How Did the Statewide Assessment and Accountability Policies of the 1990s Affect Instructional Quality in Elementary Schools?

Meredith Phillips
Jennifer Flashman

CCPR-017-07

March 2007

California Center for Population Research
On-Line Working Paper Series

How Did the Statewide Assessment and Accountability Policies of the 1990s Affect Instructional Quality in Elementary Schools?

This draft: March 2007

Meredith Phillips
and
Jennifer Flashman

UCLA

Draft. Comments welcome.

We are grateful to Jessica Schraub Norman and Ananya Sengutpa for their help assembling the accountability data; Bea Birman, Tiffani Chin, Adam Gamoran, Gloria Ladson-Billings, and Andrew Porter for their comments on this paper; and the UCLA Council on Research and, especially, the Russell Sage Foundation for their grants in support of our work with the Schools and Staffing Survey.

Introduction

In 2005, a mere 16% of low-income fourth-graders could read proficiently (as measured by national tests), compared with 42% of their middle class counterparts (Perie, Grigg, and Donahue 2005).[1] The math gap was even larger.[2] The *No Child Left Behind Act of 2001* (NCLB) represents an ambitious attempt to improve this situation. Although critics of NCLB complain about vagaries in the law, its under-funding, and the leeway given to states to choose their own standards and tests (Asimov 2005; Dobbs 2005; Rossi, 2006; Schaffer 2005), few can disagree with its ultimate purpose: "To ensure that all children have a fair, equal, and significant opportunity to obtain a high-quality education and reach, at a minimum, proficiency on challenging state academic achievement standards and state academic assessments" (Public Law 107–110, sec. 1001).

NCLB aims to accomplish its goals by holding public schools accountable for their students' achievement. Under NCLB, public schools must test all students annually in math and reading in grades 3 though 8. Schools face corrective action or complete restructuring if their students fail to meet state standards for academic progress several years in a row. NCLB also makes it easier for students in low-performing schools to attend better schools or to obtain free tutoring.

Yet the law wisely gave states several years to comply with its most critical requirements. Not until the 2005-2006 school year, for example, did states need to have a "highly qualified"

[1] For the purposes of this paper, we define low-income students as those who qualify for subsidized school meals. Students qualify for either free or reduced-price meals if their family income falls below 185 percent of the poverty line (U.S. Department of Agriculture 2002).

[2] In 2005, 19% of low-income fourth-graders scored at or above proficient in math, compared with 49% of middle class fourth-graders (Perie, Grigg, and Dion 2005).

teacher in every classroom and reading and math testing in grades 3 through 8. Consequently, it is still too early to assess the effects of NCLB on the nation's educational system.

Fortunately, however, the assessment and accountability provisions of NCLB build directly on assessment and accountability policies begun a decade earlier. Spurred in part by the *Improving America's Schools Act of 1994* (Public Law 103–382), many states increased the number of tests they required students to take and some states began imposing consequences on schools or school districts when students performed poorly on those tests. Examining the effects of these state-wide reforms of the 1990s can help us understand the likely effects of NCLB and generate ideas about how to improve NCLB when it comes up for reauthorization. Prior work has examined the impact of state accountability policies on student achievement (e.g., Carnoy and Loeb 2002; Hanushek and Raymond 2005). This paper focuses instead on how such policies influenced teachers and school administrators—especially those who work in low-income schools.

Possible Consequences of Accountability Policies

Proponents of accountability reform suspect that teachers and administrators have not been doing as much as they could to improve academic achievement, especially among traditionally low-performing students. Their solution is to attach positive and negative sanctions to students' test performance, hoping that these incentives will stimulate teachers and administrators to focus more assiduously on increasing academic achievement and equity (see, e.g., Gordon 2005; Mathews 2006). If this solution works, we should find empirical evidence that accountability reforms motivate teachers or administrators to focus additional instructional resources on improving academic achievement. For example, accountability policies may motivate teachers to work harder to ensure that children master important concepts and skills. Or

administrators may respond to accountability pressures by devoting more resources to professional development that helps teachers improve their instructional practices or by substituting instructional expenditures for non-instructional expenditures to reduce class sizes or improve teacher quality.

