

Educational Evaluation and Policy Analysis

<http://eepe.aera.net>

The Effects of NCLB on School Resources and Practices

Thomas S. Dee, Brian Jacob and Nathaniel L. Schwartz

EDUCATIONAL EVALUATION AND POLICY ANALYSIS 2013 35: 252 originally published online 28 December 2012

DOI: 10.3102/0162373712467080

The online version of this article can be found at:

<http://eepe.aera.net/content/35/2/252>

Published on behalf of



[American Educational Research Association](http://www.aera.net)

and



<http://www.sagepublications.com>

Additional services and information for *Educational Evaluation and Policy Analysis* can be found at:

Email Alerts: <http://eepe.aera.net/alerts>

Subscriptions: <http://eepe.aera.net/subscriptions>

Reprints: <http://www.aera.net/reprints>

Permissions: <http://www.aera.net/permissions>

>> [Version of Record](#) - Apr 29, 2013

[OnlineFirst Version of Record](#) - Dec 28, 2012

[What is This?](#)

The Effects of NCLB on School Resources and Practices

Thomas S. Dee

*Stanford University
National Bureau of Economic Research*

Brian Jacob

*University of Michigan
National Bureau of Economic Research*

Nathaniel L. Schwartz

Tennessee Department of Education

A number of studies have examined the impact of school accountability policies, including No Child Left Behind (NCLB), on student achievement. However, there is relatively little evidence on how school accountability reforms and NCLB, in particular, have influenced education policies and practices. This study examines the effects of NCLB on multiple district, school, and teacher traits using district-year financial data and pooled cross sections of teacher and principal surveys. Our results indicate that NCLB increased per-pupil spending by nearly \$600, which was funded primarily through increased state and local revenue. We find that NCLB increased teacher compensation and the share of elementary school teachers with advanced degrees but had no effects on class size. We also find that NCLB did not influence overall instructional time in core academic subjects but did lead schools to reallocate time away from science and social studies and toward the tested subject of reading.

Keywords: *accountability, NCLB, school finance, teacher compensation*

THE movement toward test-based school accountability is arguably the most striking and controversial change to occur over the past several decades in the governance of U.S. public

education. Beginning with the introduction of state-level accountability systems in the 1990s, schools have increasingly faced government regulations concerning the level of their students'

We would like to thank Rob Garlick, Elias Walsh, 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. All errors are our own.

achievement. The federal government brought test-based accountability to scale with the implementation of the No Child Left Behind Act (NCLB) during the 2002–2003 school year. NCLB required both annual student testing and school-level reporting that indicated whether schools were succeeding or failing to make “adequate yearly progress” (AYP) toward proficiency goals.

The research literature on school accountability, including both NCLB and earlier state-level reforms, suggests that these policies have had at least some meaningful but targeted success in improving student achievement (e.g., National Research Council, 2011). Yet there is also evidence of unintended and unproductive responses to these policies, including direct cheating on the part of teachers and administrators, as well as various attempts to shape the test-taking population to raise measured outcomes (Neal & Schanzenbach, 2010; Jacob & Levitt, 2003). Although such research suggests some broad impacts of school accountability policies, we know little about the mechanisms by which these changes are taking place. The extant evidence linking school accountability, particularly the most recent NCLB reforms, to the policies and practices within schools is quite limited. Much of the research in this area simply depends on reports from teachers and administrators about how accountability policies have influenced their school practice. This type of descriptive evidence could reflect framing biases because it includes neither data from the pre-reform era nor a credible control group to allow causal inference about the impact of particular policies.

This study provides new evidence on the question of how schools change in the face of test-based accountability. Specifically, we use detailed longitudinal data on district finances collected by the U.S. Census Bureau to examine the ways that NCLB changed patterns of revenue and spending. We complement these results with several years of pooled data from the Schools and Staffing Survey (SASS) to look at NCLB effects on multiple district, school, and teacher traits. These results provide insight into the mediating mechanisms associated with the student achievement effects of NCLB. In contrast with much previous work in this line, our

study estimates the effects of NCLB using panel data from multiple states rather than focusing on a single state and its pre-NCLB accountability reforms. Following Dee and Jacob (2011), we utilize a comparative interrupted time series (CITS) design that effectively compares the changes within states where NCLB compelled the implementation of new school accountability systems (i.e., treatment states) to the contemporaneous variation in states that had preexisting systems of consequential school accountability. In addition, we also draw our data from repeated surveys of general teacher-, school-, and district-level practices, which may have better construct validity than surveys that have specifically asked respondents to self-report the ways in which NCLB has changed school practice.

Our findings fall into three major areas: (a) finance and conventional resources, (b) time use, and (c) school climate. We find that NCLB increased district spending by roughly \$570 per student (2009 dollars). We find that the increased expenditures induced by NCLB were allocated to both direct instruction and pupil-support services. It is difficult from the available data to identify precisely how this money has been used. However, we do find evidence that teacher compensation increased meaningfully in response to NCLB (i.e., by \$5,000 per year), particularly in high-poverty school districts. Neither class sizes nor pupil–teacher ratios appear to have fallen as a result of NCLB, although we do see a significant rise in the percentage of teachers with master’s degrees.

Second, we are able to confirm reports that elementary and middle school teachers have significantly reallocated their time as a result of NCLB’s school accountability policies. We find strong evidence that the share of instructional time given to mathematics and English/language arts has increased, with corresponding decreases in the share going to science and social studies per week.

Third, we consider NCLB’s effects on several measures of school climate. Although several qualitative studies have suggested that school accountability has tended to create more test-focused environments within schools, we are unable to find evidence that NCLB has shifted principals’ priorities around student progress or influenced the disciplinary climate

within the school. It is interesting that we do find evidence that NCLB led to improvements in an index of teacher-reported student behaviors (e.g., absenteeism, tardiness, and apathy) commonly categorized as “behavioral engagement” with school. The remainder of the article proceeds as follows. The second section reviews the prior literature. The third section explains our research design, and the fourth section describes the data we use. In the fifth section, we present our results, and we discuss our conclusions in the sixth section.

Literature Review

In January 2002, President Bush signed into law NCLB, dramatically expanding the scope of federal involvement in public K–12 schools. NCLB effectively brought to scale school accountability systems like those that had been implemented in several states during the prior decade.

Causal Research on NCLB

Identifying the effects of NCLB presents a challenge. Because NCLB’s requirements were simultaneously applicable to schools nationwide, a credible “control group” is not readily apparent. Studies of NCLB’s effects on student achievement offer two potential solutions to this problem. One approach has been to examine the effects associated with being in a school that is close to AYP failure under state accountability rules. For example, Reback, Rockoff, and Schwartz (2010) adopt this approach using nationally representative data from the Early Childhood Longitudinal Study. 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. Similarly, Ballou and Springer (2008) find that comparative student performance improved in those grade-year combinations relevant for a state’s AYP calculations. Yet other recent studies using state- and city-specific data suggest the effects of school accountability may be non-existent, or even negative, for students who are not near the high-stakes proficiency threshold (Krieg, 2008; Neal & Schanzenbach, 2010).

A concern with these AYP-threshold studies, however, is that they may only capture a partial

impact of the policy because they rely on differences in sanction relevance among schools, all of which may have been influenced by NCLB to some extent. Dee and Jacob (2011) adopt an alternative approach based on a CITS design that compares deviations from *preexisting* trends following the introduction of NCLB in states with and without consequential school accountability prior to NCLB. Using state-year panel data from the low-stakes National Assessment of Educational Progress (NAEP), they find evidence that NCLB led to improvements in math performance, particularly among fourth graders (effect size = 0.23). They find that these math gains existed for both low- and high-performing students but were most pronounced among Black, Hispanic, and free-lunch-eligible students. They find no evidence that NCLB improved reading scores among fourth graders and note that the data available for eighth grade reading did not allow for a convincing examination of student progress in that area.

Causal research into NCLB’s effects on factors beyond student achievement is much more limited. We know of only three prior studies that have used regression analysis to isolate the effects of school accountability policies on district, school, and classroom practices from the potentially confounding effects of other determinants. Each of these use the AYP-threshold method described above.

Reback et al. (2010) present evidence at the national level that teachers in schools facing NCLB accountability pressure worry about their job security and (at least in reading) increase the time they allocate to test preparation. Rouse, Hannaway, Goldhaber, and Figlio (2007) used a regression-discontinuity design and data from principal surveys in Florida to examine how schools responded to pressure from the state’s accountability system. The authors 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 (e.g., block scheduling).

These studies suggest that NCLB has caused substantial changes in both the ways that schools are organized and the ways that teachers are behaving. As noted previously, however, since these studies compare schools facing different

levels of sanction threat, they may identify only what constitutes a partial impact of school accountability. Moreover, they each possess their own limitations. For example, the study by Rouse et al. (2007) relies on a comparatively rich set of outcome measures but may have attenuated external validity because it is limited to Florida. In contrast, the study by Reback et al. (2010) leverages the data from a nationally representative survey but examines relatively few school and teacher process measures.

Other Research on the Effects of School Accountability

If we extend our review beyond directly causal studies of NCLB, we find several other useful sources of information. These include both an extensive set of studies on local school accountability policies before NCLB and several more recent surveys that ask participants to review the ways that NCLB might have changed their workplace. Our review covers three major areas: (a) school and district finances, (b) instructional time, and (c) staff practices.

With respect to school and district finances, most of the research predates NCLB but suggests that the policy would be likely to raise spending and directly affect the allocation of resources. In case studies of four states that all implemented comprehensive standards-based reform and accountability programs in the 1990s, Hannaway, McKay, and Nakib (2002) find that two of the states (Kentucky and Texas) increased educational expenditures substantially more than the national average, and disproportionately allocated the increase to instruction, but that two other states (Maryland and North Carolina) did not. Hannaway and Stanislawski (2005) also 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, though it is difficult to determine whether the accountability policy caused the increased expenditures or they were merely part of a broader reform agenda. Bifulco (2010) offers additional evidence on the financial effects of accountability with the finding that pre-NCLB state accountability raised novice teacher salaries relative to veteran teachers

in the same district. This suggests the possibility that districts pursue new teacher-recruitment strategies in response to accountability.

A number of studies have looked at the relationship between school accountability and the allocation of instructional time, offering evidence that accountability causes educators to reallocate time toward tested subjects, toward specific content areas within subjects, and toward particular types of test preparation activities. As in the literature on financial resources, however, much of the work in this area has examined accountability policies predating NCLB. Equally problematic is that much of the work in this area has relied on teachers' retrospective reports of how accountability policies influenced their work, making causal attribution uncertain (see, e.g., Koretz, Barron, Mitchell, & Stecher, 1996; Koretz & Hamilton, 2006; Koretz, Mitchell, Barron, & Keith, 1996; Pedulla et al., 2003; Stecher, Barron, Kaganoff, & Goodwin, 1998; Jacob, 2005; Swanson & Stevenson, 2002; Taylor, Shepard, Kinner, & Rosenthal, 2003).¹

Survey-based studies that focus on NCLB itself find similar results. The Center on Education Policy (CEP) has studied the implementation and impact of NCLB since its inception (CEP, 2006, 2007, 2008). As part of its work, CEP not only surveyed a nationally representative sample of school districts in 2005–2006 and again in 2006–2007 but also conducted more intensive case studies of selected school districts. District officials, particularly those in urban and high-poverty districts, report that NCLB increased the instructional time they devote to math and English/language arts (ELA) and decreased the time they devote to other subjects (CEP, 2006, 2008).² In related work, researchers at RAND collected data in 2005 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 for their state assessment system. In addition, educators also claimed that they responded to NCLB by increasing the alignment between the curriculum and state standards (also see Murnane & Papay, 2010).

