned with the allegations of critics, who have alleged that NCLB constitutes an "unfunded mandate." In results not shown here, we find that the effects were fairly similar across districts characterized by baseline student-poverty levels, suggesting that NCLB did not meaningfully influence distributional equity. Moreover, in results reported elsewhere, we demonstrate that these findings are robust to the same falsification exercises and alternative specifications described earlier for the achievement analysis (Dee, Jacob and Schwartz 2010).[15]

In light of the achievement effects discussed in the previous section, a natural and policy-relevant question is to ask how the monetized benefits of those test-score gains compare to the corresponding expenditure increases presented here. Based on prior estimates that a one standard deviation increase in elementary math scores is associated with an 8 percent in adult earnings (Kruger 2003), the 0.23 SD impact of NCLB would translate into an earnings boost of 1.8 percent. Assuming a 3 percent discount rate, the present discounted value as of age 9 of a 1.8 percent increase in subsequent earnings beginning at age 18 is at least $13,300.[16] Hence, even if

[15] As discussed in related work, we do not find substantial impacts on class size, suggesting that the increase in instructional expenditures due to NCLB may have been allocated to other functions (Dee, Jacob and Schwartz 2010).

[16] This calculation uses an age-earnings profile of 18-65 year olds from the March 2007 Current Population Survey (CPS). Allowing for reasonable productivity-related growth in earnings (i.e., 2 percent) increases the monetized benefit of the test-score gains due to NCLB to roughly $25,500.

we assume that the increased expenditures due to NCLB are sustained for all eight elementary-school years, the economic benefits of the corresponding test-score gains are at least twice as large. It should be stressed that this exercise turns on multiple unstated assumptions. In particular, this back-of-the-envelope calculation ignores socially relevant benefits (e.g., the externalities of human-capital improvements) and costs (e.g., the deadweight losses associated with raising revenues). More generally, it is not clear that these expenditure increases were a relevant mediating mechanism behind NCLB's achievement effects. Nonetheless, this calculation provides suggestive evidence that the achievement and expenditure effects of NCLB could pass a cost-benefit test.

4.2 Impact of NCLB on Teachers and Classroom

One of the most prominent issues with respect to NCLB concerns the intended and unintended ways in which it may have influenced classroom practice. In particular, test-based accountability policy creates a strong incentive for educators to focus on tested content and skills. Indeed, according to many, this is the exact point. At the same time, critics have worried that such incentives may cause schools to neglect important but non-tested subjects, or to change instructional practice in a way that prioritizes narrow test preparation over broader learning. In this section, we discuss the available evidence on how school accountability programs, including NCLB, influence classroom instruction.

The most consistent and compelling finding with regard to school accountability and classroom instruction involves the allocation of instructional time. A number of studies have documented that test-based accountability programs cause educators to (1) reallocate instructional time toward tested subjects; (2) reallocate time within tested subjects toward

specific content and skills covered on the exam; and (3) increase time devoted to narrow test preparation activities that may have little broader value (Hannaway and Hamilton 2008).

In 2001, for example, researchers at the National Board on Education Testing and Public Policy surveyed a nationally representative sample of teachers, asking them a series of questions about how state-mandated testing programs influenced their practice (Pedulla et al. 2003). Teachers in states where the exam results were used to hold teachers or schools accountable reported shifting instruction toward tested subjects more than teachers in states where the exam results were used primarily for informational purposes. For example, 34% of teachers working in high-stakes testing regimes reported that the testing program had increased the time spent in tested areas "a great deal" relative to 17% of teachers in moderate-stakes regimes.

In addition, teachers in states with school accountability programs reported spending more time on a variety of activities designed to improve student test-taking skills such as taking practice tests (Pedulla et al. 2003). In states where the tests had important consequences for the schools, roughly 36 percent of elementary teachers reported spending more than 30 hours per year on test preparation activities, compared with only 12 percent of teachers in states where tests had few consequences for schools.[17]

Recent studies that focus on NCLB itself find similar results. In 2005, for example, researchers at RAND collected data from teachers, principals, and superintendents in three states (California, Pennsylvania, and Georgia) to examine how they were responding to the introduction of NCLB (Hamilton et al. 2007). Educators reported a narrowing of the curriculum and an emphasis on test preparation, particularly for "bubble kids" near the proficiency cut score

[17] Ladd and Zelli (2002) found similar results in a survey study of school principals in North Carolina during the period when the state was introducing its school-accountability program. Specifically, principals reported devoting more resources to the high-stakes subjects of math, reading and writing.

for their state assessment system. In addition, educators responded to NCLB by increasing the alignment between the curriculum and state standards.

Studies of earlier school accountability programs found a similar increase in alignment. Specifically, the programs led teachers to shift the content of their instruction within subjects (Stecher et al. 1998, Koretz et al. 1996; Jacob 2005; Koretz and Hamilton 2006). This literature emphasizes that the format and structure of the test itself can influence instruction. For example, Taylor et al. (2003) finds that testing programs with short, open-ended items leads teachers to focus greater attention on problem-solving skills.

The Center for Education Policy (CEP) has studied the implementation and impact of NCLB since its inception (CEP 2006, 2007, 2008, 2009). As part of its work, CEP not only surveyed a nationally representative sample of school districts in 2005-06 and again in 2006-07, but also conducted more intensive case studies of selected school districts. District officials report that NCLB increased the instructional time they devote to math and English language arts. About 62% of districts reported that between 2001-02 and 2006-07 they increased instruction in these subjects in elementary schools, with the largest increases coming from districts with more schools in need of improvement (CEP 2007) and in urban and high-poverty districts (CEP 2006). Moreover, the reallocation reported by officials appears substantial. For example, 80% of districts that reported increasing ELA time did so by at least 75 minutes per week, and 54% reported doing so by at least 150 minutes per week (CEP 2008). Most districts that reported increased time for ELA or math reported cuts in the time for other subjects (such as social

studies, art, music, gym, recess or lunch) rather than increases in total time in school (CEP 2008).[18]

The CEP studies also suggest that NCLB influenced classroom practice in ways that may have attenuated teacher autonomy. For example, CEP (2006, 2007) reports that schools made a concentrated effort to align their curriculum to state standards in the wake of NCLB, thus changing the focus of their curriculum to put greater emphasis on tested content and skills. Many districts also became more prescriptive during this period about what and how teachers were supposed to teach (CEP 2006).

It is worth noting that the costs and benefits of these instructional changes depend on one's objectives, and are not always clear even for a given objective. For example, many observers applaud the increasing emphasis on math and reading instruction, while others lament the decreasing attention on subjects such as art and music (Rothstein et. al. 2008).

While these studies paint a consistent picture, they need to be interpreted with some caution. All of the research described thus far relies on self-reports from teachers or administrators. Moreover, the information is based on questions that ask respondents to retrospectively assess whether certain practices have changed over time. For this reason, one might be worried about the reliability and validity of the data (Bradburn and Sudman 1988).

Few studies have implemented regression-based research designs that attempt to isolate the effects of school-accountability policies on district, school, and classroom practices from the potentially confounding effects of other determinants. One prominent exception is a recent study by Rouse et al. (2007), which used a regression-discontinuity design and data from principal

[18] A 2009 GAO study based on teacher survey data (and supplemental interviews with state officials) finds that 90 percent of elementary teachers reported no change in instructional time for arts education between 2004-05 and 2006-07 (GAO 2009). At the same time, a higher fraction of teachers in schools identified as needing improvement under NCLB reported a decline in art instruction, relative to teachers in other schools. This study used data from the Department of Education's National Longitudinal Study of No Child Left Behind (NLS-NCLB).

surveys in Florida to examine how schools responded to pressure from the state's accountability system. They find that accountability pressure leads to an increased emphasis on low-performing students (e.g., grade retention, summer school, and tutoring), increased overall instructional time, and reorganized school days. They also find suggestive evidence that accountability reduced principal control and increased the resources available to teachers. Furthermore, the school policies influenced by school accountability explain a meaningful fraction of student test-score gains, suggesting that schools responded to accountability pressure in specific ways that improved student achievement.

While this work addresses some of the concerns from previous work, it has its own set of limitations. It does not address NCLB per se, and it estimates what one might describe as the partial impact of the Florida accountability system – i.e., comparing schools more or less affected by accountability pressure. However, it is possible that the accountability system in Florida, or NCLB more generally, may have led to changes across all schools.

In recent work, we use data from the nationally representative Schools and Staffing Survey (SASS) to inform this issue (Dee, Jacob and Schwartz 2010). The SASS is a nationally representative survey of teachers and school administrators that has been conducted periodically since the early 1990s (i.e., 1994, 2000, 2004, and 2008). We use teacher survey responses to construct a variety of measures of classroom instruction and school organization. These data allow us to compare changes in teacher responses over time rather than relying on retrospective judgments on the part of teachers. It also provides more objective measures of some of the constructs – e.g., the time use questions ask about the actual number of hours per week a teacher devotes to math, rather than asking teachers to characterize their emphasis on math as "big" or "small" or "larger/smaller" relative to certain number of years ago.