Critics of accountability reform worry, however, that such policies will have unintended, negative consequences. In particular, critics contend that accountability reforms worsen teacher quality by creating working conditions that good teachers cannot tolerate (see Chenfeld 2006; Tobin and Ave 2006). Qualitative research indicates that mandated tests may reduce teachers' autonomy over their teaching strategies by forcing them to cover skills at different paces, at different times, or in a different order than they ideally would (Chin 2002). Such studies also suggest that because accountability policies dictate the skills that students must learn, they force creative teachers to abandon innovative teaching strategies and materials, causing the best teachers to become demoralized and bored (see, e.g., McNeil 2000). These studies imply that accountability reforms may compel the best teachers to leave the profession or may dissuade potentially good teachers from considering a career as a public school teacher.

Critics also worry that test-based accountability policies will encourage teachers and administrators to focus too much on improving tested skills to the detriment of other types of skills that the tests do not measure (Gallagher 2004; Marshak 2003). A number of studies suggest that mandatory assessments cause teachers to shift their instructional time toward tested subjects and skills and away from untested subjects and skills (Hamilton, Stecher, and Klein 2002; Jacob 2004; Koretz 2002). The effects of this shift may be positive if the tested subjects are more important than the untested subjects or if the tested subjects provide an important

foundation for learning the untested subjects. But the effects of the shift may be negative if narrow tests lead teachers to neglect important skills that lie beyond the scope of the tests.

Skeptics suspect that both the proponents and critics of accountability reform may assume too much about the power of top-down incentives to change public education (Wilms 2003). A weak version of this position holds that administrators may simply reallocate resources in an attempt to improve test scores--for example, reducing class sizes in tested grades or assigning high-quality teachers to tested grades. Although skeptics might argue that reallocating resources among grades will merely help raise test scores in particular grades, the effects of reallocation could be positive if the tested grades were also the grades that were especially critical for mastering skills that are prerequisites for further learning. A stronger version of the skeptical stance holds that accountability reforms will have hardly any effect on teachers' and children's day-to-day experiences, perhaps because teachers and administrators assume that current reforms, like past reforms, will soon be replaced by yet another set of "new" reforms (Tyack and Cuban 1995), or perhaps simply because of bureaucratic inertia.

Accountability and Equity

Proponents, critics, and skeptics agree, however, that if accountability reforms influence educational practice, they will probably have their largest effects on schools that serve disadvantaged students. Because poor, African American, and Latino students score lower on tests, on average, than do their middle class, white, and Asian American counterparts, and because schools tend to be segregated by ethnicity and social class, both the positive and negative incentives created by test-based accountability policies should affect predominantly poor, minority schools the most. It is impossible to know *a priori*, however, whether the overall

consequences for such schools will be good or bad (see the excellent discussion in Clotfelter, Ladd, Vigdor, and Aliaga 2004).

Relevant Literature

Previous research on the effects of pre-NCLB accountability reform on student achievement has generally found that stronger accountability policies are associated with improved test scores. For example, using cross-sectional data on accountability systems in 2000, Carnoy and Loeb (2002) found that fourth and eighth graders in states with strong accountability systems gained more in math than did their counterparts in states with weak or non-existent accountability systems. Likewise, using panel data on state accountability systems during the 1990s, Hanushek and Raymond (2005) found that students gained more in math and reading between fourth and eighth grade when they lived in states that adopted accountability systems. And, using data on school districts rather than states, Jacob (2004) found that a policy that threatened to retain students and put schools on probation (with the additional threat of reconstitution if they did not improve) led to improved reading and math scores among third, fifth, and eighth graders in the Chicago Public Schools, at least on the test that the district used to judge academic skills.