Our final area of interest is how school accountability and NCLB in particular might have affected staff practices. By setting clear and coherent benchmarks for student progress, school accountability systems were meant to motivate educators and students toward common goals and increasingly effective practices (Smith & O'Day, 1991). In general, there has been little research on whether and how NCLB has influenced such outcomes since they tend to defy easy measurement. 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). Teachers reported that teaching practices and the general focus on student learning "changed for the better" under accountability (Hamilton et al., 2007). Other studies have reported related changes in the actions and perceptions of school administrators. For example, in the RAND study, 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. District officials in the CEP study similarly reported an increase in the use of data to guide instruction (CEP, 2006).

Research Design

Following Dee and Jacob (2011), we use a CITS approach to examine the effects of NCLB on education finance as well as several measures of instructional practice and school climate. The CITS specifications we estimate effectively compare the deviation from prior outcome trends among a "treatment group" of observations to the analogous deviation for observations from a "comparison group." The intuition is that the deviations from trends in the comparison group will reflect other hard-to-observe and potentially confounding factors (e.g., the economy, other education reforms) that may have influenced student achievement in the absence of NCLB. This general strategy has a long tradition in education research (see, e.g., the discussion in Bloom, 1999, and Shadish, Cook, & Campbell, 2002), and has been used recently to evaluate reforms as diverse as

Accelerated Schools (Bloom, Ham, Melton, & O'Brien, 2001) and pre-NCLB accountability policies (Jacob, 2005).

The central challenge for any CITS design is to identify a plausible comparison group that was unaffected by the intervention under study. In the case of NCLB, this is particularly difficult. It simultaneously applied to all public schools in the United States but with particularly explicit sanctions for schools receiving federal Title I funds. Here, we rely on the fact that several states actually introduced school accountability policies similar to those catalyzed by NCLB but in different years *prior* to NCLB. The fundamental intuition behind this approach is that NCLB represented less of a "treatment" (or a nonexistent treatment) in states that had already adopted NCLB-like school accountability policies in the years prior to 2002. Stated differently, to the extent that NCLB-like accountability had either positive or negative effects on any of our outcome measures, we would expect to observe those effects most distinctly in the "treatment" states that had *not* previously introduced similar policies.³

This approach relies on the assertion that pre-NCLB school accountability policies were comparable to NCLB—that is, the two types of accountability regimes are similar in the most relevant respects. To ensure that this is the case, we categorize states according to whether the features of their pre-NCLB accountability policies closely resemble the key aspects of NCLB. Although we relied on a number of different sources to categorize pre-NCLB accountability policies across states (including studies of such policies by Carnoy & Loeb, 2002; Hanushek & Raymond, 2005; Lee & Wong, 2004), the taxonomy developed by Hanushek and Raymond (2005) is particularly salient in this context because it most closely tracked the key school accountability features of NCLB.

We reviewed their coding with information from a variety of sources including the Quality Counts series put out by *Education Week* ("Rewarding Results, Punishing Failure," 1999), the state-specific Accountability and Assessment Profiles assembled by the Consortium for Policy Research in Education (Goertz & Duffy, 2001), annual surveys on state assessment programs fielded by the Council of Chief State School

Officers, information from state department of education websites, LexisNexis searches of state and local newspapers, and conversations with academics and state officials in several states. Our review generally confirmed their coding for the existence and timing of these state “consequential accountability” policies and indicated that these state policies did closely resemble the frameworks required under NCLB.⁴ Table A1 lists the states that we determined had implemented a consequential accountability policy prior to NCLB.

Following the intuition of the CITS research design we have outlined, we estimate the following regression model,

$$Y_{st} = \beta_0 + \beta_1 YEAR_t + \beta_2 NCLB_t + \beta_3 (YR_SINCE_NCLB_t) + \beta_4 (T_s \times YEAR_t) + \beta_5 (T_s \times NCLB_t) + \beta_6 (T_s \times YR_SINCE_NCLB_t) + \beta_7 X_{st} + \mu_s + \varepsilon_{st} \quad (1)$$

where Y_{st} is an outcome measure observed for state s in year t , $YEAR_t$ is a trend variable (defined as $YEAR - 1989$ so that it starts with a value of 1 in 1990), and $NCLB_t$ is a dummy variable equal to 1 for observations from the NCLB era. For the majority of our analysis, we follow the conventional assumption that the NCLB era began in the academic year 2002–2003, the first full academic year after the legislation was signed in January 2002.⁵ $YR_SINCE_NCLB_t$ is defined posttreatment as $YEAR - 2002$, so that this variable takes on a value of 1 for the 2002–2003 year and 0 for all previous years. X_{st} represents covariates varying within states over time. The variables, μ_s and ε_{st} represent state fixed effects and a mean-zero random error respectively.

T_s is a time-invariant state-level variable that measures the treatment imposed by NCLB. In a most basic application, T_s would simply be a dummy variable that identifies whether a given state had *not* instituted consequential accountability prior to NCLB. Such a 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). In such a basic specification, the total estimated NCLB effect as of 2008 would be $\hat{\beta}_5 + 6 \times \hat{\beta}_6$.

Although this simple case highlights the intuition behind our approach, there are ways in

which it is probably more accurate to view the “treatment” provided by the introduction of NCLB in the framework of a dosage model. In particular, slightly more than half of the states that introduced consequential school accountability prior to NCLB did so just 4 years or fewer prior to NCLB’s implementation. Given the number of states that implemented consequential accountability shortly before the implementation of NCLB, the simple binary definition of T_s defined above could lead to attenuated estimates of the NCLB effect. That is, the comparison group includes some states for which the effects of prior state policies and NCLB are closely intertwined.

To address this concern, we define T_s in our primary specification 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 1991–1992 academic year and the onset of NCLB. Hence, states with no school accountability at all prior to NCLB would have the highest value for the treatment measure, T_s (i.e., 11). In contrast, Illinois, which adopted its policy in the 1992–1993 school year, would have a value of only 1. Texas would have a value of 3 because its policy started in 1994–1995, and Vermont would have a value of 8 because its program started in 1999–2000. Our identification strategy implies that the larger the value of this treatment variable, the greater the potential impact of NCLB. In specifications based on this construction of T_s , we define the impact of NCLB as of 2008 (e.g., the most recent SASS data available) and relative to a state that introduced consequential accountability in 1997 (i.e., $6 \times \beta_5 + 36 \times \beta_6$).⁶

When using data from the SASS, available in only four unique periods rather than annually, we modify our approach slightly to include both the CITS described above as well as conventional “difference in differences” specifications that do not condition on pre-NCLB trends. Specifically, we estimate the following model,

$$Y_{st} = \beta_0 + \beta_1 NCLB_{2004} + \beta_2 NCLB_{2008} + \beta_3 (T_s \times NCLB_{2004}) + \beta_4 (T_s \times NCLB_{2008}) + \beta_5 X_{st} + \mu_s + \varepsilon_{st}$$

where Y_{st} is an outcome measure observed for state s in year t . $NCLB_{2004}$ is a dummy indicator

for observations from the 2003–2004 SASS survey and $NCLB_{2008}$ is a dummy indicator for observations from the 2007–2008 SASS survey. As in our CITS specification, T_s is a time-invariant state-level variable that measures the treatment imposed by NCLB. To compare results across the CITS specification and the difference in differences (DD) specification, we present our DD results as the total estimated NCLB effect as of 2008 relative to a state that implemented consequential accountability in 1997—that is, $6 \times \beta_4$. Since it is difficult to say without more data whether the CITS or the DD approach provides the most accurate approach, we choose to present both results. It is reassuring that we found this approach generated quite similar results with the exception of one finding, which we discuss along with our other CITS results.

There are several important threats to causal inference in our study design. In particular, our key identifying assumption is that the deviations from prior outcome trends within the comparison states (i.e., those with lower values of T_s) provide a valid counterfactual for what would have happened in “treatment states” if NCLB had not been implemented. To be clear, this assumption is not violated by the presence of time-invariant state traits, nor is it violated by the presence of pre-NCLB trends that are related to treatment status, T_s . However, the internal validity of this identification strategy would be violated if there were unobserved determinants of our outcome measures that varied both contemporaneously with the onset of NCLB and uniquely with respect to treatment status, T_s . For example, this might occur if the socioeconomic traits of students and their families tended to change as NCLB was implemented in treatment states, but not in control states.

The presence or absence of such unique time-specific and state-specific unobservables cannot be definitively established. However, we provide indirect evidence on this important question by reporting the results of auxiliary regressions like Equation 1 but where the dependent variables are state-year measures of observed traits that are themselves significantly associated with district spending levels and thus may influence our outcome measures (e.g., parental education, poverty rate, and median

household income). The estimated “effect” of NCLB on these measures provides evidence on whether observables appear to vary along with the adoption of NCLB in a manner that could confound our key CITS inferences.

We see no significant variation in any tested observables when using our yearly finance data (see Table 5 below). In contrast, we do find a somewhat puzzling apparent increase in median household income when we restrict our analysis to only the four available years of SASS data in our CITS specification (Table B2). This suggests that the lack of more frequent yearly reporting from the SASS renders inferences based on these data somewhat less reliable. These findings of apparently causal impacts on other observed traits multiply when we use a DD rather than CITS approach, giving us reason to prefer the CITS estimates. It is reassuring, however, that our main SASS results do not change significantly when we control for time-varying state characteristics such as median household income.⁷

In addition to this evidence and the previously discussed robustness checks, we also assess the sensitivity of our CITS results to different pretrend specifications. Since we do not know how to best model the initial trends in our CITS estimations, we include results using quadratic rather than linear pretrends. Similarly, to avoid the possibility that our data might be skewed by initial trends that did not carry through into the NCLB years, we estimate results that do not include data from 1995 and 1996.

Finally, it is important to consider how to interpret the resulting estimates. First, our estimates will capture only the impact of the *accountability* provisions of NCLB and, moreover, only those accountability provisions that were unique to treatment states. These estimates will not reflect the impact of other NCLB provisions such as Reading First or the “highly qualified teacher” provision, which were effectively new policy mandates for all states. Second, under the maintained assumption that NCLB was effectively irrelevant in states with prior consequential accountability systems, our estimates will identify the effects of NCLB-induced school accountability provisions that are specific to those 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, our estimates may understate the average treatment effects of school accountability. Similarly, our CITS will also understate the general effects of school accountability if NCLB amplified the effects of school accountability within the comparison states (e.g., the focus on subgroup performance might have strengthened the preexisting accountability systems in certain states). An alternative concern is that the accountability systems within comparison states may have been weakened as they were adjusted in response to NCLB (e.g., states may have been forced to abandon a successful state-developed school performance system and focus instead of AYP). To the extent this occurred, our CITS approach would instead overstate the effects of NCLB. We suspect this concern is not empirically relevant because the school reporting and performance sanctions occasioned by NCLB (e.g., the possibility of school reconstitution or closure) were strong relative to prior state accountability policies. However, we also examined this issue empirically by identifying how the proficiency standards adopted in various states may have changed after the implementation of NCLB. As detailed in Appendix A (available online at <http://epa.sagepub.com/supplemental>), this analysis confirmed that NCLB did not seem to have a substantive impact in states that had previously adopted consequential school accountability.

Data

The Common Core of Data

Our first source of data is the Common Core of Data's Local Education Agency (School District) Finance Survey, also known as the F-33 survey. The survey has been administered annually since 1994–1995 and includes detailed financial data for all school districts in the United States, along with district enrollment totals. Unfortunately, the data are available only at the district level and do not allow further disaggregation to determine how district budgets were divided among different schools within a district. However, even though many of NCLB's

accountability provisions are focused at the school level, the structure of school finances often requires a districtwide response, and thus district effects remain relevant.