Figure 7a shows the fraction of elementary and middle school teachers with a MA degree by year. As with the previous the figures, we show the trends separately for states that did and did not adopt school accountability programs prior to NCLB. In states with prior accountability programs, roughly 43 percent of teachers had a MA degree prior in 1994 compared with 37 percent of teachers in other states. Following the introduction of NCLB, the fraction of teachers with a MA degree jumped notably in states without prior accountability so that in 2004 the rates were equal across both groups of states. And by 2008, teachers in states without prior accountability were slightly more likely to have a MA degree.

Table 4 presents regression estimates based on the model shown in equation (1), with the inclusion of state fixed effects and the following teacher and school controls to increase statistical precision: dummies for the teacher's race, school level, gender, assignment and grade level, quartic functions of school enrollment, school percentage minority, and school percentage free lunch, as well as an interaction between percent minority and percent free lunch. As above, standard errors are clustered by state. The estimate in row 1 of Table 4 indicates that NCLB increased the fraction with MA degree by roughly .056 from of a baseline of .41, representing an increase of roughly 14 percent. Given that many districts require teachers to have a MA degree for permanent certification, it is possible that this effect reflects the response of states to the NCLB provision requiring schools to have “highly qualified” teachers in every classroom. The fact that states with prior accountability policies also had a substantially higher fraction of teachers with a MA degree suggests that even state-adopted programs contained some provisions regarding teacher qualifications.

Figures 9b-c show trends in time use for our sample of elementary school teachers and principals, separately for states that did and did not adopt school accountability programs prior to

NCLB. Figure 9b shows the amount of instructional time (in hours per week) teachers report for core academic subjects. Figure 9c shows the fraction of time during the week that self-contained teachers spent teaching math and English where the denominator is the total time spent on the four core subjects (math, English, social studies and science). Figures 9d shows this ratio specifically with respect to math alone.

These figures suggest that NCLB did *not* lead to meaningful increases in the total amount of instructional time devoted to core subjects. However, these figures suggest that instructional time allocated to math and ELA increased following the introduction of NCLB. Moreover, the effects seem to be larger in states that had not instituted school accountability prior to this time, consistent with NCLB leading to this change.

Table 4 presents regression estimates based on the model shown in equation (1), with the inclusion of state fixed effects and the following teacher and school controls to increase statistical precision: dummies for the teacher's race, school level, gender, assignment and grade level, quartic functions of school enrollment, school percentage minority, and school percentage free lunch, as well as an interaction between percent minority and percent free lunch. As above, standard errors are clustered by state.

The estimates shown in row 4 of Table 4 indicate that NCLB increased the fraction of time teachers spend on math and English by 0.036 percentage points, off of a baseline of 74 percent. This implies an additional 45 minutes per week of math/ELA instruction for teachers who spend 20 instructional hours on these two subjects. It appears that this increase was driven primarily by an increase in time devoted to English. NCLB does not appear to have increased the fraction of time devoted to math at all. This is particularly interesting given that we find substantial achievement effects in math but not in reading.

4.2 School organization, climate or culture.

The prior literature provides some evidence that test-based accountability policies including NCLB have spurred other useful changes in school-wide instructional practice. In the RAND study, for example, school and district administrators reported that NCLB increased the use of formative assessment as an instructional tool and increased the technical assistance and professional-development opportunities offered to schools. In earlier survey work, researchers found that teachers in high-stakes environments found test results more useful, and were more likely to use test information to inform their practice, than colleagues in low-stakes environments (Pedulla et al. 2003). Similarly, teachers in the RAND study reported that their state's accountability system under NCLB led them to search for more effective teaching practices and, in nearly all cases, had led to positive changes in their schools (Hamilton et al. 2007). Interestingly, for example, teachers reported that teaching practices and the general focus on student learning "changed for the better" under accountability (Hamilton et al. 2007). District officials in the CEP study reported an increase in the use of data to guide instruction (CEP 2006).

Unfortunately, the SASS has not routinely collected data on many of the school and teacher practices that are of interest, limiting our capacity to isolate the effects of NCLB on some of these outcomes. However, the SASS has collected consistent data on several relevant school-level traits. For example, the principals who responded to the SASS were asked to choose from a list of 9 educational goals their top 3 priorities. Figure 8a shows the comparative trend data for the share of principals who indicated that either academic excellence or basic skills was their top goal. This measure of instructional focus did not converge after NCLB across states defined by whether they already had school accountability. This pattern suggests that NCLB did not

generate a detectable increase in instructional focus, a result confirmed by the regression results in Table 4.

Teachers in the SASS provided scaled responses to questions about whether principals and fellow teachers enforced rules for student conduct. Figure 8b shows the comparative trends for the standardized responses to this question and suggests the lack of an NCLB effect (see Table 4, also). Figure 8c shows the trend data for a standardized measure of how teachers viewed the student culture within their school. More specifically, this variable reflects the extent to which teachers felt that tardiness, absenteeism, apathy, etc. were a problem within their schools The trend data in Figure 8c suggest a decline in this measure of student culture for states that introduced school accountability because of NCLB. The regression results in Table 4 indicate that the effect size of this increase is 0.23.

This estimated effect is over twice as large in high-poverty schools (Dee, Jacob, and Schwartz 2010). While it is possible that student culture did deteriorate in these schools, it is also possible that teachers implicitly adjust their responses to this question based on the expectations facing the students and the school. For example, to the extent that NCLB increased the expectations for academic achievement in states without prior school accountability policies, it is possible that teachers in these states could have evaluated student culture more critically following the introduction of NCLB.

4.3 Summary

The evidence presented above suggests that NCLB has had both desirable and undesirable effects on teachers, classroom practice, and school culture. Unfortunately, the lack of objective measures of many important instructional practices limit our ability to examine many of the most plausible mechanisms through which accountability may have operated to improve

student achievement. Moreover, the analysis presented above does not allow us to identify which, if any, of the factors we identify as improving (e.g., per-pupil spending, student engagement, teacher qualifications, instruction time devoted to English) might explain the achievement effects we document.

5. Conclusions

Eight years have passed since No Child Left Behind dramatically expanded the federal role in public schooling. Given the national scope of the policy, it is difficult to reach definitive conclusions with regard to its impact. Nonetheless, evidence from a variety of data sources utilizing several plausible comparison groups suggests that NCLB had a positive effect on elementary student performance in mathematics, particularly at the lower grades. The benefits appear to be concentrated among traditionally disadvantaged populations, with particularly large effects among Hispanic students. On the other hand, the existing evidence suggests that NCLB did not have a comparable effect on reading performance.

We find compelling evidence that NCLB increased per-pupil district expenditures on both direct instruction and instructional support, a mediating mechanism which may explain the corresponding achievement gains. By 2008, for example, the policy appears to have increased annual per pupil spending by roughly $733 in states that did not have any school accountability program prior to NCLB. We also presented evidence that these expenditure increases were modest relative to the present discounted value of the corresponding test score gains. We also discussed evidence suggesting that NCLB influenced teachers and schools in several potentially important ways. It appears that NCLB has led elementary schools to increase instructional time devoted to math and reading, although the majority of evidence on this point comes from teacher

and/or administrator survey data that is subject to some potential biases. Similarly, teachers report that NCLB has encouraged schools to spend time on narrow test preparation activities. However, we also found evidence that NCLB led to increases in teacher-reported measures of student engagement. Unfortunately, a lack of richly detailed data that lends itself to credible identification strategies makes it difficult to assess whether NCLB influenced curriculum and instructional in more fundamental ways.

Nonetheless, the extant body of evidence can provide some guidance to the ongoing debate over the proposed reauthorization of NCLB. In March of 2010, the Obama administration released an NCLB "blueprint" that outlined proposed features of a reauthorization (Klein and McNeil 2010). This proposal calls for continued annual reporting of school-level test-based student assessments. However, it allows for flexibility in how states calculate school effectiveness (e.g., through the use of value-added modeling). The blueprint also calls for the use of non-test accountability indicators, especially measures of college and career readiness (e.g., attendance, course completion, and school climate). Another potentially critical feature of this proposal involves changing how measures of school performance are linked to consequences. The blueprint also proposes to give states increased flexibility in how they might intervene in low-performing schools, only mandating specific consequences for the very lowest-performing schools and those schools with persistently large achievement gaps. It is not clear how states would respond to this flexibility. However, the literature on pre-NCLB accountability policies suggests that simply reporting accountability measures that were unconnected to explicit consequences did not drive improvements in student achievement (Hanushek and Raymond 2004). This suggests that the targeted achievement gains attributable to NCLB could be at risk under state reforms that decoupled performance measures from meaningful consequences.

REFERENCES

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

Associated Press. "Diocese of Tucson Becomes 2nd to File for Bankruptcy," September 21, 2004.

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

Bandeira de Mello, V., Blankenship, C., and McLaughlin, D.H. (2009). *Mapping State Proficiency Standards Onto NAEP Scales: 2005-2007* (NCES 2010-456). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

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

Bifulco, Robert. "The Influence of Performance-Based Accountability on the Distribution of Teacher Salary Increases," Education Finance and Policy Spring 2010, Vol. 5, No. 2: 177–199.