These studies and others have also examined a range of likely explanations for these effects on student achievement—from increased special education placements or preemptive retention to changes in teachers' instruction practices or teacher quality. The evidence on special education placements or retention is mixed, with some studies showing effects of accountability reforms (e.g., Jacob 2004, Deere and Strayer 2001; Cullen and Reback 2002; Figlio and Getzler 2002) and others finding no effects (e.g., Koretz and Barron 1998; Hanushek and Raymond 2005). A number of studies have concluded, however, that teachers respond strategically to

accountability pressures by focusing their instruction on tested material (for reviews, see Hamilton, Stecher, and Klein 2002 and Koretz 2002; also see suggestive evidence based on test score gains in tested and non-tested subjects in Jacob 2004 and Deere and Strayer 2001).

A final set of studies has examined whether accountability policies influence class size, teacher quality, or teachers' attitudes, especially in schools that serve disadvantaged children. Two studies, based on data from two different states, yield somewhat different answers about the impact of accountability on teacher quality. Clotfelter, Ladd, Vigdor, and Aliaga (2004) found that the introduction of an accountability system in North Carolina (and/or the official labeling of schools as "low-performing") increased teacher turnover in low-performing schools but did not cause a statistically significant decline in the proportion of high-quality teachers (as measured by experience and college quality) in low-performing schools. In contrast, Boyd, Lankford, Loeb, and Wyckoff (2005) found that the introduction of mandatory testing in fourth grade in New York, combined with school-level consequences for poor test performance, *reduced* the relative turnover rate of fourth-grade teachers. In addition, teachers who entered fourth grade after the reform tended to be *more* experienced and, in the lowest-performing schools, were more likely to have attended highly competitive colleges.

Two other studies used the same national data source that we use in this paper, the Schools and Staffing Survey (SASS), to investigate the relationship between accountability policies and changes in teachers' attitudes and quality between 1994 and 2000. Lee and Wong (2004) found that differences among states in their accountability policies during the 1990s were not associated with changes in class sizes or in-field teaching. Similarly, Loeb and Estrada (2005) found that differences among states in accountability policies in 2000 were not associated with changes in teachers' academic skills (as measured by the competitiveness of their

undergraduate institutions) or with changes in teachers' reports about their job satisfaction or control over the curriculum. And neither of these studies found that the effects of accountability differed between poor and non-poor schools.

Our paper builds on this past research. Like Lee and Wong (2004), we examine the effects of state accountability policies on class size and, like Loeb and Estrada (2005), we examine the effects of accountability on teachers' academic skills (as measured by the competitiveness of teachers' colleges) and their perceptions of autonomy. In addition, we estimate the effects of accountability policies on the time teachers devoted to school-related activities outside normal school hours, teachers' exposure to several types of professional development, and teacher turnover, experience, credentials, and educational attainment. Our methodology also differs from previous work because we estimate the effects of *changes* in state accountability policies (rather than levels of state accountability policies) on various outcomes. This strategy allows us to control for any unchanging aspects of states that might be associated with differences among states in both their accountability policies and their teachers.

In contrast to Clotfelter et al. (2004) and Boyd et al. (2005), we draw on evidence from all states rather than just one. This multi-state approach has one large disadvantage: we have less detailed, and more errorful, data both on the accountability policies in each state as well as on teachers' career paths. But our approach also has several advantages. First, because different states implemented different accountability policies over this time period, our methodology allows us to compare the effects of different types of accountability policies (e.g., those that mandated tests, those that imposed consequences on students, those that imposed consequences on teachers, and those that imposed consequences on schools). Second, to the extent that other state education policies, such as teacher certification testing or class size reduction, changed over

the same time period, our multi-state analyses can better distinguish the concomitant effects of these policies from the effects of accountability policies. Third, and perhaps most important, our work provides a more general test of how accountability policies may have influenced schools and teachers in states across the nation. Thus, our results may help policymakers predict how educators nationwide will respond to NCLB.