All of the monetary data available in the pooled F-33 surveys were converted to real 2009 dollars using the Consumer Price Index. Our analytical sample consists of district-by-year observations of all regular, operational, unified (K–12) school districts and for each school year between 1994–1995 and 2007–2008. This conventional sample construction creates more comparable units of observation by excluding districts that operate only elementary or secondary schools, districts that are purely administrative in nature, and agencies that operate only charter schools. For similar reasons, we excluded Hawaii and the District of Columbia where the entire jurisdictions are collapsed into single districts. The financial data in the F-33 surveys also include extreme outliers, which appear to reflect both the idiosyncratic nature of some agencies and reporting or coding errors (e.g., miscoded decimal places). Following the practice in earlier research using these data (e.g., Murray, Evans, & Schwab, 1998), we excluded extreme outliers by dropping observations where real revenues per pupil were greater than 150% of the state-specific 95th percentile value or less than 50% of the state-specific 5th percentile value. (i.e., roughly one half of 1% of the district-year panel observations). We also recoded as missing the pupil–teacher ratios and instructional salaries where the variable was greater than 150% of the state's 95th percentile or less than 50% of the state's 5th percentile.

Using the sample described above, we merged the district data with other publicly available data to add some additional information into the set. From the Common Core of Data's Public Elementary/Secondary School Universe Survey Data, we merged in the percentage of students designated as Black or Hispanic within the district as well as the total instructional full-time equivalents. This latter variable allowed us to calculate average instructional compensation as well as pupil–teacher ratios. Outliers for these values were recoded as missing according to the formula described above. From the School District Demographics System provided by the National Center for

TABLE 1
Common Core of Data District Sample Characteristics

	Full sample	State Quartile 1 (lowest poverty rate)	State Quartile 2	State Quartile 3	State Quartile 4 (highest poverty rate)
District enrollment (hundreds)	719.38	264.86	317.77	422.05	1,816.10
% Black or Hispanic	33.37	18.00	25.67	36.02	56.36
Poverty rate (2000 census)	0.12	0.06	0.10	0.14	0.21
Total current K–12 expenditures	8.73	8.28	8.37	8.43	9.79
Instructional expenditures	5.41	5.14	5.15	5.17	6.15
Social service expenditures	2.94	2.81	2.85	2.88	3.23
Other elementary/secondary expenditures	0.37	0.34	0.37	0.38	0.42
Total revenue	10.17	9.87	9.79	9.74	11.18
Federal revenue	0.75	0.44	0.62	0.78	1.20
State and local revenue	9.43	9.43	9.17	8.96	9.98
Average teacher salary	63.63	62.34	61.39	61.15	68.99
Pupil–teacher ratio	17.05	17.10	17.14	16.99	16.95
<i>N</i> (district-years)	146,238	36,844	36,622	36,565	36,207

Note. All revenue and expenditure variables are per pupil and are reported in thousands of 2009 dollars.

Education Statistics, we brought in district poverty rates as recorded on the 2000 census. Less than 0.1% of the sample was missing either student ethnicity or year 2000 poverty rates, and we dropped these observations from our sample.

Our final analytical sample consists of 142,607 district-by-year observations, reflecting roughly 10,000 unified school districts observed over 14 school years. We provide basic descriptive statistics on this sample in Table 1. We conducted our analyses both on the full sample of data as well as on subsamples designated by poverty-rate quartile within state. Because we used poverty rates from the 2000 census to create these subsamples, these quartile groups remain stable throughout the period of our study.

The Schools and Staffing Survey

Our second source of data is the SASS. Administered every 4 years, the SASS is the nation's largest comprehensive source of data on school organization and staff perceptions. During each round of data collection, the National Center for Education Statistics surveys

schools, teachers, and principals, creating national and state-representative data files. The primary sampling unit for the survey is the school, and schools are selected and assigned sampling weights based on sector, location, school level, and school population. Once chosen, schools provide teacher listings and teachers are similarly stratified and assigned sampling weights based on their subject areas and experience levels. Prior research has used the SASS data to identify trends in teacher qualifications, teacher autonomy, and various labor market outcomes (e.g., Ingersoll, 1999, 2006; Liu, 2007).

To create a longitudinal panel of data that allows us to identify trends before and after NCLB, we made use of survey questions that were repeated across four SASS administrations: 1993–1994, 1999–2000, 2003–2004, and 2007–2008. We divided these data into a series of stratified samples based on teacher and school characteristics. We chose to focus our primary analysis on teachers and principals working in either elementary or middle schools.⁸ Although most of NCLB's regulations apply equally to all

school levels, its testing requirements are concentrated in elementary and middle school grades and its achievement effects appear to have been concentrated among younger students (Dee & Jacob, 2011).⁹

Within the group of elementary and middle teachers and principals, we keep all public school principals but limit our teacher sample to full-time, public school teachers with a main assignment in mathematics, ELA, or general elementary. We drop 8% of teacher-year observations and 6% of principal-year observations that are missing school information. Our final sample, summarized in Table 2, includes approximately 36,000 teacher-year observations and 16,500 principal-year observations. All our statistical estimates are weighted using the appropriate teacher and principal weights.

Outcomes

Our outcomes are divided into four categories based on our review of the literature: (a) expenditures and revenue, (b) conventional resources, (c) use of instructional time, and (d) school climate. We examine total current K–12 expenditures per pupil and the allocation of these expenditures across three broad functions defined by the F-33 (instructional, support services, and other) as well as district-level revenues by source (i.e., federal, and state or local). We also examine several conventional measures of instructional resources: pupil–teacher ratios and total teacher compensation (both from the Common Core of Data [CCD]) and class sizes and the fraction of teachers who hold a master’s degree (both from the SASS).

We then look at the allocation of instructional time within the academic subjects. Because teachers who teach departmentalized classes do not answer questions about academic time use on the SASS, we limit our time-use analysis to only teachers who teach in self-contained or team-taught classrooms. For these teachers, we consider the number of hours per week spent on core academic subjects including math, ELA, science, and social studies.

Finally, we look at outcomes representing teachers’ and principals’ perceptions of their school environment. Principals surveyed by the SASS are asked to order their three most important goals from a list of nine possible choices

that range from “building basic literacy skills (reading, math, writing, speaking)” or “encouraging academic excellence” to “promoting specific moral values.” As an indicator of the extent to which NCLB increased schools’ focus on student achievement, we created a variable indicating the percentage of principals who chose either basic literacy skills or academic excellence as their most important goal. The SASS also asks teachers to use 5-point scales to answer a series of questions on topics ranging from their colleagues’ enforcement of school rules to the extent that student absenteeism causes problems in the school. We aggregated related questions into a school-discipline composite variable and a behavioral-engagement composite variable. These composites are the sum of teachers’ responses on related questions within years, standardized using base-year means and standard deviations. In addition to providing details about all SASS variables, Appendix C (available online at <http://epa.sagepub.com/supplemental>) provides a full description of each composite outcome and the process by which it was created.

Our measure of teachers’ perceptions of the school disciplinary environment combines two survey questions to teachers, one about whether rules are enforced by other teachers in the building even outside their own classrooms and one about the extent of the principal’s support for school rules. Responses were standardized using base-year data and thus begin at zero. Our measure of teachers’ perceptions of student engagement combines questions about whether various student factors, including apathy, tardiness, class cutting, absenteeism, and coming to class unprepared, cause problems for the school.

Results

In the following sections, we compare treatment and control states as described previously. For each set of outcomes, we present a series of figures that graphically illustrate our analysis approach as well as regression estimates that formalize the intuition presented in the figures.¹⁰

District Expenditure and Revenue Results

Our analysis suggests that school accountability under NCLB significantly increased

TABLE 2

Schools and Staffing Survey (SASS) Teacher and Principal Sample Characteristics

	Full sample	High-poverty sample
Teachers		
% Male	10.9	11.4
Age	42.1	41.6
% Black	7.9	13.7
% Hispanic	6.5	12.0
School enrollment (hundreds)	5.9	5.9
% of school approved for free/reduced lunch	44.3	73.8
% of school designated as minority	39.0	62.1
% departmentalized	21.7	20.2
Academic time per week (math + ELA + science + social studies)	21.5	21.8
Percentage of academic time given to math	25.4	25.4
Percentage of academic time given to ELA	49.9	50.7
Percentage of academic time given to science	12.0	11.7
Percentage of academic time given to social studies	12.7	12.1
Teachers' perception of school discipline (standardized using base year)	-0.1	-0.1
Teachers' perception of student culture (standardized using base year)	0.1	0.0
Class size	22.3	22.1
Salary	41,999	41,001
% MA	43.7	40.4
Observations (teacher-years)	36,169	14,543
Principals		
% Male	48.9	43.4
Age	49.0	49.3
% Black	11.0	19.2
% Hispanic	5.8	10.5
School enrollment (hundreds)	4.8	4.8
% of school approved for free/reduced lunch	43.4	72.9
% of school designated as minority	34.9	57.0
% prioritizing skills/excellence	85.4	86.6
Observations (principal-years)	16,552	6,474

Note. Means combine SASS results from 1993–1994, 1999–2000, 2003–2004, and 2007–2008 and are calculated using appropriate weights. The teacher sample includes all elementary and middle school full-time teachers whose main assignment is “math,” “ELA,” or “general elementary.” The principal sample includes all elementary and middle school teachers. For both groups, the high-poverty sample is defined as those teachers and principals who teach in schools where more than 50% of students are approved for free lunch.

districts' spending. Figure 1 illustrates this effect. In Figure 1a, for example, we see that total per-pupil expenditures rose more quickly from 1994 to 2002 in states that adopted pre-NCLB accountability policies. But following the introduction of NCLB, spending grew more quickly in the states where NCLB mandated accountability for the first time. Figures 1b and 1c show comparable results for the two largest categories of total expenditures, instructional spending, and support service spending. In contrast, a treatment effect is not apparent in the

small residual category of other K–12 expenditures (i.e., Figure 1d).

Table 3 presents regression estimates that formalize the intuition presented in the figures above. All models include state fixed effects and show standard errors clustered at the state level. The first row of the table shows our preferred specification, which includes a series of district-level covariates and weights the estimate by district enrollment totals. According to this specification, by 2008 NCLB increased total current expenditures in states with no pre-NCLB

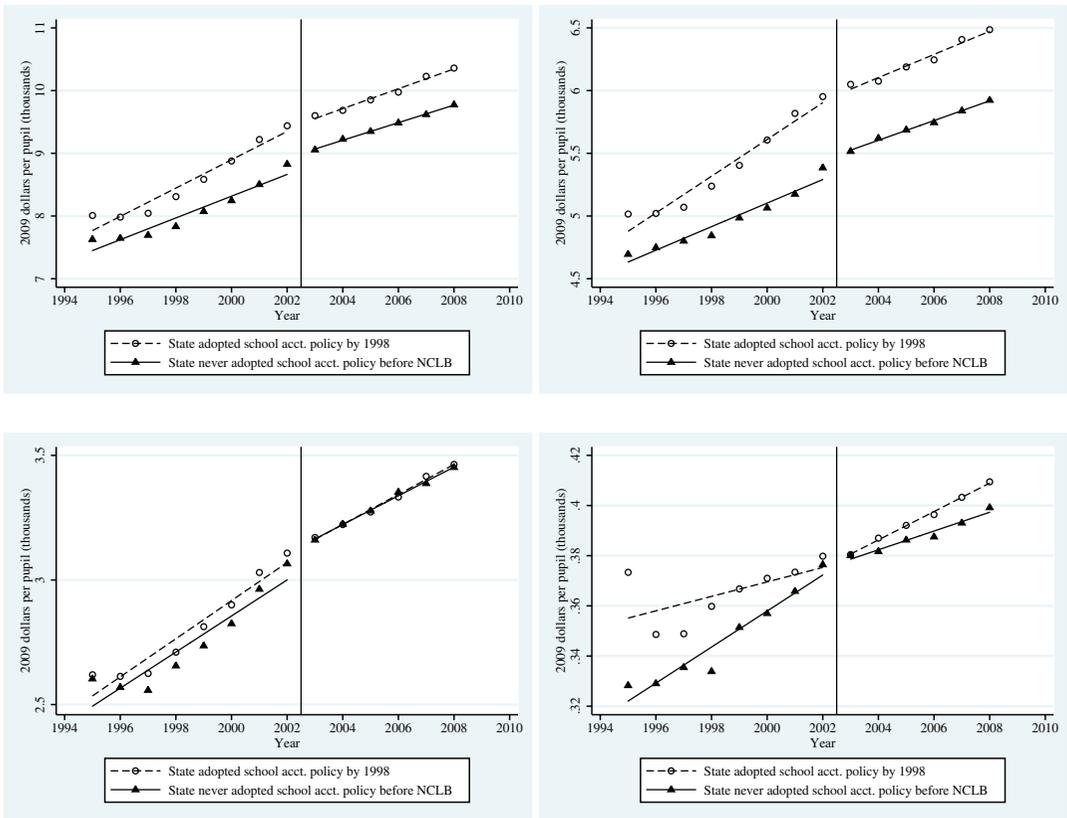


FIGURE 1. Trends in district expenditure outcomes by timing of accountability policy.