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

Bloom, Howard S., 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.

Bradburn, Norman and Seymour Sudman. *Polls and Surveys: Understanding What They Tell Us.* Jossey-Bass, 1988.

Braun, Henry, & Qian, Jiahe. (2008). *Mapping state standards to the NAEP scale*. Report No: ETS RR-08-57. Princeton, NJ: Educational Testing Service.

Bryk, Anthony, Valerie E. Lee and Peter B. Holland. *Catholic Schools and the Common Good*. Harvard University Press, 1993.

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.

Carroll, Matt, Sacha Pfeiffer, Michael Rezendes and Walter V. Robinson. "Church Allowed Abuse by Priest for Years, Aware of Geoghan Record, Archdiocese Still Shuttled Him from Parish to Parish," *The Boston Globe*, January 6, 2002, A1.

Center on Education Policy. From the Capital to the Classroom: Year 4 of the No Child Left Behind Act. Washington, DC, March 2006.

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

Center for Education Policy (2007). "Choices, Changes and Challenges: Curriculum and Instruction in the NCLB Era." A report in the series From the Capital to the Classroom: Year 5 of the No Child Left Behind Act.

CEP (2008). "Instructional Time in Elementary Schools: A Closer Look at Changes for Specific Subjects." A report in the series From the Capital to the Classroom: Year 5 of the No Child Left Behind Act.

Cullen, Julie Berry and Randall Reback (2006). "Tinkering Toward Accolades: School Gaming Under a Performance Accountability System." in Improving School Accountability: Check-Ups or Choice, Advances in Applied Microeconomics, T. Gronberg and D. Jansen, eds., 14 (Amsterdam: Elsevier Science).

Dee, Thomas S. and Jacob, Brian A. (forthcoming). "The Impact of No Child Left Behind on Student Achievement). *Journal of Policy Analysis and Management*.n (Earlier version available as National Bureau of Economic Research Working Paper Number 15531).

Dee, Thomas S., Jacob, Brian A. and Nathaniel Schwartz (2010). "The Effect of No Child Left Behind on School Finance, Organization and Practice." Working paper.

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

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.

Figlio, David. (2006) "Testing, Crime and Punishment." *Journal of Public Economics* 90(4): 837-851.

Fredricks, Jennifer A., Phyllis C. Blumenfeld, and Alison H. Paris. 2004. School Engagement: Potential of the Concept, State of the Evidence. *Review of Educational Research* 74(1): 59-109.

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.

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.

Laura S. Hamilton, Brian M. Stecher, Julie A. Marsh, Jennifer Sloan McCombs, Abby Robyn, Jennifer Lin Russell, Scott Naftel, Heather Barney. *Standards-Based Accountability Under No Child Left Behind Experiences of Teachers and Administrators in Three States*. Rand Corporation, 2007.

Hannaway, Jane, and Yasser Nakib. 2002. *School Business Affairs* 68(5):12–19.

Hannaway, Jane, Shannon McKay, and Yasser Nakib. 2002. "Reform and Resource Allocation: National Trends and State Policies." In *Developments in School Finance 1999-2000*, edited by William Fowler. Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Hannaway, Jane, and Maggie Stanislawski. 2005. "Responding to Reform: Florida School Expenditures in the 1990s" (mimeo).

Hanushek, Eric A. and Margaret E. Raymond. "The Confusing World of Educational Accountability," National Tax Journal 54(2), June 2001, pp. 365-384.

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.

Heckman, James J. "The Economics, Technology and Neuroscience of Human Capability Formation," Proceedings of the National Academy of Sciences, 104(33): 13250-13255, (2007).

Hess, Frederick M. and Michael J. Petrilli. *No Child Left Behind Primer*. New York: Peter Lang Publishing, 2006.

Hess, Frederick M. and Michael J. Petrilli. "Wrong Turn on School Reform," Policy Review, February/March 2009.

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.

Jacob, B. and Levitt, S. (2003). "Rotten Apples: An Investigation of the Prevalence and Predictors of Teacher Cheating." *Quarterly Journal of Economics*. 118(3): 843-877.

Klein, Alyson. "Focus Turns to Congress after High Court's Denial of Challenge to NCLB Law," Education Week, June 16, 2010.

Klein, Alyson and Michele McNeil. "Administration Unveils ESEA Reauthorization Blueprint," Education Week, March 17, 2010.

Koretz, Daniel (2008). Measuring Up: What Educational Testing Really Tells Us. Harvard University Press, Cambridge, MA.

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.

Krueger, Alan B. 2003. Economic Considerations and Class Size. *Economic Journal* 113: F34-F63.

Ladd, Helen F. "Holding Schools Accountable Revisited," 2007 Spencer Foundation Lecture in Education Policy and Management, Association for Public Policy Analysis and Management, https://www.appam.org/awards/pdf/2007Spencer-Ladd.pdf, Accessed November 8, 2009.

Ladd, Helen F. and Arnaldo Zelli. "School-Based Accountability in North Carolina: The Responses of School Principals," Educational Administration Quarterly 38, 2002, 494-529.

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.

National Center for Education Statistics. *Mapping 2005 State Proficiency Standards Onto the NAEP Scales* (NCES 2007–482). U.S. Department of Education, National Center for Education Statistics, Washington, D.C.: U.S. Government Printing Office.

Neal, Derek and Diane Whitmore Schanzenbach (2010). "Left Behind by Design: Proficiency Counts and Test-Based Accountability." *Review of Economics and Statistics*, 92(2): 263-283.

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

Olson, Lynn. "States Strive Toward ESEA Compliance," *Education Week*, December 1, 2002.

Olson, Lynn. "In ESEA Wake, School Data Flowing Forth," *Education Week*, December 10, 2003.

Olson, Lynn. "Taking Root," *Education Week*, December 8, 2004.

Pedulla, Joseph J., Lisa M. Abrams, George F. Madaus, Michael K. Russell, Miguel A. Ramos, and Jing Miao. 2003. *Perceived Effects of State-Mandated Testing Programs on Teaching and Learning: Findings from a National Survey of Teachers*. Chestnut Hill, MA: National Board on Education Testing and Public Policy.

Palmer, Scott R. and Arthur L. Coleman. "The No Child Left Behind Act: Summary of NCLB Requirements and Deadlines for State Action," Council of Chief State School Officers, November 2003 (http://www.ccsso.org/content/pdfs/Deadlines.pdf, Accessed November 13,

2009).

Rothstein, Richard, Rebecca Jacobsen and Tamara Wilder (2008). "Grading Education: Getting Accountability Right." Teachers College Press, New York City.

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

Rouse, Cecilia Elena, Jane Hannaway, Dan Goldhaber, and David Figlio. "Feeling the Florida Heat? How Low-Performing Schools Respond to Voucher and Accountability Pressure," CALDER Working Paper 13, November 2007.

Spillane, James P., John B. Diamond, Patricia Burch, Tim Hallett, Loyiso Jita and Jennifer Zoltners. "Managing in the Middle: School Leaders and the Enactment of Accountability Policy," Educational Policy 16, 2002: 731-762.

Swanson, Christopher B. and David Lee Stevenson. "Standards-Based Reform in Practice: Evidence on State Policy and Classroom Instruction from the NAEP State Assessments," Educational Evaluation and Policy Analysis 24, 2002: 1-27.

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.

Taylor, Grace, Lorrie Shepard, Freya Kinner, and Justin Rosenthal. 2003. *A Survey of Teachers' Perspectives on High-Stakes Testing in Colorado: What Gets Taught, What Gets Lost.* CSE Technical Report 588. Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards and Student Testing.

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.

U.S. Department of Education, Office of Planning, Evaluation and Policy Development, Policy and Program Studies Service, Private School Participants in Federal Programs Under the *No Child Left Behind Act* and the *Individuals with Disabilities Education Act,* Washington, D.C., 2007.

U.S. Department of Education. "Title I Implementation – Update on Recent Evaluation Findings." U.S. Department of Education, Office of Planning, Evaluation and Policy Development. Stullich, Andrew Abrams, Elizabeth Eisner, Erica Lee, 2009.

Whitehurst, Grover. “Don’t Forget Curriculum,” Brown Center Letters on Education #3, October 2009.

Winters, Marcus A., Julie R. Trivitt, and Jay P. Greene (Forthcoming). “The Impact of High-Stakes Testing on Student Proficiency in Low-Stakes Subjects: Evidence from Florida’s Elementary Science Exam.” *Economics of Education Review*.