Methods

Data

We constructed a dataset by combining data on state accountability policies from the State Student Assessments Annual Survey (Council of Chief State School Officers 2001; Council of Chief State School Officers/North Central Regional Education Laboratory 1995) for the 1993-1994 and 1999-2000 school years (hereafter CCSSO) with data reported by teachers in the Schools and Staffing Survey (SASS) for the same school years. The CCSSO survey, which was filled out by the state directors of assessment, provides the best data for our purposes both because it contains detailed questions about many different types of assessment and accountability and because it provides reasonably comparable data throughout the 1990s. The SASS data provide nationally representative data on teachers' characteristics, attitudes, instructional practices, class sizes, and exposure to professional development.

In contrast to previous studies that have used the SASS data (e.g., Lee and Wong 2004 and Loeb and Estrada 2005), we restricted our sample to the teachers for whom we expected accountability policies to have the largest effects: Full-time public school teachers who taught in regular schools or in schools with a "special program emphasis" (such as charter or magnet schools) and who taught English, math, science, history, or general elementary subjects. We excluded part-time teachers, teachers who taught in special education, vocational-technical, or

alternative schools, and teachers who taught non-academic subjects such as art, music, or physical education.[3] We also restricted our sample to teachers who taught in grades 1 through 6. We decided to focus on elementary school teachers both because we suspected that accountability pressures would be greatest for these teachers and because the majority of elementary school teachers teach multiple subjects, and we wanted to investigate whether accountability policies affected the percentage of time teachers allocated to various subjects.[4]

Defining "Accountability"

States implemented various types of school accountability policies over the 1990s. All of these policies involved mandatory testing in some subjects and grades. In conjunction with testing, some states implemented "student accountability" policies that threatened to prevent students from being promoted to the next grade if they did not meet certain performance standards on the tests. A few states developed "teacher accountability" policies that linked rewards and punishments for teachers to students' test performance. Other states implemented "school accountability" policies that ranged from simply reporting schools' test performance to school dissolution and takeover.

Most previous studies of state accountability policies have tried to capture this policy variation in their accountability measures, but no consensus has developed about the best way to measure "accountability." Carnoy and Loeb (2002) and Loeb and Estrada (2005) used 1999-2000 data from the Consortium for Policy Research in Education (CPRE) report (Goertz and Duffy 2001) to create a 0-5 scale of accountability strength based on information about the

[3] Past studies have not restricted their samples to teachers of core subjects in regular schools (Lee and Wong 2004, Loeb and Estrada 2004). Of the public school teachers in the pooled SASS sample, over 40% listed a non-academic subject as their main assignment field and another 10% taught in special education, vocational-technical, or alternative schools.

[4] Note, however, that we do not restrict our sample to elementary schools per se. Instead, we control for whether the teacher taught in an elementary school in our regressions. Consequently, we can generalize our results to all full-time teachers of regular subjects who teach in grades 1 through 6, regardless of whether they teach in schools that also contain middle or high school grades.

extent of mandatory testing, the seriousness of consequences for schools, and the requirement of exit exams for high school students. Schiller and Muller (2003), who examined the effects of state accountability policies on students' mathematics coursework, used several separate measures that they constructed from questions asked on the 1993 National Longitudinal Study of Schools survey. Their measures included the number of mandated tests in math, reading, science, and history/social studies in high school, the consequences of testing for students (including retention, remediation, and high school graduation), and the number of positive and negative consequences for schools (including financial incentives, official recognition, accreditation, waivers from testing/reporting/other regulations, negative publicity, or loss of control to a higher educational authority).

Lee and Wong (2004) combined data from three different policy surveys administered during the 1990s to characterize the overall strength of accountability policy in a state during the decade. Their measure included testing, student accountability, teacher accountability, and school accountability data from the 1995-1996 CCSSO survey, testing and school accountability data from the 1999 Quality Counts report, and the CPRE policy index devised by Carnoy and Loeb (2002) for 1999-2000.