Note. Data drawn from the Common Core of Data's Local Education Agency (School District) Finance Survey. Sample is composed of all noncharter, unified LEAs, excluding Hawaii, the District of Columbia, and all zero-enrollment districts. Estimates are weighted by district enrollment. Outlier districts where real revenues per pupil were greater than 150% of the state-specific 95th percentile value or less than 50% of the state-specific 5th percentile value were also excluded. Total elementary/secondary expenditures is equal to the total of instructional, support service, and other elementary or secondary expenditures.

accountability by \$570 per pupil relative to states that adopted school accountability in 1997. It is important that most of this increase in spending (\$430) went directly toward instructional use, although spending for student and staff support services also rose. The remaining rows in the table show results without additional district-level covariates and without weighting by enrollment. We also present results using log-spending outcomes. Regardless of specification, the results are qualitatively the same.¹¹

How did public schools pay for these increased expenditures? A provision in NCLB states that states and school districts would not be required "to spend any funds or incur any costs not paid for under this act." However, many states and school districts have argued that

the legislation did, in practice, constitute an unfunded mandate. In fact, a survey of superintendents and principals found that nearly 90% agreed with the "unfunded mandate" characterization of NCLB (Olson, 2003). And several school districts and the state of Connecticut pursued legal challenges to NCLB, emphasizing this point (Hoff & Walsh, 2008)

To examine this question, we estimated separate CITS models for models of per-pupil revenues from federal and state/local sources. The revenue results, which are presented in the final three columns of Table 3, suggest that NCLB led to no substantial change in federal revenue for those states without prior accountability systems. The point estimates on federal revenue are quite small and not statistically significant. The

TABLE 3

District Expenditure and Revenue Results

	Real per-pupil expenditures				Real per-pupil revenue		
	Total current K–12 expenditures	Instructional expenditures	Support service expenditures	Other current K–12 expenditures	Total revenue	Federal revenue	State/local revenue
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Actual outcomes, with covariates, weighted by enrollment							
Effect by 2007–2008	0.570** (0.237)	0.430*** (0.137)	0.155 (0.112)	–0.015 (0.025)	0.490* (0.281)	0.042 (0.031)	0.448 (0.288)
Actual outcomes, no covariates, weighted by enrollment							
Effect by 2007–2008	0.579** (0.231)	0.436*** (0.135)	0.158 (0.108)	–0.015 (0.025)	0.498* (0.278)	0.039 (0.031)	0.458 (0.284)
Actual outcomes, no covariates, no weights							
Effect by 2007–2008	0.550** (0.257)	0.399** (0.183)	0.190** (0.085)	–0.038 (0.045)	0.436 (0.340)	0.028 (0.043)	0.408 (0.332)
Log outcomes, with covariates, weighted by enrollment							
Effect by 2007–2008	0.061** (0.025)	0.081*** (0.024)	0.041 (0.034)	–0.048 (0.056)	0.040 (0.026)	0.073 (0.049)	0.038 (0.028)
Mean for 1999–2000 school year	8.360	5.209	2.786	0.365	9.816	0.660	9.155
<i>SD</i> for 1999–2000 school year	2.061	1.428	0.772	0.111	2.284	0.473	2.250
<i>N</i> (district-years)	146,238	146,238	146,238	145,639	146,238	146,153	146,238

Note. Each cell is a separate regression. The dependent variables are defined per pupil, in 2009 dollars and in natural log form where specified. Total current K–12 expenditures is the sum of instructional expenditures, support service expenditures, and other expenditures. Total revenue is the sum of federal and state/local revenue. The total NCLB effect by 2008 is relative to a state with school accountability starting in 1997. All specifications include state fixed effects. Covariates, where specified, include a quadratic in the percentage of the district that is Black or Hispanic, a quadratic in the district poverty rate from the 2000 census, and an interaction between poverty rate and the percentage Black or Hispanic. Standard errors (in parentheses) are clustered at the state level.

* $p < .1$. ** $p < .05$. *** $p < .01$.

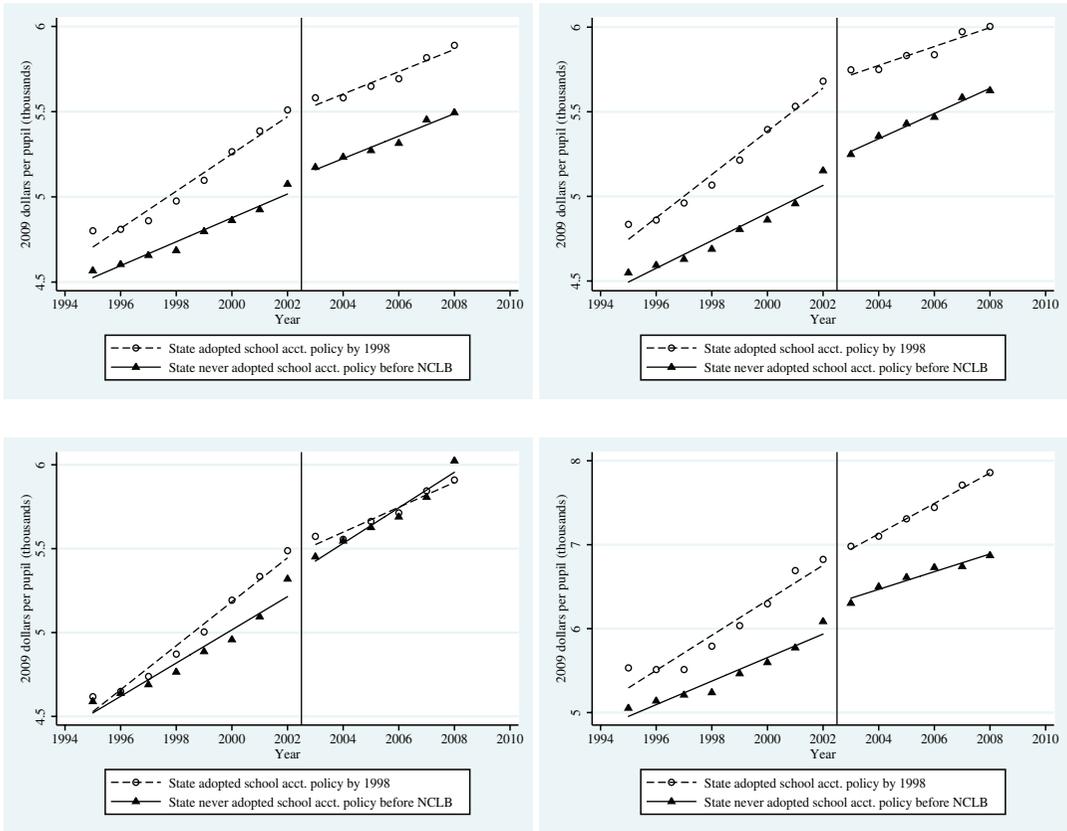


FIGURE 2. Trends in instructional expenditures by timing of accountability policy and power status.
 Note. Data drawn from Common Core of Data. See note to Figure 1 for sample details. Poverty quartiles defined within state according to the district's relative poverty level on the 2000 census.

precision of this estimate implies that the upper bound on the 95% confidence interval for the change in federal revenues is only about \$100 per pupil, less than 20% of the corresponding increase in district spending. In contrast, these results suggest that NCLB increased state and local revenues per pupil by \$448, a point estimate that is just shy of weak statistical significance ($p = .126$). Overall, these results suggest that most, if not all, of the spending increases catalyzed by NCLB in states without prior accountability systems were paid for at the state and local levels.

In Figure 2 and Table 4, we investigate heterogeneity in our district expenditure and revenue results by poverty levels within each school district. Our poverty breakdowns rely on the district's poverty rate according to the 2000 decennial census, a designation we hold constant

across all years to generate consistent quartiles across our sample. Quartiles are defined separately within each state. One of the primary objectives of NCLB was to reduce inequities in student performance by race and socioeconomic status. However, our results imply that expenditure and revenue results were not driven by any single quartile but took place across the board. One way of understanding these results is that NCLB's accountability demands are binding even on low-poverty districts, perhaps because even the low-poverty districts tend to have some members of each recognized student subgroup and therefore needed to respond to the new policies. Another is that the law prompted an across-the-board funding response regardless of its targeted population. Our point estimates indicate that the absolute dollar increase in poor districts may have been larger than in other

TABLE 4

District Expenditure and Revenue Results by Poverty Quartile

	Poverty Quartile 1 (lowest poverty rate)	Poverty Quartile 2	Poverty Quartile 3	Poverty Quartile 4 (highest poverty rate)
	(1)	(2)	(3)	(4)
Total current K–12 expenditures				
Effect by 2007–2008 school year	0.560** (0.260)	0.733*** (0.284)	0.573** (0.234)	0.561** (0.245)
Mean for 1999–2000 school year	7.977	8.019	8.073	9.283
SD for 1999–2000 school year	1.964	1.909	1.722	2.222
N (district-years)	36,844	36,622	36,565	36,207
Instructional expenditures				
Effect by 2007–2008 school year	0.434*** (0.155)	0.524*** (0.175)	0.442*** (0.140)	0.406*** (0.132)
Mean for 1999–2000 school year	4.966	4.962	4.987	5.850
SD for 1999–2000 school year	1.285	1.280	1.168	1.653
N (district-years)	36,844	36,622	36,565	36,207
Support service expenditures				
Effect by 2007–2008 school year	0.155 (0.114)	0.229** (0.109)	0.170 (0.104)	0.141 (0.154)
Mean for 1999–2000 school year	2.684	2.703	2.721	3.018
SD for 1999–2000 school year	0.780	0.729	0.669	0.816
N (district-years)	36,844	36,622	36,565	36,207
Other current K–12 expenditures				
Effect by 2007–2008 school year	–0.029* (0.017)	–0.020 (0.026)	–0.039* (0.023)	0.015 (0.042)
Mean for 1999–2000 school year	0.328	0.354	0.366	0.414
SD for 1999–2000 school year	0.101	0.102	0.104	0.115
N (district-years)	36,844	36,622	36,565	36,207
Total revenue				
Effect by 2007–2008 school year	0.499* (0.273)	0.642** (0.322)	0.422 (0.270)	0.620** (0.315)
Mean for 1999–2000 school year	9.545	9.508	9.376	10.694
SD for 1999–2000 school year	2.219	2.136	1.989	2.433
N (district-years)	36,844	36,622	36,565	36,207
Federal revenue				
Effect by 2007–2008 school year	0.027 (0.021)	0.021 (0.028)	–0.010 (0.042)	0.100* (0.058)
Mean for 1999–2000 school year	0.378	0.540	0.684	1.067
SD for 1999–2000 school year	0.266	0.315	0.326	0.562
N (district-years)	36,844	36,622	36,565	36,207
State and local revenue				
Effect by 2007–2008 school year	0.472* (0.275)	0.622* (0.323)	0.432 (0.274)	0.521 (0.335)
Mean for 1999–2000 school year	9.167	8.968	8.692	9.627
SD for 1999–2000 school year	2.280	2.179	2.010	2.347
N (district-years)	36,844	36,622	36,565	36,207

Note. Within-state poverty quartiles are designated using school district poverty rates provided by the 2000 census school district demographics database. These results are based on the comparative interrupted time series (CITS) specification described in the text without additional covariates. Standard errors (in parentheses) are clustered at the state level.