Table 1 - The Estimated Effects of NCLB on Student Achievement on the NAEP

	Dependent Variable			
Grade-Subject Sample	Mean NAEP Score (1)	Percentage At or Above Basic (2)	75th percentile NAEP Score (3)	90th percentile NAEP Score (4)
4th Grade Math (39 states, n=227)	7.244**	10.090**	6.634**	5.205**
	(2.240)	(3.145)	(1.902)	(1.916)
Mean outcome before NCLB in states without prior accountability	224	64	244	259
8th Grade Math (38 states, n=220)	3.704	5.888**	4.340**	2.537
	(2.464)	(2.680)	(2.189)	(2.404)
Mean outcome before NCLB in states without prior accountability	272	64	296	314
4th Grade Reading (37 states, n=249)	2.297	2.359	2.258**	2.097**
	(1.441)	(1.592)	(0.938)	(0.805)
Mean outcome before NCLB in states without prior accountability	216	61	240	258
8th Grade Reading (34, states, n=170)	-2.101	-3.763	1.289	1.172
	(2.070)	(2.561)	(2.249)	(2.897)
Mean outcome before NCLB in states without prior accountability	261	73	282	299

Notes: Each cell is a separate regression. See text for model details. ***p<0.01, ** p<0.05, * p<0.1.

Table 2 - The Estimated Effects of NCLB on NAEP Math Scores by Subgroup

	4th Grade				8th Grade			
	Mean NAEP Score		Percentage At or Above Basic		Mean NAEP Score		Percentage At or Above Basic	
Subgroup	Ordinary Least Squares (1)	Weighted Least Squares (2)	Ordinary Least Squares (3)	Weighted Least Squares (4)	Ordinary Least Squares (5)	Weighted Least Squares (6)	Ordinary Least Squares (7)	Weighted Least Squares (8)
White	5.953**	5.074**	7.278**	7.597**	2.863	1.828	4.740*	4.253
	(1.990)	(2.159)	(3.016)	(3.531)	(2.561)	(3.680)	(2.639)	(3.134)
Mean outcome before NCLB in states without prior accountability	232	233	76	77	281	282	74	76
Black	4.582	15.378**	8.431	22.690**	9.261	8.826	9.977	10.004
	(5.436)	(3.710)	(6.693)	(6.199)	(6.774)	(8.999)	(7.886)	(11.955)
Mean outcome before NCLB in states without prior accountability	203	202	35	33	241	242	28	28
Hispanic	12.409**	11.625**	12.499*	25.883**	20.031**	8.219**	22.006**	18.692**
	(4.540)	(1.572)	(6.334)	(2.779)	(5.766)	(4.135)	(4.618)	(4.666)
Mean outcome before NCLB in states without prior accountability	204	204	40	36	246	247	36	36
Free Lunch Eligible	6.934*	9.734**	11.186*	17.256**	10.702*	15.761**	12.773*	23.432**
	(3.604)	(2.836)	(5.769)	(4.986)	(6.155)	(5.631)	(7.328)	(6.398)
Mean outcome before NCLB in states without prior accountability	212	212	49	49	257	256	47	46
Not Free Lunch Eligible	3.916	2.603	5.388	6.832**	2.199	0.992	3.152	2.392
	(3.102)	(2.907)	(4.435)	(3.118)	(3.924)	(4.171)	(4.045)	(3.478)
Mean outcome before NCLB in states without prior accountability	232	234	76	78	279	281	72	74

Notes: Each cell is a separate regression. WLS weights by student enrollment. See text for model details. ***p<0.01, ** p<0.05, * p<0.1.

Table 3 - The Estimated Effects of NCLB on Per-Pupil Expenditures by Function and Revenues by Source

Dependent Variable	Dependent Mean for 1999-2000 school year (SD)	Estimated NCLB Impact
Current K-12 Expenditures (per-pupil)	$8,357 (2,109)	$733*** (206)
Instructional Expenditures (per-pupil)	$5,171 (1,379)	$504*** (115)
Support-Service Expenditures (per-pupil)	$2,819 (831)	$256*** (99)
Other Expenditures (per-pupil)	$367 (112)	-$28 (23)
Federal Revenues (per-pupil)	$619 (491)	$11 (30)
State and Local Revenues (per-pupil)	$9,199 (2,380)	$616** (266)

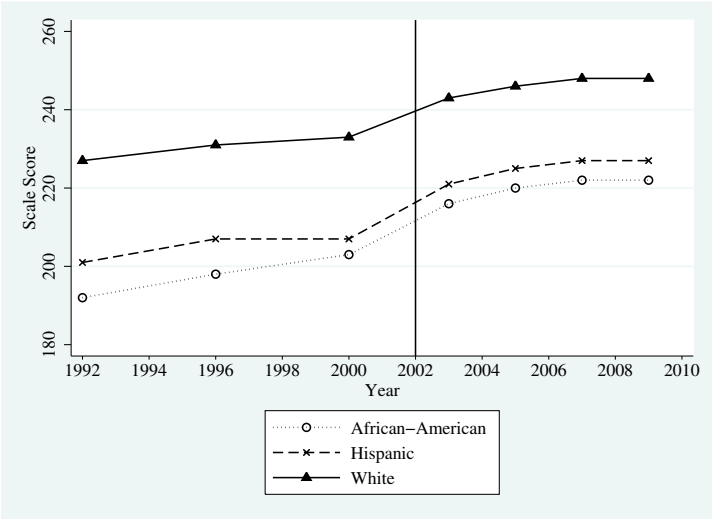
Notes: Each cell is a separate regression based on roughly 140,000 district-year observations. Outcomes are in 2009 dollars. See text for model details. *** p<0.01 ** p<0.05 * p<0.1.

Table 4 - The Estimated Effects of NCLB on Teacher, Classroom, and School Traits, Pooled Schools and Staffing Surveys (SASS)

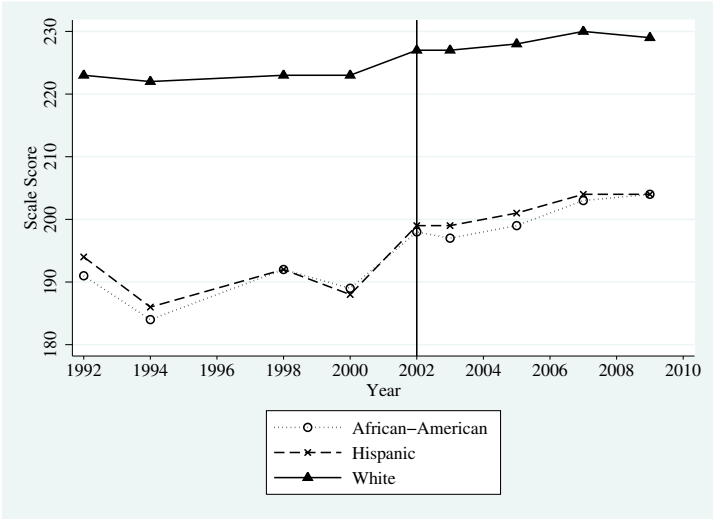
Dependent Variable	Dependent Mean for 1999-2000 school year (SD)	Estimated NCLB Impact
School Resources		
Class Size	22.120	-0.328
	4.990	(0.500)
Master's Degree	0.412	0.056**
	0.492	(0.028)
Instructional Time		
Total hours per week spent on academic subjects (math, English social studies, and science)	21.758	-0.307
	6.445	(0.684)
% Time on Math and English	0.737	0.036***
	0.130	(0.012)
% Time on Math	0.261	0.013
	0.118	(0.010)
School Climate		
Principal considers academic excellence or basic skill acquisition to be the highest priority	0.875	-0.003
	0.331	(0.037)
Teachers' perceptions of school discipline	-0.003	0.074
	0.989	(0.115)
Teachers' perceptions of student engagement	0.059	0.220***
	0.990	(0.056)

Notes: Each cell is a separate regression. See text for model details. ***p<0.01, **p<0.05 *p<0.1.

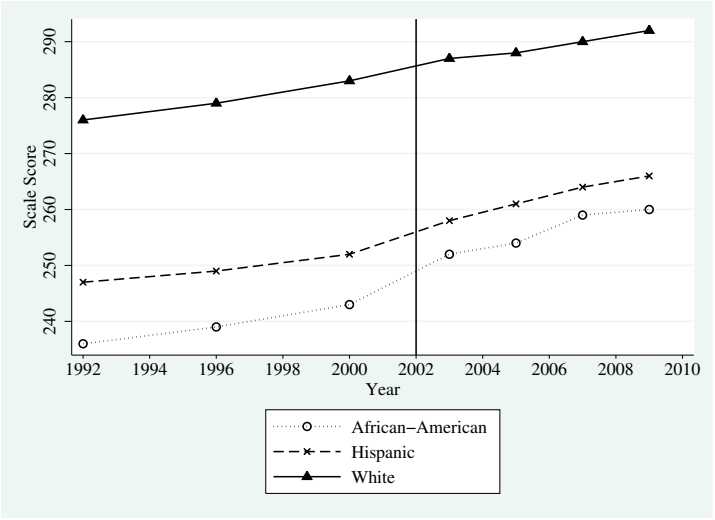
Figure 1: Mean scaled score on the main NAEP for all public schools