Hanushek and Raymond (2005) used data from a survey and analysis done by CREDO (Fletcher and Raymond 2002) that allowed them to characterize states as "report card" states, meaning that they publicly reported schools' test results, or as "consequential" states, meaning that they reported test results and attached at least one school consequence to those results. Hanushek and Raymond (2005) then used these measures to describe the proportion of time that the state had the school accountability policies in place.

We followed Schiller and Muller's (2003) approach of using multiple measures to capture potentially different aspects of state accountability systems. We used the CCSSO data to create six different "accountability" measures for each state, including the number of grades in which a state mandated tests; whether the state required the use of assessments for student retention decisions; whether the state had any teacher accountability policy; the sum of positive and negative consequences for schools; an "accountability strength" measure that uses Carnoy and Loeb's (2002) methodology to develop a scale that ranges from 0 to 5; and an indicator variable that approximates the current NCLB policy by combining rules about testing, score disaggregation, and consequences for schools.[5] Note, however, that our approximation to NCLB policy is quite imperfect because during the 1990s no states had policies that exactly resembled NCLB and very few states had accountability policies that required testing in many grades, score disaggregation, and serious school consequences. Moreover, our measure does not indicate whether students had access to supplemental educational services or school transfer options, nor does it contain information on whether states required test score disaggregation and reporting at the school level.

Table 1 shows descriptive statistics for each of these measures (as well as for the state covariates that we include in our models). Appendix table 1 describes the content of these measures in more detail. Although accountability policies became more prevalent in the late 1990s, table 1 shows that some accountability policies were already in place by 1994.[6] States have required student testing for decades (Hamilton, Stecher, and Klein 2002) and in 1994 the typical state mandated tests in nearly two grades. Similarly, the typical state by 1994 had a little over two school accountability policies already in place and ranked about a 2 on the 0-5

[5] See the Data Appendix for details on how we constructed these measures.
[6] Throughout this paper, when we refer a particular year, it denotes the last year of the school year (e.g., 1994 refers to the 1993-1994 school year).

accountability strength scale. However, few states (only 13%) retained students based on school performance in 1994, and only 2 of the 48 states in our sample (4%) had any teacher accountability policy.

By 2000, accountability had increased in all states, although the average change (at least as measured by the CCSSO data) was not dramatic. Between 1994 and 2000, mandatory testing increased by nearly one more grade on average, an additional six percent of states based retention decisions on tests, an additional eight percent of states implemented at least one teacher accountability policy, the number of school accountability policies increased by about a fourth of a policy, and the accountability strength index increased by about a third of a point on the 0 to 5 scale. Figure 1 displays these trends graphically, showing data for the intervening years as well.[7]

Even though table 1 and figure 1 show that all types of accountability increased between 1994 and 2000, states that adopted accountability policies in one domain were not necessarily the same states that adopted policies in another domain. Table 2 shows correlations among the changes in our accountability measures (as well as their correlations with other state covariates). Somewhat surprisingly, changes in mandatory testing were not correlated with changes in other types of accountability. And, although changes in student retention and teacher accountability were moderately correlated with each other, neither was correlated with changes in school accountability or changes in accountability strength. Changes in school accountability and accountability strength correlated .78, however, indicating that both reflect similar types of policy changes. Changes in NCLB accountability correlated modestly with changes in teacher accountability policies (r=.38).

[7] For simplicity, we dichotomized the extent of testing, school accountability, and accountability strength measures before graphing them. We coded the first two variables as 1 if the state had a value of 3 or higher, and coded the accountability strength variable as 1 if the state had a value of 4 or higher (because 4 or higher represents "strong" accountability).

SASS Outcomes

In both 1994 and 2000, teachers in the SASS responded to numerous survey questions about themselves, their teaching, and their classrooms. Some of these survey questions measure the types of desirable consequences that accountability advocates hope will materialize in the wake of accountability reform; others measure the undesirable consequences that critics fear.