* $p < .1$. ** $p < .05$. *** $p < .01$.

districts, although these differences are not significant at conventional levels. These results do suggest that federal revenues increased by a modest, weakly significant amount in the poorest school districts. The increases in state and local revenues attributable to NCLB are also weakly significant in the wealthier districts.

To examine the robustness of these broad findings, we carried out several specification checks,

included in Table 5 and Table B2 in Appendix B (available online at <http://epa.sagepub.com/supplemental>). For example, the trend data in Figure 1 indicate that there was a slight downward trend in school spending in the first years of our district-based sample. To examine the empirical relevance of these nonlinear pretrends, the results in Panel B of Table B2 present our key findings when the first 2 years of sample data

TABLE 5
Common Core of Data Falsification Exercise

	State poverty rate	State median household income	Employment- population ratio \times 100	Fraction in nonpublic schools	% Black	% Hispanic
	(1)	(2)	(3)	(4)	(5)	(6)
Effect by 2007–2008 school year	–0.497 (0.630)	1067.525 (1193.613)	–0.154 (0.279)	–0.137 (0.241)	0.254 (0.368)	–0.029 (0.181)
Mean for 1999–2000 school year	11.914	52710.530	49.581	17.283	17.556	14.609
SD for 1999–2000 school year	2.948	6283.845	2.581	1.407	11.228	14.776
<i>N</i> (district-years)	146,238	146,238	146,238	146,238	146,223	146,223

Note. Each cell is a separate regression. The dependent variables are defined per pupil, in 2009 dollars, and in natural log form where specified. Total current K–12 expenditures is the sum of instructional expenditures, support service expenditures, and other expenditures. Total revenue is the sum of federal and state/local revenue. The total NCLB effect by 2008 is relative to a state with school accountability starting in 1997. All specifications include state fixed effects. Standard errors (in parentheses) are clustered at the state level.

are excluded. The results are quite similar to our full-sample results (Panel A in Table 3), suggesting that these noisy, pre-NCLB trends are not a source of bias. To examine this issue further, Panel C in Table B2 presents results that allow for quadratic pre-NCLB trends. These results are also broadly similar to our baseline results, though these additional covariates do lead to a loss of statistical precision.

As discussed earlier, the CITS design assumes that there were no unobserved determinants of outcome measures that occurred differentially in our treatment versus comparison states at the same time that NCLB was implemented. Table 5 shows the results of several falsification exercises meant to test this assumption. Specifically, we show the estimated “effects” of NCLB on several observed state characteristics, using our CITS specification and our district-year observations. A finding that NCLB appeared to have a significant “effect” on these measures (i.e., poverty rate, median household income, employment-population ratio, fraction of students in public schools, percentage Black, percentage Hispanic) would suggest the existence of possibly confounding factors. The results in Table 5 consistently indicate that this is not the case. That is, we do not find a statistically significant NCLB effect on any of these observables. Although the absence of an NCLB effect on these observables is not definitive, these results do suggest that the estimated effects in Tables 3 and 4 reliably identify the impact of NCLB on per-pupil expenditures and revenues.

Instructional Resources

The key components of per-pupil instructional spending in a district are the pupil–teacher ratio and teacher salaries. The analyses discussed below examine the effect of NCLB school accountability on four measures of instructional resources that are available in the extant data: district-level measures of instructional staff salaries and pupil–teacher ratios from the CCD, and teacher-level measures of class size and master’s degrees from the SASS. Although we do not observe significant results in our teacher–student variables (either pupil–teacher ratio or class size), we do find evidence that NCLB significantly raised both teacher compensation rates and the fraction of teachers with a master’s degree.

In Figure 3a, we see that compensation levels in states that had and had not implemented consequential accountability were on separate tracks before the implementation of NCLB, with the states that had consequential accountability offering teachers several thousand dollars more in compensation. After the law took effect, compensation rates converged to a more similar level across all states although the trends still differed. Our regression estimates, presented in Table 6, suggest that NCLB increased total compensation by roughly \$5,000. Similarly, as Figure 3d shows, before NCLB, about 10 percentage points (about 20%) fewer teachers possessed master’s degrees in states without consequential

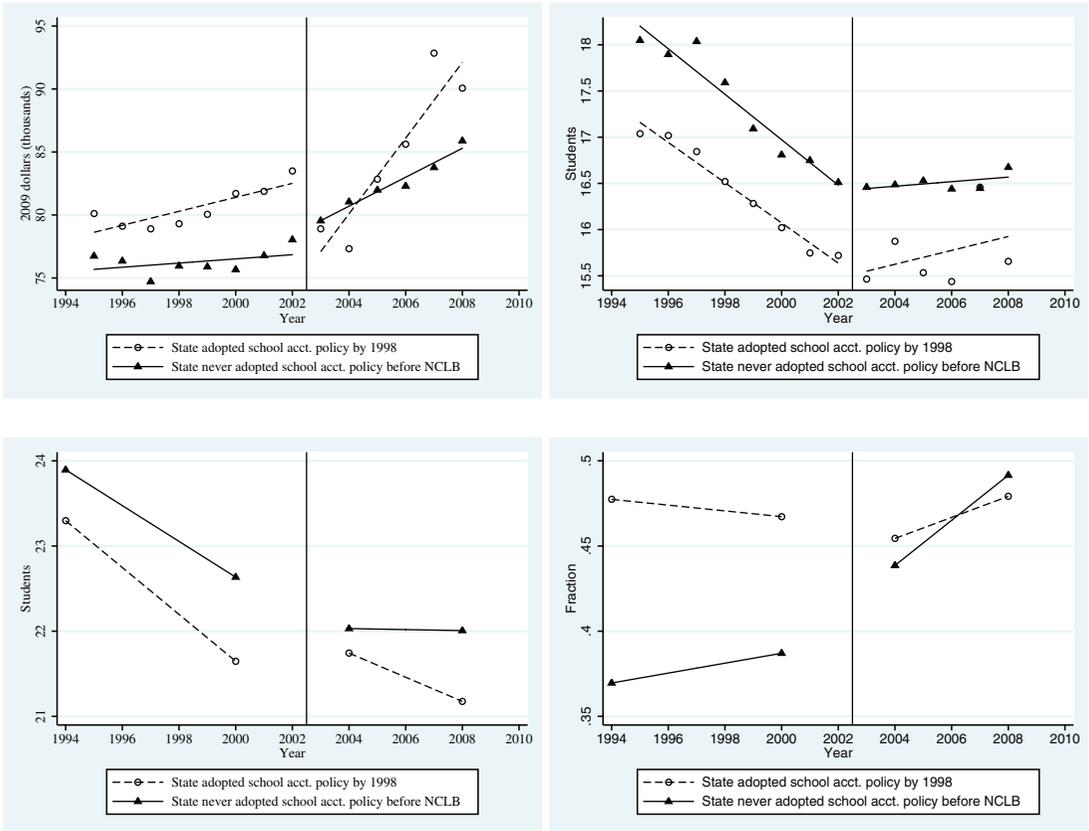


FIGURE 3. *Trends in school resources by timing of accountability.*

Note. Data for (a) and (b) drawn from the Common Core of Data. See note to Figure 1 for sample details. Data for (c) and (d) drawn from the Schools and Staffing Survey (SASS). SASS sample is composed of full-time elementary and middle school principals and full-time elementary and middle school teachers with a main assignment in either mathematics, English/language arts, or general elementary.

accountability than in states with consequential accountability. After the law took effect, the rates became similar across both groups of states. The corresponding regression results in Table 7 suggest that NCLB raised the percentage of teachers with master's degrees by approximately 6 percentage points. In contrast, although pupil-teacher ratios and class sizes have been falling, the trend is similar across states (Figures 3b and 3c). We are unable to rule out the hypothesis that neither pupil-teacher ratios nor average class sizes changed as a result of the introduction of NCLB (see the regression results in Tables 6 and 7).

Columns 2 through 5 of Table 6 report the estimated NCLB effects on the CCD measures (instructional salaries and pupil-teacher ratios) by poverty quartile. It is interesting that although

these results are somewhat similar across poverty quartiles, it does appear that the increases in teacher compensation attributable to NCLB were particularly large in high-poverty districts. And Panel B of Table 7 reports, the estimated NCLB impact among SASS respondents teaching in schools where the percentage of students on free or reduced-price lunches exceeds 50%. The one exception involves these higher-poverty schools. The estimates in Table 7 suggest that NCLB led to a particularly large increase in teacher qualifications in these schools (i.e., 16 percentage points).

Taken at face value, these results suggest that the increased instructional expenditures catalyzed by NCLB were allocated primarily to teacher compensation, both overall and to pay for more teachers with advanced degrees, rather

TABLE 6

The Estimated Effects of NCLB on District Resources

Panel A: Teacher total compensation (salary + benefits)—District-level analysis using Common Core of Data (CCD) data					
	Compensation (all)	Compensation (Poverty Quartile 1)	Compensation (Poverty Quartile 2)	Compensation (Poverty Quartile 3)	Compensation (Poverty Quartile 4)
	(1)	(2)	(3)	(4)	(5)
Effect by 2007–2008 school year	5.067* (2.888)	3.613 (3.114)	5.992 (4.287)	5.447** (2.307)	6.982** (3.474)
Mean for 1999–2000 school year	79.577	76.655	76.665	76.310	87.629
SD for 1999–2000 school year	20.338	18.236	18.173	17.498	23.720
N (district-years)	140,775	35,483	35,256	35,195	34,841
Panel B: Pupil–teacher ratios—District-level analysis using CCD data					
	Pupil–teacher ratio (all)	Pupil–teacher ratio (Poverty Quartile 1)	Pupil–teacher ratio (Poverty Quartile 2)	Pupil–teacher ratio (Poverty Quartile 3)	Pupil–teacher ratio (Poverty Quartile 4)
	(6)	(7)	(8)	(9)	(10)
Effect by 2007–2008 school year	–0.151 (0.491)	–0.508 (0.511)	–0.399 (0.888)	0.219 (0.451)	0.310 (0.507)
Mean for 1999–2000 school year	16.986	17.031	17.167	16.972	16.798
SD for 1999–2000 school year	2.692	2.440	2.823	2.832	2.740
N (district-years)	140,800	35,505	35,259	35,199	34,837

Note. Outcomes are calculated with district-level data from 1995 to 2008 from the Common Core of Data. The total NCLB effect by 2008 is relative to a state with school accountability starting in 1997. Covariates include quartic functions of the district's total student enrollment, the percentage of Black or Hispanic students, and the district poverty rate in 2000, as well as an interaction between the poverty rate and the percentage Black or Hispanic. Standard errors (in parentheses) are clustered at the state level.