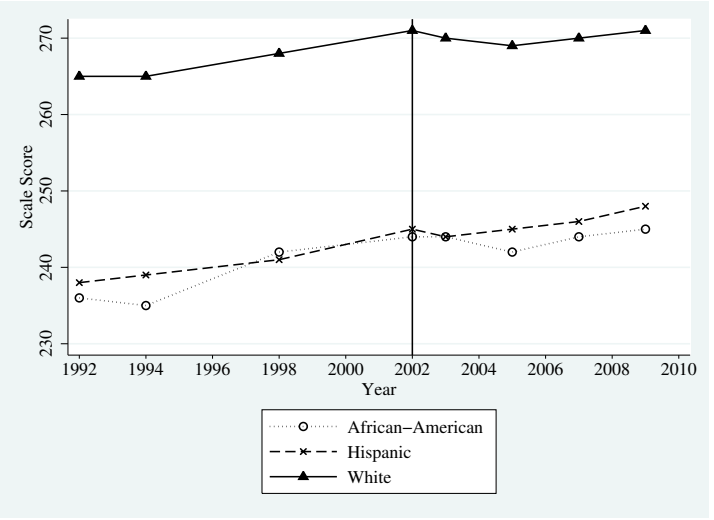
(a) Grade 4 mathematics



(b) Grade 4 reading

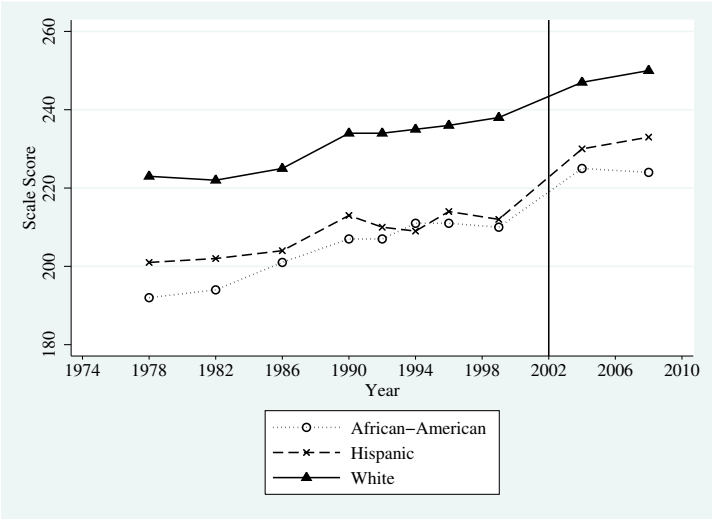


(c) Grade 8 mathematics

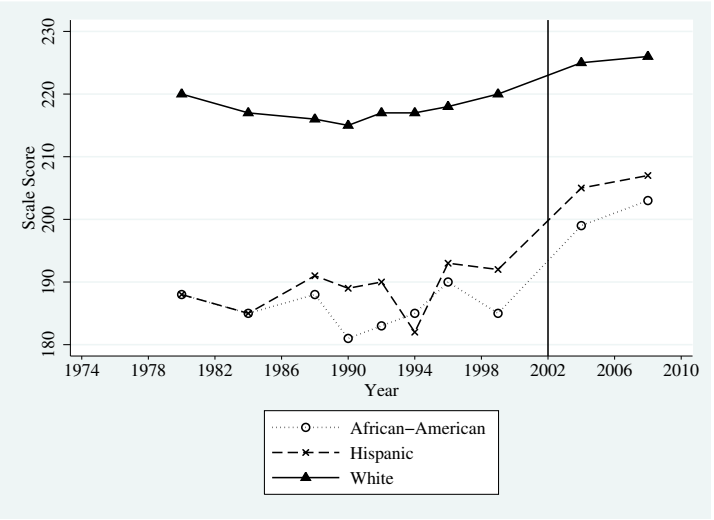


(d) Grade 8 reading

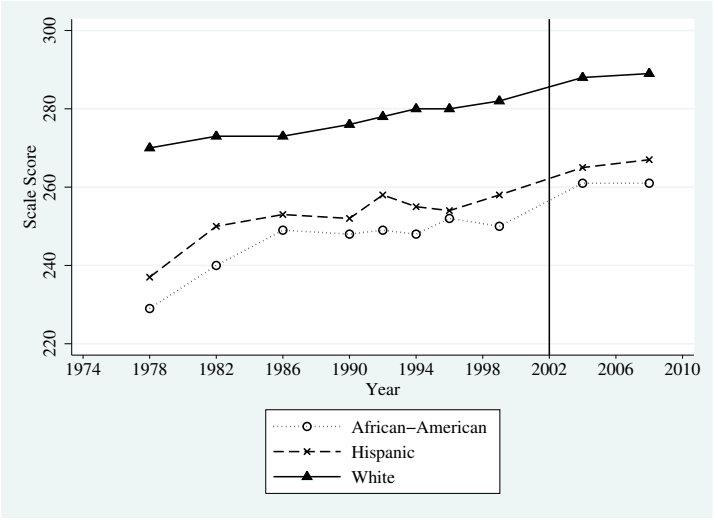
Figure 2: Mean scaled score on the Long-Term Trend NAEP for all public schools