Desirable consequences

We measured teachers' effort by teachers' responses to two questions that asked them to report the hours per week they spent on school-related activities outside of school hours, both with and without students. Because both of these questions asked about school-related activities that may not be related to student learning (e.g., the question on time with students not only asked about time on tutoring, but also about time spent coaching, on field trips, and transporting students), these variables are imperfect measures of the construct we really care about: the time teachers spent on activities that are likely to improve student achievement. Moreover, these variables do not allow us to capture any changes in the effectiveness or efficiency of classroom instruction, even though teachers may respond to accountability pressures by covering more material and tolerating fewer distractions (see Chin 2002).

We measured administrators' efforts to improve student achievement in two ways. First, we examined teachers' reports of how much professional development they received during the past year.[8] Current evidence suggests that intense and content-focused professional development may help raise student achievement (Angrist and Lavy 2001; Cohen and Hill 2000; Jacob and Lefgren 2002; for a review, see Kennedy 1999). Second, we examined teachers' reports about

[8] We actually estimate the effect of accountability policies on the likelihood that teachers spent nine or more hours in each of these types of professional development in the past year. We defined exposure to professional development in this way because response options to the SASS professional development questions changed over time and the only comparable response category was "8 hours or less."

the size of their classes. Research indicates that students learn more in smaller classes, at least in the early grades, and that disadvantaged students benefit from smaller classes more than advantaged students do (for reviews, see Ferguson 1998 and Ehrenberg et al. 2001).

Undesirable consequences

SASS teachers answered two survey questions about how much control they felt they had in their classroom over selecting instructional materials, content, and teaching techniques. They also answered a question about how much influence they perceived teachers at their school as having over establishing curriculum. We standardized these items within each year and then combined them to measure teachers' perceptions of their instructional autonomy.

We measured teacher stability with a question that asked whether teachers had taught at their current school the previous year. And we measured teacher quality using four variables: years of teaching experience, the average SAT scores of the teacher's undergraduate college, whether or not the teacher had a regular (or advanced) teaching credential (as opposed to no credential or a provisional, probationary, temporary, or emergency credential), and whether or not the teacher had at least a Masters degree. These measures of teacher quality are admittedly poor proxies for instructional practices, although most studies show that students learn less when taught by novice teachers (see, e.g., Greenwald, Hedges, and Laine 1996; and Rivkin, Hanushek, and Kain 1998) and learn more when taught by teachers with stronger math and verbal skills (as proxied by the average test scores of the college they attended) (see, e.g., Ehrenberg and Brewer 1995; Ferguson and Ladd 1996). Other things equal, though, having a teacher with a credential or a graduate degree does not seem to improve student achievement (see, e.g., Goldhaber and Brewer 2000; Walsh 2001a).

Ambiguous consequences

We measured ambiguous consequences both in terms of teachers' instructional responses to accountability policies and administrators' resource allocation decisions. Teachers reported the amount of time they spent per week on reading/language arts, math, science, and history, and we investigated whether teachers spent a greater percentage of their time teaching a particular subject when their state mandated an additional test in that subject. We also examined whether administrators reduced class sizes or placed higher-quality teachers in grades in which mandatory state testing was required.

Analytic Approach

It is difficult to estimate the causal effect of state policies on educational outcomes because states adopt policies for a variety of reasons, and those reasons themselves may be correlated with educational differences among states. To take just one example, suppose that states with higher per-pupil expenditures are less likely than other states to adopt accountability reforms. Suppose, too, that high-expenditure states tend to attract better teachers because those states pay more. If an analysis of the effects of accountability on teacher quality did not take educational spending differences among states into account, it would incorrectly conclude that stronger accountability reforms led to reductions in teacher quality. Studies that use only one year of data try to avoid these problems by controlling statistically for confounding variables (see, e.g., Carnoy and Loeb 2002). However, is it difficult to control for a large number of state-level variables (because the total number of states is relatively small) and it is impossible to know whether confounding variables that have been omitted from the model still bias the estimated policy effects.