* $p < .1$. ** $p < .05$.

than toward more teachers and smaller classes. Although the class size results from the SASS are relatively imprecise, the confidence intervals still suggest that class-size reductions alone cannot explain the achievement gains attributed to NCLB. We cannot reject the hypothesis that NCLB reduced class sizes by as much as 1.1 students. Given prior evidence that class-size reductions of roughly 7 students lead to an achievement gain of 0.20 standard deviations (Krueger, 2003), we would expect a class-size reduction of this size to raise achievement by no more than 0.031 standard deviations (i.e., $[1.1/7] \times 0.20$), a fraction of the Grade 4 math score gain attributed to NCLB (Dee & Jacob, 2011).

Use of Instructional Time

As noted in our literature review, there is descriptive evidence from several teacher and school surveys suggesting that NCLB has caused teachers to shift their instructional focus toward tested subjects. We are able to confirm many of these reports by using our CITS analysis strategy to examine teachers' reports of their time use across survey years. These data allow us to compare changes in teacher responses over time rather than relying on retrospective judgments on the part of teachers. The data also provide usefully objective measures of some of the constructs—for example, the time use

TABLE 7

The Estimated Effects of NCLB on Classroom Resources

Panel A: All teachers		
	Class size	Fraction of teachers with a master's degree
	(1)	(2)
CITS model		
Effect by 2007–2008 school year	–0.328 (0.500)	0.056** (0.028)
DD model		
Effect by 2007–2008 school year	–0.254 (0.245)	0.084*** (0.024)
Mean for 1999–2000 school year	22.120	0.412
SD for 1999–2000 school year	4.990	0.492
N (teacher-years)	27,454	36,100
Panel B: Teachers in schools with >50% free lunch		
	Class size	Fraction of teachers with a master's degree
	(3)	(4)
CITS model		
Effect by 2007–2008 school year	1.085 (0.898)	0.160*** (0.057)
DD model		
Effect by 2007–2008 school year	–0.067 (0.352)	0.082** (0.033)
Mean for 1999–2000 school year	22.075	0.371
SD for 1999–2000 school year	5.287	0.483
N (teacher-years)	11,367	14,513

Note. CITS = comparative interrupted time series; DD = difference in differences. The total NCLB effect by 2008 is relative to a state with school accountability starting in 1997. Covariates include 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 percentage minority and percentage free lunch. All specifications include state fixed effects. Standard errors (in parentheses) are clustered at the state level.

** $p < .05$. *** $p < .01$.

questions ask about the actual number of hours per week teachers devote to math, rather than asking teachers to characterize their emphasis on math as “big” or “small” or “larger/smaller” relative to a certain number of years ago.

Figure 4 illustrates some of these results by showing the unadjusted national trends in several measures for the sample of elementary school teachers in 1994, 2000, 2004, and 2008, separately for states that did and did not adopt school accountability programs prior to NCLB. Figure 4a shows the fraction of teachers in the school who are departmentalized—that is, they instruct several classes of different students in one or more subjects (rather than teaching the

same group of students all day in multiple subjects, which is referred to as a self-contained teacher). Figure 4b shows in hours per week the amount of instructional time that nondepartmentalized teachers report for core academic subjects. Figure 4c shows the fraction of time during the week that nondepartmentalized teachers spent teaching math and ELA where the denominator is the total time spent on the four core subjects (math, ELA, social studies, and science). Figure 4d shows this ratio specifically with respect to reading.

These figures suggest that NCLB did not lead to meaningful increases in departmentalized instruction or in the total amount of instructional

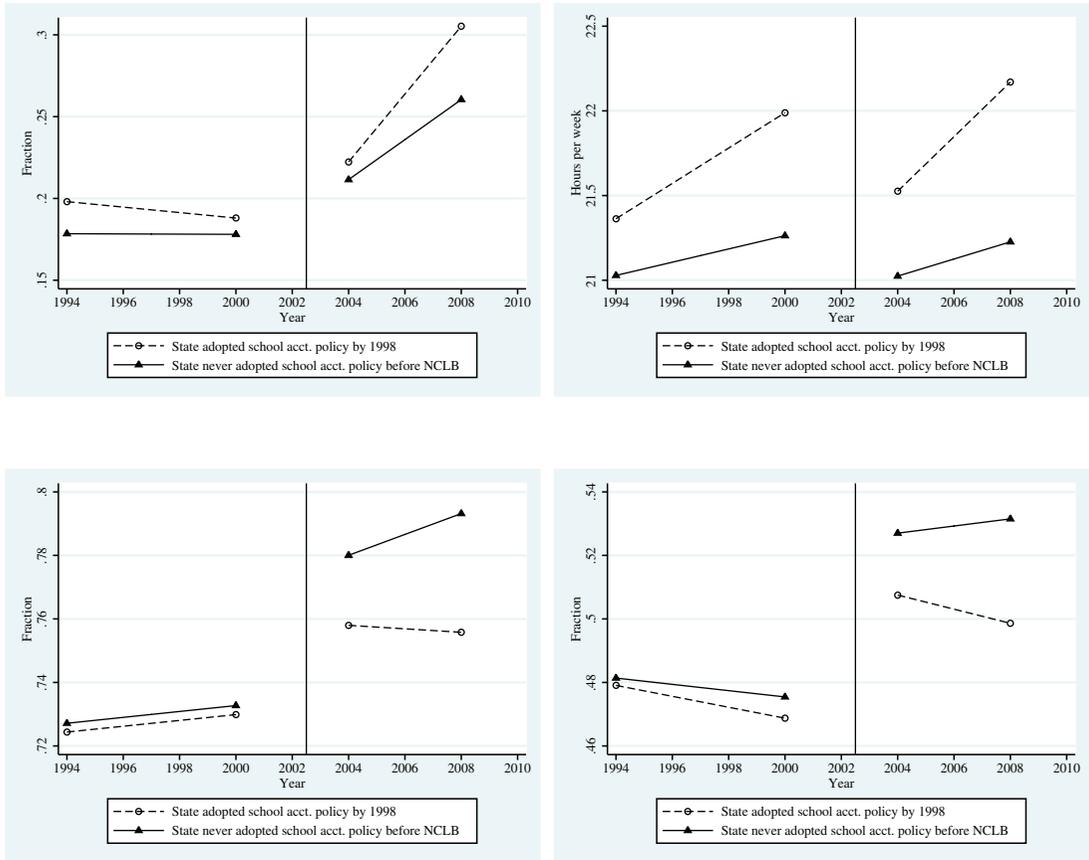


FIGURE 4. Trends in school time use by timing of accountability.
 Note. Data drawn from the Schools and Staffing Survey. See note to Figure 3 for sample details. Graphs (b) through (d) represent only self-contained and team teachers.

time for core subjects. However, Figure 4c shows that the share of instructional time allocated to math and ELA increased following the introduction of NCLB, particularly in states that had not instituted school accountability prior to this time. Similarly, we can see corresponding and similar-sized drops in the share of time allocated to science and social studies (figures not shown). More specifically, the regression results in Table 8 indicate that NCLB increased the share of time given to math and ELA by 3.6 percentage points (roughly 5%). The magnitude of the effect is even greater in schools where more than half of students were approved for free lunch, with an increase of 4.2 percentage points. To put this effect size in perspective using the structure of the original question from the SASS, this increase implies an additional 45 minutes per week of math/ELA instruction or 50 minutes per week in

high-poverty schools for teachers who spend 20 instructional hours on these two subjects.¹² It is interesting that the trends in Figure 4d (and the corresponding estimates in Table 8) suggest that the overall increase in math and ELA instruction combined was driven primarily by an increase in time devoted to ELA. The estimated effects of NCLB on the fraction of time devoted to math, though positive, are smaller and statistically insignificant. This heterogeneity is particularly interesting in light of the prior evidence that the achievement gains attributable to NCLB are concentrated in math, not reading (Dee & Jacob, 2011).

School Climate

Our final set of results corresponds to a series of variables that we have grouped under the

TABLE 8
The Estimated Effects of NCLB on the Use of Instructional Time

Panel A: All teachers							
	Fraction departmentalized	Total academic time per week	Math and ELA time/total academic time	Math time/total academic time	ELA time/total academic time	Science time/ total academic time	Social studies time/total academic time
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
CITS model							
Effect by 2007–2008 school year	0.002 (0.015)	–0.307 (0.684)	0.036*** (0.012)	0.013 (0.010)	0.023* (0.013)	–0.017* (0.009)	–0.019*** (0.005)
DD model							
Effect by 2007–2008 school year	–0.003 (0.017)	–0.490 (0.359)	0.030** (0.013)	0.000 (0.005)	0.029* (0.016)	–0.014* (0.008)	–0.016*** (0.006)
Mean for 1999–2000 school year	0.180	21.758	0.737	0.261	0.476	0.125	0.137
SD for 1999–2000 school year	0.384	6.445	0.130	0.118	0.156	0.089	0.089
N (teacher-years)	36,169	27,460	27,436	27,436	27,436	27,436	27,436
Panel B: Teachers in schools with >50% free lunch							
	Fraction departmentalized	Total academic time per week	Math and ELA time/total academic time	Math time/total academic time	ELA time/total academic time	Science time/ total academic time	Social studies time/total academic time
	(8)	(9)	(10)	(11)	(12)	(13)	(14)
CITS model							
Effect by 2007–2008 school year	–0.015 (0.037)	–1.128 (1.050)	0.042* (0.022)	0.005 (0.019)	0.037*** (0.014)	–0.007 (0.017)	–0.035*** (0.010)
DD model							
Effect by 2007–2008 school year	–0.027* (0.015)	–0.382 (0.428)	0.033 (0.022)	–0.007 (0.005)	0.040* (0.024)	–0.013 (0.013)	–0.020** (0.009)
Mean for 1999–2000 school year	0.155	21.733	0.747	0.265	0.482	0.120	0.133
SD for 1999–2000 school year	0.362	6.887	0.131	0.117	0.160	0.083	0.084
N (teacher-years)	14,543	11,299	11,288	11,288	11,288	11,288	11,288

Note. CITS = comparative interrupted time series; DD = difference in differences. The total NCLB effect by 2008 is relative to a state with school accountability starting in 1997. When specified, covariates include 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 percentage minority and percentage free lunch. All specifications include state fixed effects. Standard errors (in parentheses) are clustered at the state level.

* $p < .1$. ** $p < .05$. *** $p < .01$.

general heading of school climate outcomes. Unfortunately, the SASS has not routinely collected data on all the school and teacher practices that are of interest, so our capacity to isolate the effects of NCLB on some of these outcomes is limited. However, the SASS has collected consistent data on several relevant school-level traits. These outcomes help to answer the question of how teachers' and principals' perceptions of their schools have changed in response to NCLB. For example, the principals who responded to the SASS were asked to choose from a list of nine educational goals their top three priorities (see Appendix C). Figure 5a shows the comparative trend data for the share of principals who indicated that either academic excellence or basic skills was a top goal as opposed to the other options such as promoting personal growth, human relations, multicultural awareness, and so on. This measure of instructional focus did not follow a clear trend across treatment and control states after NCLB, suggesting that NCLB did not generate a detectable increase in instructional focus. This result is confirmed by the regression results in Table 9. However, it is worth noting that a very large fraction of principals in all years and all states indicated that these "academic" goals were their top priority. It is possible that NCLB affected the intensity of a principal's focus on this top goal, even if it did not raise the goal in the principal's priority ranking.

Teachers in the SASS answered questions about whether principals and fellow teachers enforced rules for student conduct. The trends in Figure 5b and regression estimates in column 2 of Table 9 indicate that NCLB did not have any noticeable impact on this measure. However, teachers in the SASS also answered a series of questions about their view of various student behaviors and attitudes. These items asked teachers to what extent they thought the following things were a problem in their school: student tardiness, student absenteeism, student class cutting, student dropping out, student apathy, and student unpreparedness to learn when coming to school. As outlined in Appendix C, we standardized and aggregated these measures into a composite measure of a trait we call student engagement, where positive values reflect greater student engagement.