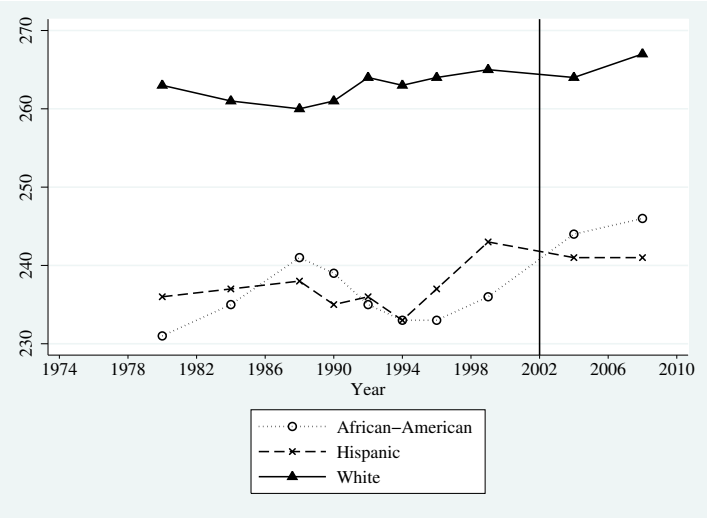
(a) Age 9 mathematics



(b) Age 9 reading

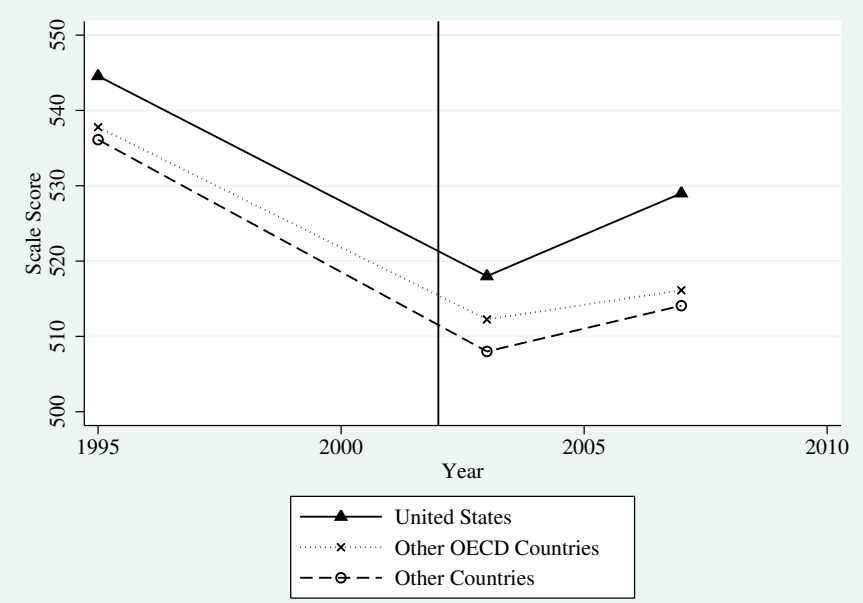


(c) Age 13 mathematics

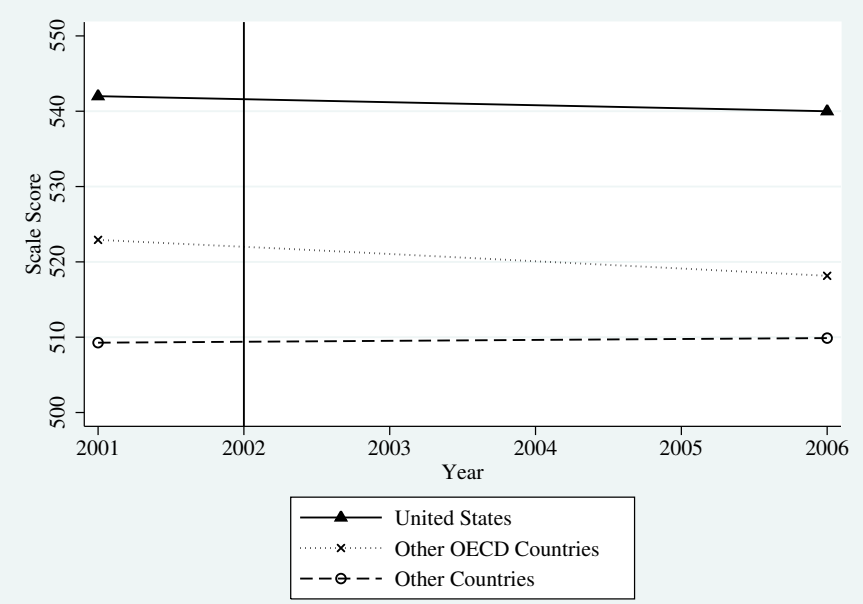


(d) Age 13 reading

Figure 3: Achievement Trends in United States and Other Countries



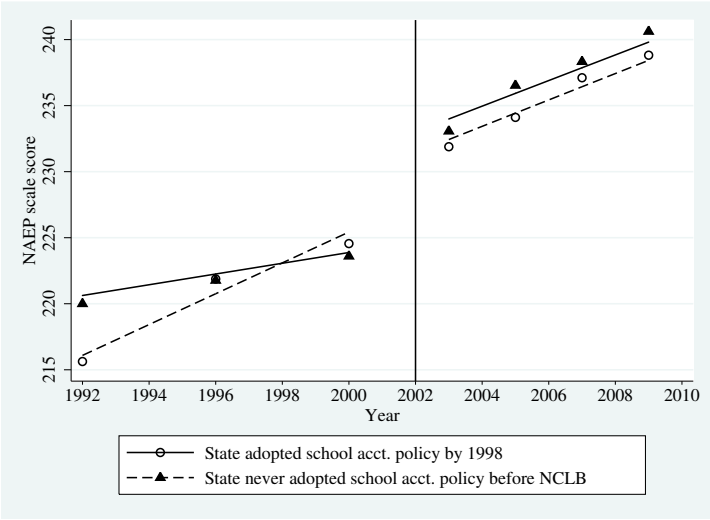
(a) Grade 4 mathematics, all countries (TIMSS)



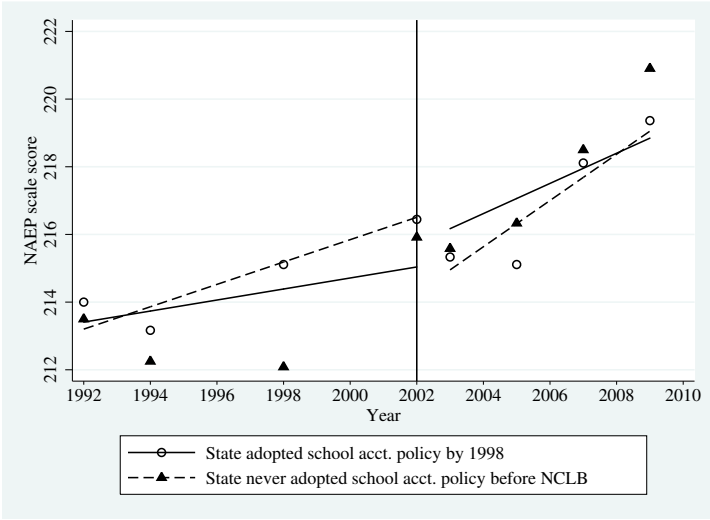
(b) Grade 4 reading, all countries (PIRLS)

Notes: For the TIMSS data, other countries include all countries that participated in all three TIMSS surveys: Australia, England, Hungary, Iran, Japan, Latvia, Netherlands, New Zealand, Norway, Scotland, Singapore, and Slovenia. The other OECD countries exclude Iran, Latvia, Singapore, and Slovenia, as well as the U.S. For the PIRLS data, other countries include all countries that participated in all both PIRLS surveys: Bulgaria, England, France, Germany, Hong Kong, Hungary, Iran, Israel, Italy, Kuwait, Latvia, Lithuania, Macedonia, Moldova, Morocco, Netherlands, New Zealand, Norway, Romania, Russia, Scotland, Singapore, Slovak Republic, Slovenia, and Sweden. The other OECD countries exclude Bulgaria, Hong Kong, Iran, Israel, Latvia, Lithuania, Macedonia, Moldova, Morocco, Romania, Russia, Singapore, and Slovenia, as well as the U.S.

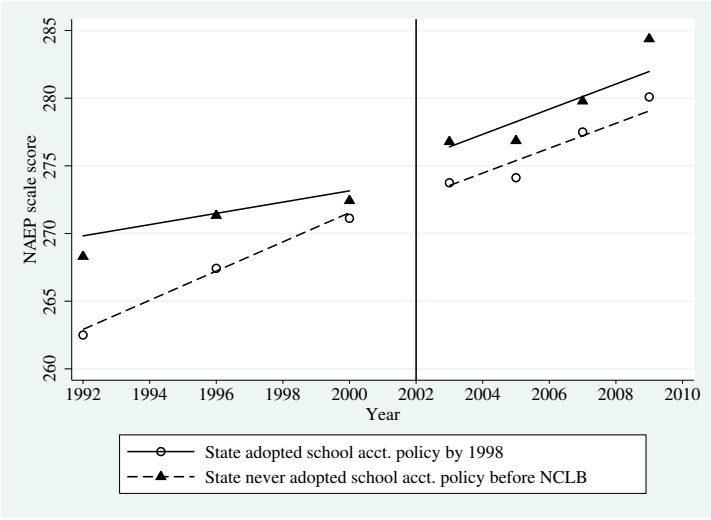
Figure 4: Trends in Achievement in the Main NAEP by Timing of Accountability



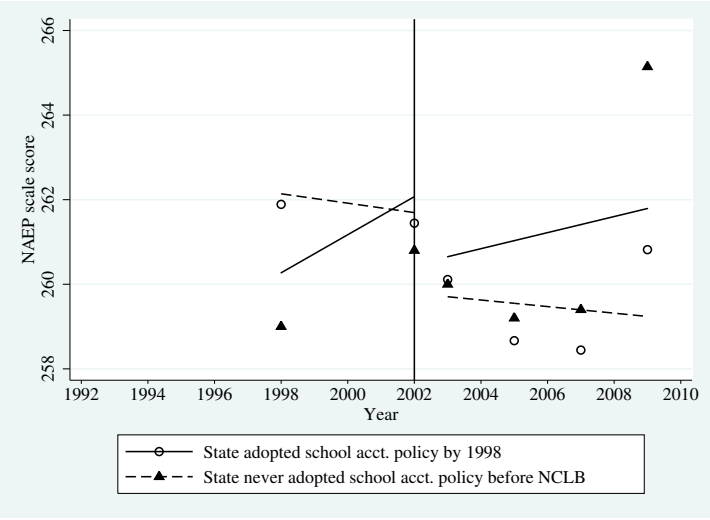
(a) Average scale score, Grade 4 mathematics