Fixed effects models improve on cross-sectional models by eliminating the confounding effects of all state variables that do not change over time. We use such models in this paper. We estimate the impact of accountability reform by pooling two cross-sections of the SASS (from the 1993-1994 and 1999-2000 school years) and regressing our various outcome variables on teacher demographics, school characteristics, district characteristics, and measures of accountability policies in each state in a given year.[9] We add to these models indicator variables for each state and an indicator variable for year. These models estimate the impact of within-state changes in accountability on within-state changes in class size, teachers' perceptions of instructional autonomy, teacher quality, and so on, holding constant the overall secular change in these outcomes.[10] Fixed effects models do not, however, account for other within-state changes that may confound the effects of changes in accountability policies. We try to deal with this problem by controlling for changes in other state-level variables that may have influenced changes in our outcomes.

Although the benefits of this "within-state" strategy recommend it above cross-sectional strategies, it has limitations. Although a six-year period is a reasonably long period over which to examine accountability changes, some state accountability policies changed only modestly or not all over this time period. This lack of change limits our ability to detect statistically significant effects. Moreover, any mistakes we have made in measuring state accountability policies make it even harder for us to detect statistically significant effects. We caution readers that the standard errors in our tables, which we adjust for dependencies among teachers within

[9] We originally planned to pool more than two cross-sections, because SASS also collected data from teachers in 1987-1988 and 1990-1991. However, we could not find school accountability data for those earlier years that were comparable to the accountability data that we used for the mid and late 1990s.

[10] These models make the most sense for outcomes that reflect teachers' experiences during the school year, such as teachers' sense of autonomy or the amount of time they devote to teaching particular subjects. In contrast, the teacher quality and class size models are only valid to the extent that teachers or administrators made their employment or resource allocation decisions based on relatively accurate expectations about the accountability policies that would be in place in their state during the 1999-2000 school year.

states, are often quite large. Thus, we do not have enough statistical power to detect some potentially important effects of accountability policies.

Most of our models take the following general form[11]:

$$(1) \qquad y_{ist} = \lambda X_{ist} + \gamma LP_{st} + \beta A_{st} + \mu_s + \alpha_t + \varepsilon_{ist},$$

where y_{ist} is an outcome variable for teacher *i* in state *s* at time *t*; X_{ist} are teacher characteristics and the characteristics of teachers' schools and districts. These include teachers' gender, ethnicity, age (as well as a quadratic for age)[12], and grade taught; the natural log of school size, the percentage of students in the school who are eligible for subsidized meals (and a quadratic for subsidized meals), the percentage of minority students in the school (and a quadratic for percent minority), indicator variables for whether the school is located in an urban, rural, or suburban area, and indicators for whether the school is a charter school or a magnet school; the natural log of district size, the percentage of students in the district who are eligible for subsidized meals, and the percentage of minority students in the district.[13] LP_{st} are state characteristics that may influence teacher labor markets, including state per pupil expenditures, state unemployment rate, and the natural log of state school enrollments, as well as state education policies that may also affect teachers, including class size reduction policies and

[11] For continuous outcomes, we estimate linear regressions. For dichotomous outcomes, we estimate logistic regressions. Although we report logit coefficients and standard errors in our logistic regression tables, we discuss odds and probabilities in the text.

[12] We exclude teachers' age and age-squared from our equations when we estimate the impact of accountability on teacher quality (i.e., experience, education, credentials, college quality, and stability) because accountability policies may influence those outcomes *via* their effect on the age distribution of teachers. For example, if accountability policies encourage older teachers to retire earlier than they otherwise would, those earlier retirements may leave schools with a less experienced, less well-educated teacher workforce. Under such a scenario, if we controlled for age, we would risk understating the effects of accountability on teacher quality.