An extensive literature in educational psychology characterizes measures like these as "student engagement" and argues that this broad multidimensional construct is an important mediating determinant of student achievement. More specifically, this literature defines student engagement as a "fusion of behavior, emotion, and cognition" that implies an active commitment to education (e.g., Fredricks, Blumenfeld, & Paris, 2004; Glanville & Wildhagen, 2007). This literature has identified at least two specific dimensions of student engagement, both of which are captured in the SASS measure used here. Actions related to classroom participation (e.g., attendance, tardiness, disruptiveness) are commonly characterized as behavioral engagement, whereas measures related to students' affective reactions to school (e.g., interest, motivation, sense of belonging) constitute psychological or emotional engagement.

Figure 5c shows for this measure the trend data, which suggest a noticeable improvement in this measure of student engagement for states that introduced school accountability because of NCLB. The regression results in Table 9 indicate that the effect size of this statistically significant increase is 0.22 in the full sample. In the sample of schools with greater than half of students approved for free lunch, the effect of NCLB on student engagement is substantially larger (i.e., effect size = 0.55). These results suggest that the introduction of consequential accountability catalyzed by NCLB led to improvements in school climates along fundamental, noncognitive dimensions that are clear antecedents to cognitive achievement (e.g., less absenteeism, tardiness, and apathy). In results available on request, we find that NCLB appeared to have a comparable effect on each of the items within this composite measure.

However, it should be noted we find that this result is less striking in conventional DD specifications, which do not condition on pretreatment trends unique to treatment status. More specifically, those specifications suggest that NCLB did not have statistically significant effects on student engagement. We cannot definitively establish whether the CITS or DD specifications generate more reliable point estimates in this context. However, the differences in pre-NCLB student-engagement trends across

TABLE 9

The Estimated Effects of NCLB on School Climate Outcomes

Panel A: All teachers			
	Fraction of principals with skills/excellence as #1 goal	Teachers' perceptions of school discipline	Teachers' perceptions of student engagement
	(1)	(2)	(3)
CITS model			
Effect by 2007–2008 school year	–0.003 (0.037)	–0.074 (0.115)	0.220*** (0.056)
DD model			
Effect by 2007–2008 school year	0.055 (0.036)	0.044 (0.032)	0.017 (0.020)
Mean for 1999–2000 school year	0.875	0.003	0.059
SD for 1999–2000 school year	0.331	0.989	0.990
N (teacher-years)		36,169	36,169
N (principal-years)	16,552		
Panel B: Teachers in schools with >50% free lunch			
	Fraction of principals with skills/excellence as #1 goal	Teachers' perceptions of school discipline	Teachers' perceptions of student engagement
	(4)	(5)	(6)
CITS model			
Effect by 2007–2008 school year	0.014 (0.066)	–0.003 (0.105)	0.552*** (0.111)
DD model			
Effect by 2007–2008 school year	0.026 (0.044)	0.094* (0.047)	0.018 (0.018)
Mean for 1999–2000 school year	0.905	–0.048	–0.069
SD for 1999–2000 school year	0.293	0.988	0.934
N (teacher-years)		14,543	14,543
N (principal-years)	6,474		

Note. CITS = comparative interrupted time series; DD = difference in differences. The total NCLB effect by 2008 is relative to a state with school accountability starting in 1997. Covariates include 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 percentage minority and percentage free lunch. All specifications include state fixed effects. Standard errors are clustered at the state level.

* $p < .1$. *** $p < .01$.

treatment and comparison states (i.e., Figure 5c) are consistent with the motivation for preferring the CITS approach to the DD specification.

Conclusions

NCLB marked a dramatic expansion of federal involvement in public elementary and secondary by effectively compelling all states to introduce test-based school accountability sys-

tems. Our results indicate that NCLB led to district-level increases in school spending of nearly \$600 per pupil, which were funded by increases in state and local (as opposed to federal) revenue. Although we do not find evidence that the increased spending catalyzed by NCLB led to smaller class sizes, we find that the legislation increased both teacher compensation and the share of elementary school teachers with advanced degrees.

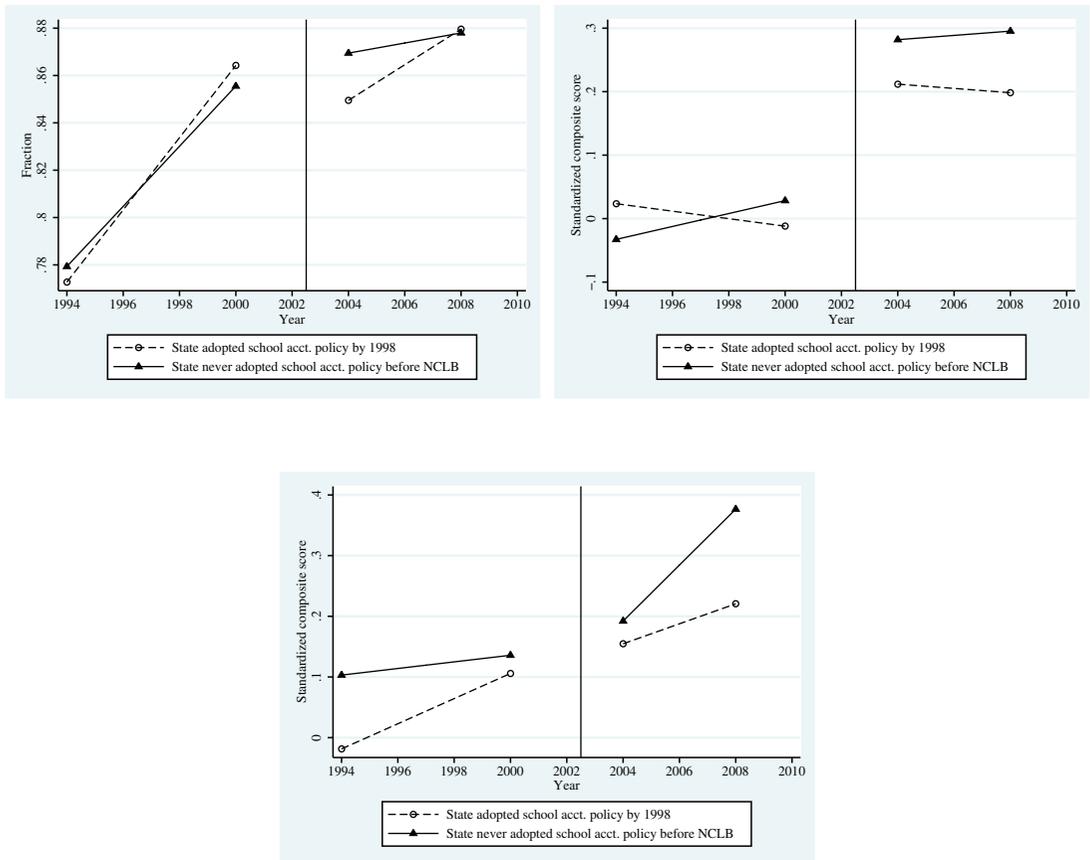


FIGURE 5. Trends in school culture outcomes by timing of accountability.

Note. Data drawn from the Schools and Staffing Survey. See note to Figure 3 for sample details.

Finally, our results suggest that NCLB influenced some school and classroom practices. We find that NCLB led schools to reallocate time away from science and social studies and toward math and reading. However, these effects were fairly modest and concentrated on reading (where NCLB did not have detectable achievement effects) rather than on math (where NCLB appears to have improved performance). The legislation also appears to have led to a striking improvement in teacher-reported measures of the behavioral engagement of students (e.g., reducing tardiness, absenteeism, and apathy). Together, these results suggest that the achievement effects generated by NCLB may have been related to the increased resources directed to

both teachers and pupil support services and related to changes in school culture that promoted basic dimensions of student engagement (e.g., attendance, punctuality, interest).

Given the increased expenditures, it is reasonable to ask whether NCLB would pass a simple cost-benefit test. Based on prior estimates that a 1-standard-deviation increase in elementary math scores is associated with an 8% increase in adult earnings (Krueger, 2003), the 0.23-standard-deviation impact of NCLB (Dee & Jacob, 2011) would translate into an earnings boost of 1.8%. Assuming a 3% discount rate, the present discounted value as of age 9 of a 1.8% increase in subsequent earnings beginning at age 18 is at least \$13,300.¹³ This

calculation implies that the test score gains attributable to NCLB are quite large relative to the corresponding expenditure increases of \$600 per student-year, even if we assume that the spending increases resulting from NCLB are sustained for eight elementary school years.

However, the many caveats associated with this “back-of-the-envelope” calculation should be noted. In particular, this exercise ignores both socially relevant benefits (e.g., the positive externalities of human-capital improvements) and costs (e.g., the deadweight losses associated with raising revenues). Also, this calculation should not be understood to suggest that increased spending was necessarily the relevant mediating mechanism behind NCLB’s achievement effects. Nonetheless, this calculation provides suggestive evidence that the achievement and expenditure effects of NCLB could easily pass a cost–benefit test.

Notes

1. 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.

2. A 2009 Government Accountability Office (GAO) study based on teacher survey data (and supplemental interviews with state officials) finds that 90% of elementary teachers reported no change in instructional time for arts education between 2004–2005 and 2006–2007 (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 U.S. Department of Education’s National Longitudinal Study of No Child Left Behind.

3. We considered using Catholic schools as an alternative control group. However, by checking how the observed traits of Catholic schools changed before and after NCLB’s introduction, we uncovered a striking empirical pattern. Catholic-school enrollment exhibited a distinct negative trend break for academic year 2002–2003 (NCLB’s first year) and a sharper downward trend after that. Further investigation suggests that this enrollment drop was a result of intense publicity surrounding the sex abuse scandal in Boston and then elsewhere nationwide. Because this scandal generated potentially nonrandom out-migration from

Catholic schools that happened to be contemporaneous with NCLB, we think this is unlikely to be a genuinely credible basis for identification.

4. However, there are also a few notable distinctions between our classification of consequential accountability states (Table A1) and the coding reported by Hanushek and Raymond (2005). For more details on these discrepancies, see the appendices in Dee and Jacob (2011).

5. Appendix A discusses dating the start of NCLB in more detail and the results of sensitivity analyses based on assuming that NCLB began in earlier or later years.

6. There are several other ways that one might think about the question of treatment intensity; however, we have not been able to find an alternative that is both conceptually compelling and feasible from a data perspective. In the absence of a satisfying alternative measure of intensity, we can demonstrate that our results are not dependent on our primary measure of years without consequential accountability by using only the binary treatment described in the previous paragraph. Findings (available on request) remain qualitatively the same.

7. The one exception is that we no longer see a significant increase in the percentage of teachers with a master’s degree in our main sample when we condition on household income, although we still see the effect in our high-poverty sample.

8. In our data, both types of schools were defined according to National Center for Education Statistics convention; elementary schools are designated as those with at least one grade lower than four and no grades higher than eight, and middle schools are designated as those with grades no lower than five and no grades higher than eight.

9. We also dropped from our data set all teachers and principals who worked within schools where more than 80% of students were classified with physical or mental disabilities.

10. The figures omit states that adopted school accountability programs between 1999 and 2001 because the impacts of these state programs might be confounded with the introduction of NCLB in 2002. In the regression estimates, we measure the intensity of the treatment in terms of the years without prior state accountability (see the discussion following Equation 1) to allow us to include these “late-adopting” states.

11. The total effect we report here is driven by both a level shift immediately following the introduction of NCLB as well as a slight increase in the time trend following NCLB (see Table B1).