(b) Average scale score, Grade 4 reading

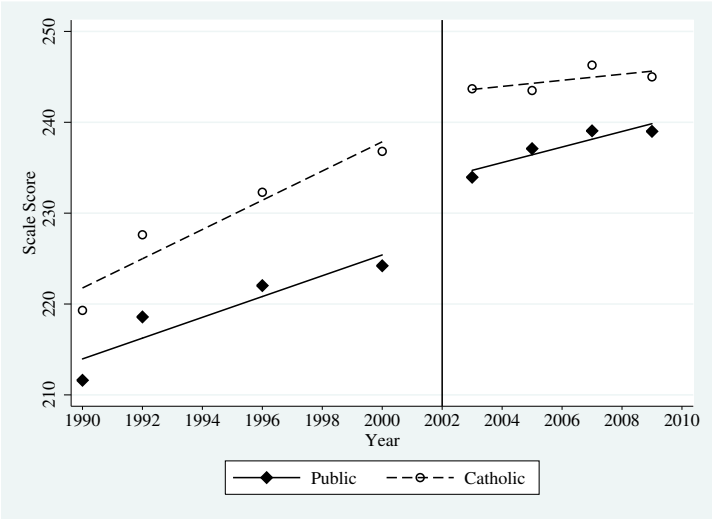


(c) Average scale score, Grade 8 mathematics

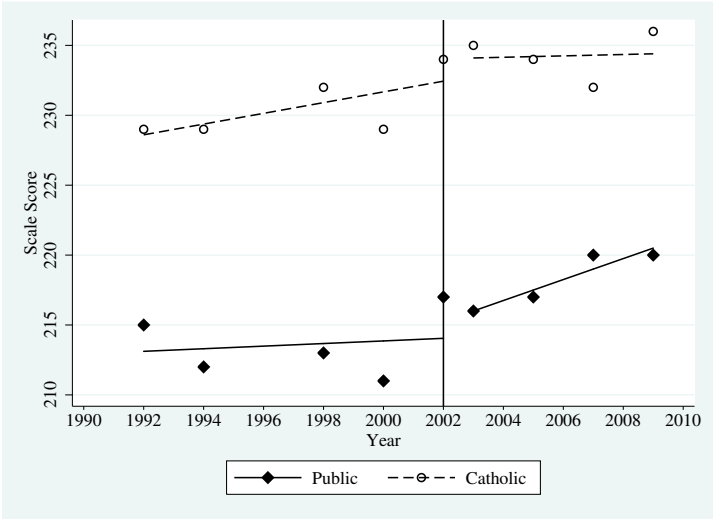


(d) Average scale score, Grade 8 reading

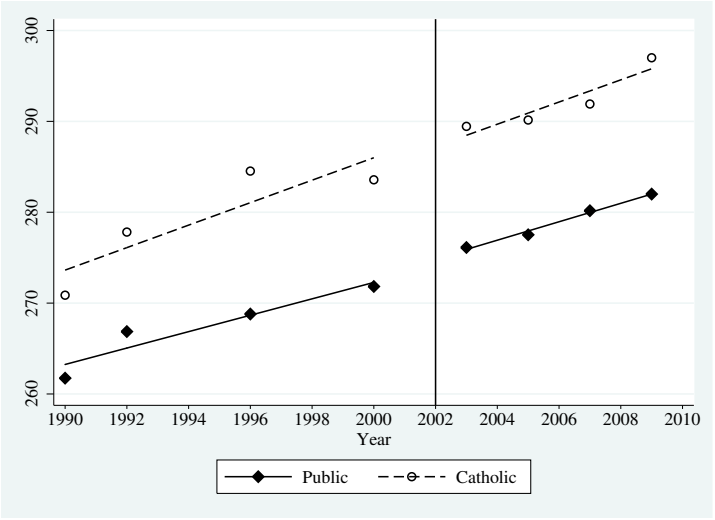
Figure 5: Main NAEP Achievement Trends in Public versus Catholic Schools



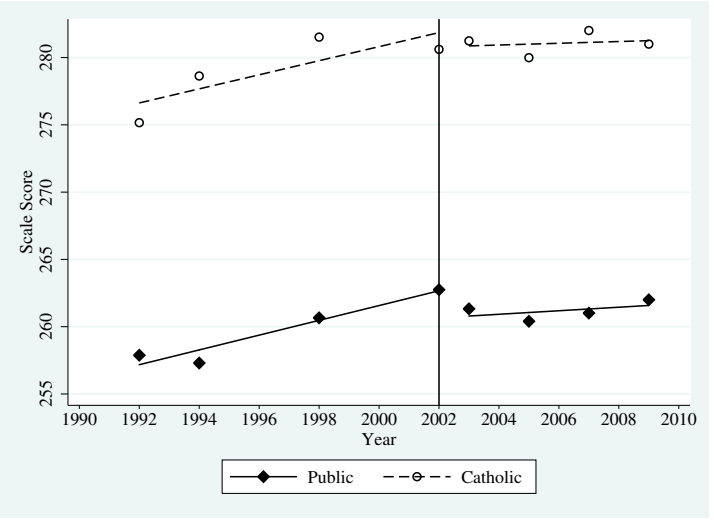
(a) Grade 4 mathematics



(b) Grade 4 reading

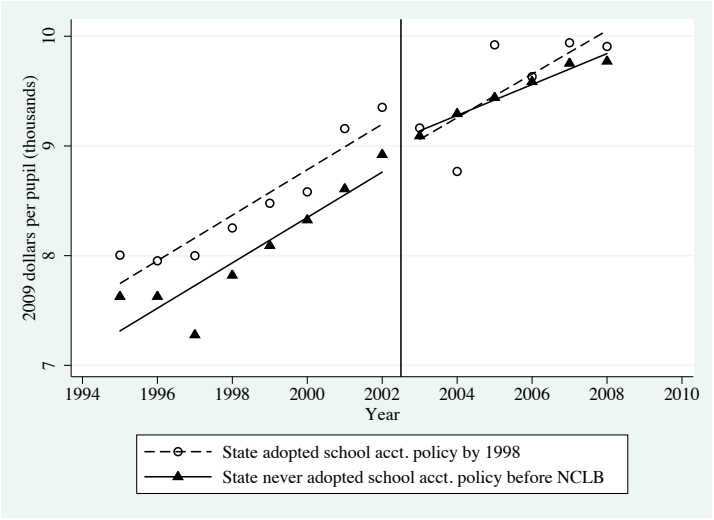


(c) Grade 8 mathematics

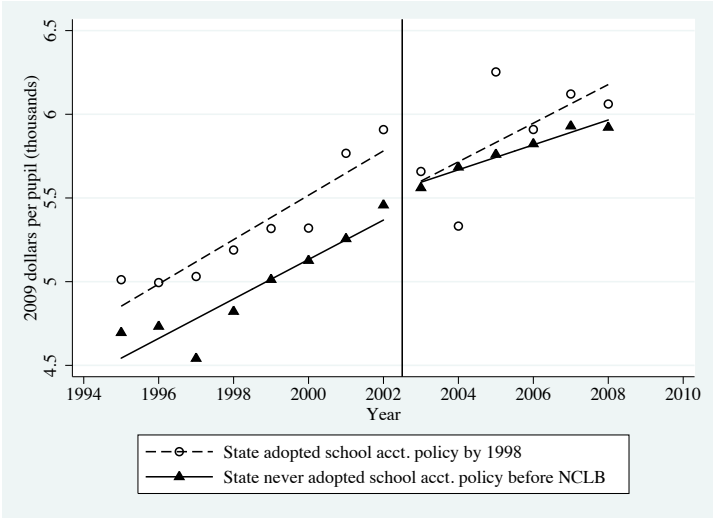


(d) Grade 8 reading

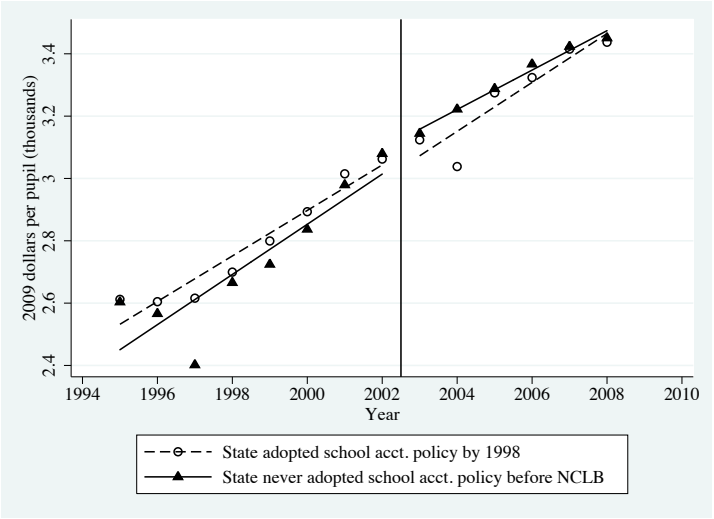
Figure 6: Trends in District Expenditures by Timing of Accountability



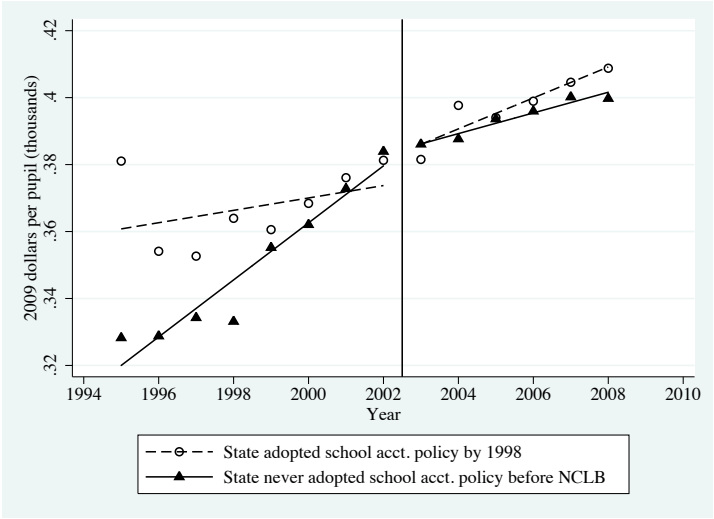
(a) Total Elementary/Secondary Expenditures