12. If a teacher’s math/ELA time rises 3.6% from a base of 20 hours per week, this will add 0.72 hours

per week or 43 minutes per week of extra math/ELA instructional time.

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

14. Only about 20 states have National Assessment of Educational Progress (NAEP) equivalence measures prior to NCLB for a combination of reasons, including that (a) many states did not administer state-representative NAEP prior to 2003, (b) many states did not report proficiency levels as part of their state testing regime, (c) many states did not test the two NAEP grades (i.e., Grades 4 and 8), and (d) the authors of the report calculated equivalence measures for only a subset of states with available data prior to 2003 because these early years were viewed as a “trial run” for developing the equating procedures. Even in the years 2003, 2005, and 2007, NAEP equivalence measures are available for only a limited set of states because (a) not all states tested fourth and eighth graders and (b) a handful of states did not have sufficient NAEP data in a given grade, year, or subject to justify the equivalence exercise.

15. For example, our point estimates suggest that states without prior accountability lowered proficiency standards in fourth grade math by 8 to 10 NAEP scale points between 2003 and 2007, although these estimates are very imprecise and not statistically different from zero.

References

- Ballou, D., & Springer, M. G. (2008). *Achievement trade-offs and No Child Left Behind*. Working paper.
- Bandeira de Mello, V., Blankenship, C., & McLaughlin, D. *Mapping state proficiency standards onto NAEP scales: 2005-2007*. Retrieved October 30, 2012, from <http://nces.ed.gov/nationreportcard/pubs/studies/2010456.asp>
- Bifulco, R., Jr. (2010). The influence of performance-based accountability on the distribution of teacher salary increases. *Education Finance and Policy*, 5, 177–199.
- Bloom, H. S. (1999). *Estimating program impacts on student achievement using short interrupted time series*. Working paper, Manpower Demonstration Research Corporation, New York, NY.
- Bloom, H. S., Ham, S., Melton, L., & O'Brien, J. (2001). *Evaluating the accelerated schools approach a look at early implementation and impacts on student achievement in eight elementary schools*. New York, NY: Manpower Demonstration Research Corporation.
- Braun, H., & Quian, J. (2008) *Mapping state standards to the NAEP scale*. ETS RR-08-57. Educational Testing Service.
- Carnoy, M., & Loeb, S. (2002). Does external accountability affect student outcomes? A cross-state analysis. *Educational Evaluation and Policy Analysis*, 24(4), 305–331.
- Center on Education Policy. (2006). *From the capital to the classroom: Year 4 of the No Child Left Behind Act*. Washington, DC: Diane Stark Rentner, Caitlin Scott, Nancy Kober, Naomi Chudowsky, Victor Chudowsky, Scott Joftus, Dalia Zabala.
- Center on Education Policy. (2007). *Choices, changes and challenges: Curriculum and Instruction in the NCLB era. From the capital to the classroom: Year 5 of the No Child Left Behind Act*. Washington, DC: Jennifer McMurrer.
- Center on Education Policy. (2008). *Instructional time in elementary schools: A closer look at changes for specific subjects. From the capital to the classroom: Year 5 of the No Child Left Behind Act*. Washington, DC: Jennifer McMurrer.
- Dee, T. S. & Jacob, B. A. (2011). The impact of the No Child Left Behind Act on student achievement. *Journal of Policy Analysis and Management*, 30(3), 418–446.
- Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. 2004. School engagement: Potential of the concept, state of the evidence. *Review of Educational Research*, 74(1), 59–109.
- Glanville, J. L., & Wildhagen, T. (2007). The measurement of school engagement. *Educational and Psychological Measurement*, 67(6), 1019–1041.
- Goertz, M. E., & Duffy, M. E. (2001). *Assessment and accountability systems in the 50 states: 1999–2000* (CPRE Research Report RR-046). Philadelphia, PA: Consortium for Policy Research in Education.
- Government Accountability Office. (2009). *Access to arts education: Report to congressional requesters* (GAO-09-286). Washington, DC: Author.
- Hamilton, L. S., Stecher, B. M., Marsh, J. A., McCombs, J. S., Robyn, A., Russell, J. L., . . . Barney, H. (2007). *Standards-based accountability under No Child Left Behind: Experiences of teachers and administrators in three states*. Santa Monica, CA: RAND.
- Hannaway, J., McKay, S., & Nakib, Y. (2002). Reform and resource allocation: National trends and state policies. In W. Fowler (Ed.), *Developments in school finance 1999–2000* (pp. 57–76). Washington,

- DC: U.S. Department of Education, National Center for Education Statistics.
- Hannaway, J., & Stanislawski, M. (2005). *Responding to reform: Florida school expenditures in the 1990s*. Unpublished manuscript.
- Hanushek, E. A., & Raymond, M. E. (2005). Does school accountability lead to improved student performance? *Journal of Policy Analysis and Management*, 24(2), 297–327.
- Hess, F. M., & Petrilli, M. J. (2006). *No child left behind*. New York, NY: Peter Lang Publishing.
- Hoff, D. J., & Walsh, M. (2008). Sparring continues over NCLB legal ruling. *Education Week*, 27(21), 19.
- Ingersoll, R. M. (1999). The problem of underqualified teachers in American secondary schools. *Educational Researcher*, 28(2), 26–37.
- Ingersoll, R. M. (2006). *Who controls teachers' work? Power and accountability in America's schools*. Cambridge, MA: Harvard University Press.
- Jacob, B. (2005). Accountability, incentives and behavior: Evidence from school reform in Chicago. *Journal of Public Economics*, 89(5-6), 761–796.
- Jacob, B., & Levitt, S. (2003). Rotten apples: An investigation of the prevalence and predictors of teacher cheating. *Quarterly Journal of Economics*, 118(3), 843–877.
- Koretz, D. M., Barron, S. I., Mitchell, K. J., & Stecher, B. M. (1996). *The perceived effects of the Kentucky Instructional Results Information System (KIRIS) (MR-792-PCT/FF)*. Santa Monica, CA: RAND.
- Koretz, D., & Hamilton, L. S. (2006). Testing for accountability in K–12. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 531–578). Westport, CT: American Council on Education/Praeger.
- Koretz, D. M., Mitchell, K. J., Barron, S. I., & Keith, S. (1996). *The perceived effects of the Maryland School Performance Assessment Program (CSE Technical Report 409)*. Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing.
- Krieg, J. M. (2008). Are students left behind? The distributional effects of the No Child Left Behind Act. *Education Finance and Policy*, 3(2), 250–281.
- Krueger, A. B. (2003). Economic considerations and class size. *Economic Journal*, 113, F34–F63.
- Ladd, H. F., & Zelli, A. (2002). School-based accountability in North Carolina: The responses of school principals. *Education Administration Quarterly*, 38(4), 494–529.
- Lee, J., & Wong, K. K. (2004). The impact of accountability on racial and socioeconomic equity: Considering both school resources and achievement outcomes. *American Educational Research Journal*, 41(4), 797–832.
- Liu, X. (2007). The effect of teacher influence at school on first-year teacher attrition: A multilevel analysis of the Schools and Staffing Survey for 1999–2000. *Educational Research and Evaluation*, 13(1), 1–16.
- Murnane, R. J., & Papay, J. P. (2010). Teachers' views on No Child Left Behind: Support for the principles, concerns about the practices. *Journal of Economic Perspectives*, 24(3), 151–166.
- Murray, S. E., Evans, W. N., & Schwab, R. M. (1998). Education-finance reform and the distribution of education resources. *The American Economic Review*, 88(4), 789–812.
- National Center for Education Statistics. (2007). Mapping 2005 state proficiency standards onto the NAEP Scales. NCEP 2007-482. Washington, DC: U.S. Government Printing Office.
- National Research Council. (2011). *Incentives and test-based accountability in public education* (Committee on Incentives and Test-Based Accountability in Public Education, M. Hout, & S. W. Elliott, Eds.). Washington, DC: National Academies Press, Board on Testing and Assessment, Division of Behavioral and Social Sciences and Education.
- Neal, D., & Schanzenbach, D. W. (2010). Left behind by design: Proficiency counts and test-based accountability. *Review of Economics and Statistics*, 92(2), 263–283.
- Olson, L. (2002). States strive toward ESEA compliance. *Education Week*, December 1. Retrieved October 30, 2012, from <http://www.edweek.org/ew/articles/2002/12/11/15esea.h22.html>
- Olson, L. (2003). In ESEA wake, school data flowing forth. *Education Week*, December 10. Retrieved October 30, 2012, from <http://www.edweek.org/ew/articles/2003/12/10/15nclb.h23.html>
- Palmer, S. R., & Coleman, A. L. (2003). The No Child Left Behind Act: Summary of NCLB requirements and deadlines for state action. Council of Chief State School Officers. Retrieved November 13, 2009, from www.ccsso.org/content/pdfs/Deadlines.pdf
- Pedulla, J. J., Abrams, L. M., Madaus, G. F., Russell, M. K., Ramos, M. A., & Miao, J. (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.
- Reback, R., Rockoff, J. E., & Schwartz, H. L. (2010). *The effects of No Child Left Behind on school services and student outcomes*. Working paper.

- Rewarding results, punishing failure. (1999). *Education Week*. Retrieved October 30, 2012, from <http://www.edcounts.org/archive/sreports/qc99/>
- Rouse, C. E., Hannaway, J., Goldhaber, D., & Figlio, D. (2007). *Feeling the Florida heat? How low-performing schools respond to voucher and accountability pressure* (CALDER Working Paper 13). Cambridge, MA: National Bureau of Economic Research.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton-Mifflin.
- Smith, M. S., & O'Day, J. (1991). Systemic school reform. In S. H. Fuhrman, & B. Malen (Eds.), *The politics of curriculum and testing* (*Politics of Education Association Yearbook*, 1990) (pp. 233-267). London, UK: Taylor & Francis.
- Stecher, B. M., Barron, S. I., Kaganoff, T., & Goodwin, J. (1998). *The effects of standards-based assessment on classroom practices: Results of the 1996-97 RAND survey of Kentucky teachers of mathematics and writing* (CSE Technical Report 482). Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing, RAND Education.
- Swanson, C. B., & Stevenson, D. L. (2002). Standards-based reform in practice: Evidence on state policy and classroom instruction from the NAEP state assessments. *Educational Evaluation and Policy Analysis*, 24(1), 1-27.
- Taylor, G., Shepard, L., Kinner, F., & Rosenthal, J. (2003). *A survey of teachers' perspectives on high-stakes testing in Colorado: What gets taught, what gets lost* (CSE Technical Report 588). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Authors

THOMAS S. DEE is a Professor of Education at Stanford University, 520 Galvez Mall, Stanford, CA 94305, and a Research Associate of the National Bureau of Economic Research; tdee@stanford.edu. His research interests include education policy and the economics of education with an emphasis on at-risk students and the use of quasi-experimental and experimental research designs.

BRIAN JACOB is the Walter H. Annenberg Professor of Education Policy, Professor of Economics, and Director of the Center on Local, State and Urban Policy (CLOSUP) at the Gerald R. Ford School of Public Policy, University of Michigan, 5208 Weill Hall, 735 S. State Street, Ann Arbor, MI 48109; bajaboc@umich.edu. His primary fields of interest are labor economics, program evaluation, and the economics of education. His current research focuses on urban school reform and teacher labor markets.

NATHANIEL L. SCHWARTZ is the director of research and policy for the Tennessee Department of Education, 710 James Robertson Parkway, Nashville, TN 37243; nate.schwartz@tn.gov. His research focuses on instructional reform and teacher and school accountability.

Manuscript received July 18, 2011

Revision received June 22, 2012

Accepted October 15, 2012