(b) Instructional Expenditures



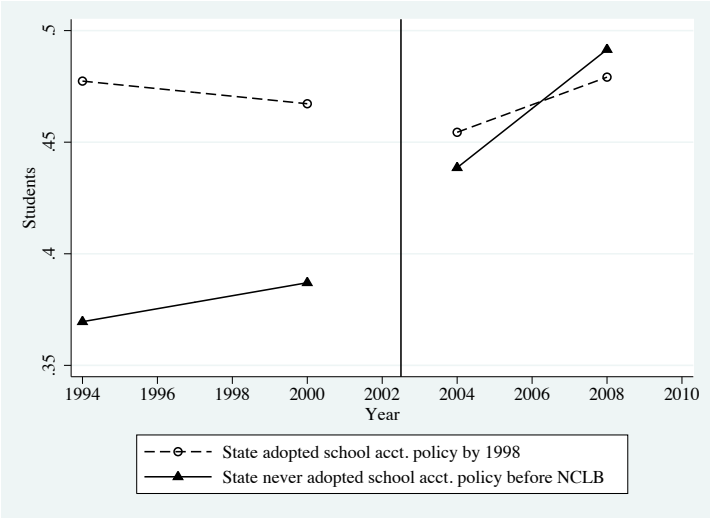
(c) Support Service Expenditures



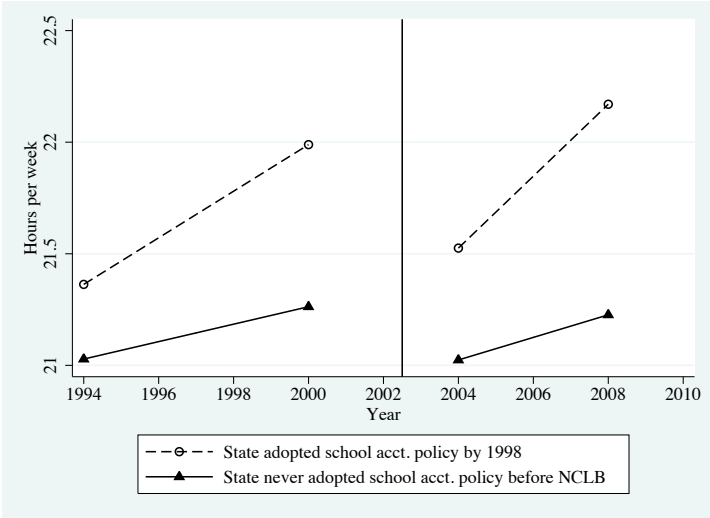
(d) Other Elementary/Secondary Expenditures

Notes: Data drawn from the Common Core of Datas Local Education Agency (School District) Finance Survey. Sample is composed of all non-charter, unified LEAS, excluding Hawaii, the District of Columbia, and all zero-enrollment districts. Estimates are weighted by district enrollment.

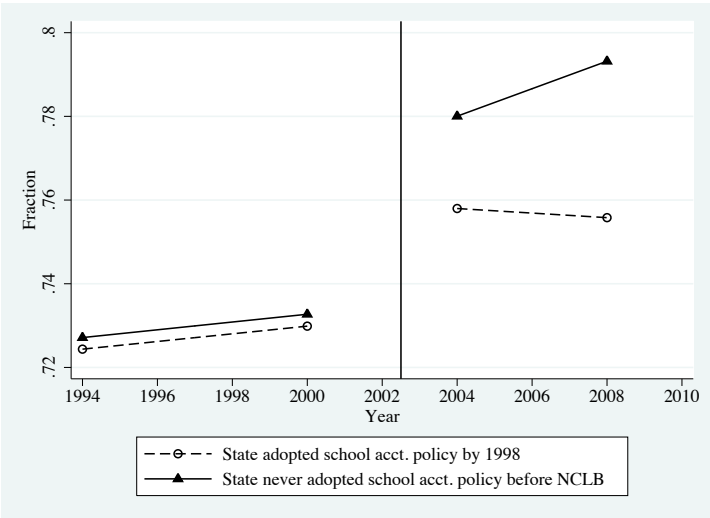
Figure 7: Trends in School Resource and Time Use Use by Timing of Accountability



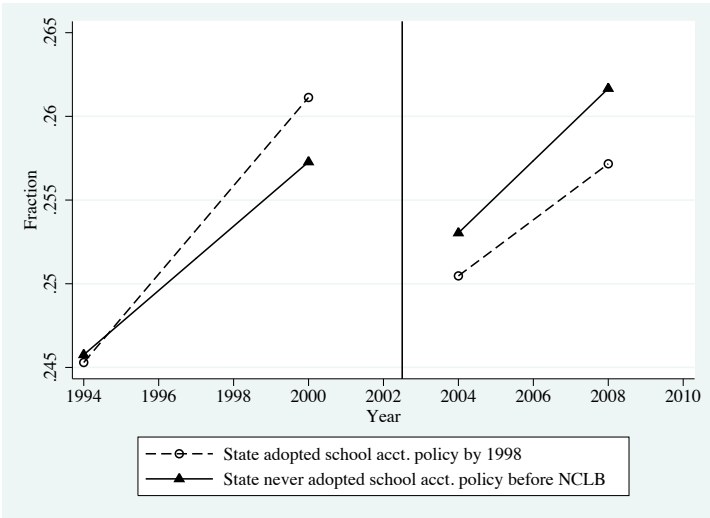
(a) Fraction of teachers with a Masters Degree



(b) Total academic time (math+English+science+social studies)



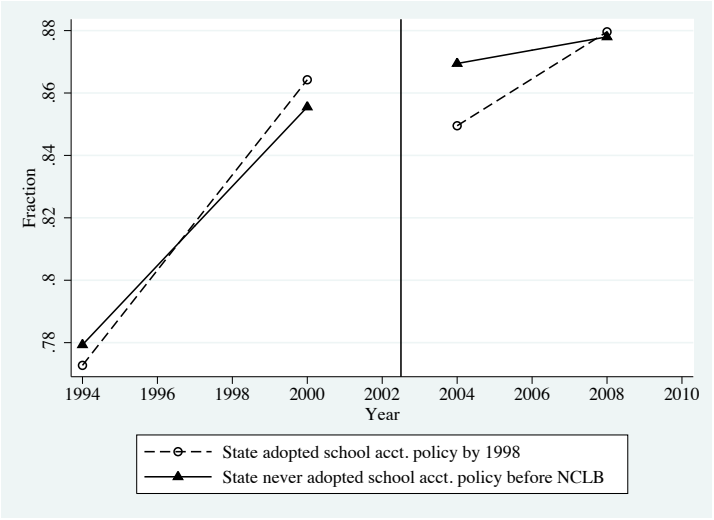
(c) Ratio of math and English time to total academic time



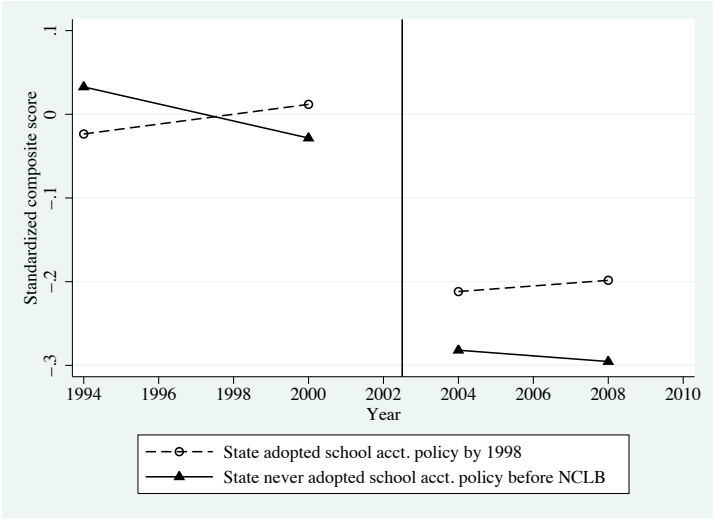
(d) Ratio of math time to total academic time

Notes: Data drawn from the Schools and Staffing Survey. Sample is composed of full-time elementary and middle school teachers with a main assignment in either Mathematics, English/Language Arts, or General Elementary. Means are weighted using NCES-generated weights and adjusted for individual and school covariates, as well as state-fixed effects. Graphs (b), (c) ,and (d) are limited to teachers teaching in non-departmentalized (e.g. self-contained or team-taught) classrooms.

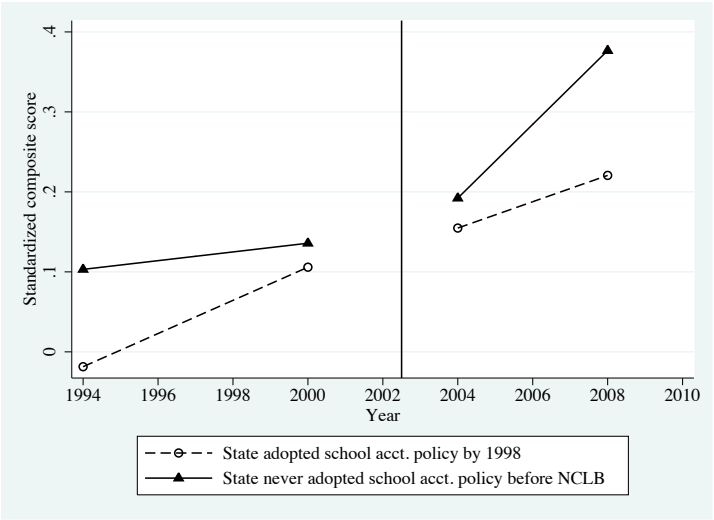
Figure 8: Trends in School Culture Outcomes by Timing of Accountability



(a) Fraction of principals who consider academic excellence or basic skills to be their number one goal



(b) Teachers' perception of school discipline (higher score indicates greater enforcement of rules)



(c) Teachers' perceptions of student engagement (higher score indicates greater engagement)

Notes: Data drawn from the Schools and Staffing Survey. Data for graph (a) is composed of full-time elementary and middle school principals. See notes for figure 7 for sample definition for graphs (b) and (c).