[13] Because the SASS data do not separately identify charter schools in 1994, our dummy for charters reflects a comparison between designated charters in 2000 and all other schools. Appendix table 4 shows descriptive statistics for the teacher-level variables in our models.

teacher certification test requirements. A_{st} are state accountability policies and μ_s and α_t are state and year indicators.[14]

Near the end of this paper, we use two additional models to examine whether teachers or school administrators reallocated resources within schools in response to assessment and accountability policies. To assess whether teachers spent more time teaching a subject if the state mandated more testing in that subject, we estimate:

$$(2) \qquad y_{ist} = \lambda X_{ist} + \gamma LP_{st} + \beta TestSubject_{st} + \phi TestOtherSubjects_{st} + \mu_s + \alpha_t + \varepsilon_{ist},$$

Equation 2 is analogous to equation 1, except that our dependent variables are the percentages of time teachers report spending on reading, math, science, or social studies instruction, and we replace the accountability policies, A_{st}, with $TestSubject_{st}$, which is the sum of the number of grades with a mandatory test in a particular subject (e.g., reading). We also add a control, $TestOtherSubjects_{st}$, for the number of mandatory tests in grades 1-6 in the other core subjects (e.g., math, history, and science).[15] This model tries to identify the effects of changes in mandatory testing of a particular subject on time spent teaching that subject, holding constant the testing regimen for other subjects.

Our third model addresses the question of whether administrators reallocated classroom resources in response to testing and accountability policies. Specifically, we examine whether teachers who taught in grades that were subjected to mandatory state tests were more likely than their counterparts who taught in grades without mandatory tests to have smaller classes, more

[14] We weight all our models with the teacher weights so that our estimates are nationally representative of teachers in grades 1-6 who teach in regular or "special program emphasis" schools in each year, and we estimate robust standard errors. Because we estimate our models at the teacher level, but policy variation occurs at the state level, we correct the standard errors by clustering at the state level (Bertrand, Duflo, and Mullainathan 2004).

[15] If a state mandated multiple tests in the same subject in any particular grade, we counted all tests in that subject and grade as one test.

experience, better academic skills, credentials, or MA degrees. These equations take the following form:

$$(3) \qquad y_{ist} = \lambda X_{ist} + \gamma LP_{st} + \beta Test_{gst} + \phi Teachtestedgrade_{ist} + \alpha_t + \varepsilon_{ist}$$

These models include the same variables in LP_{st} and X_{ist} as in equations 1 and 2. Note that the teacher variables in all our models include indicators for the grade taught by each teacher. *Test* represents a set of six indicator variables for whether the state, in a given year, mandates a test in at least one core subject in grades 1 through 6. *Teachintestedgrade* is a variable indicating whether a teacher teaches in a tested grade. We estimate these models separately for states with weak to moderate accountability polices (states that scored below a "4" on the Accountability Strength index) and states with strong accountability policies (states that scored "4" or higher on the Accountability strength index).

To assess whether accountability reforms differentially affected teachers who taught in low-income schools, we estimate each of the models described above separately for various groups of teachers. Our first set of models compares effects for teachers in schools in which fewer than 40% of the students qualified for subsidized meals ("non-poor schools") and teachers in schools in which at least 40% of the students qualified for subsidized meals ("poor schools").[16] Our second set of models refines these classifications to also consider the percentage of minority students at the school. Specifically, we compare outcomes for teachers in "non-poor, white" schools; "non-poor, minority" schools; "poor, white" schools; and "poor, minority" schools, where we define "poor" by the same 40% cut-off and "white" as fewer than 20% non-white students.[17] These subsets of schools differ in predictable ways. For example, "minority"

[16] We define "poor" schools as those in which 40% or more of the students are eligible for subsidized meals because NLCB uses the same cut-off to define schools as having a "high concentration of low-income students."
[17] The 20% white cut-off divides our weighted sample of teachers roughly in half.

schools, regardless of their poverty status, are more likely than their "white" counterparts to be located in urban areas. And "poor, white schools" are much more likely to be located in rural areas.

Results

Desirable Consequences

Teacher effort

Proponents of accountability reform hope that accountability policies will encourage teachers to work harder to raise student achievement. Table 3 indicates that teachers responded to at least a couple of the accountability reforms of the 1990s by spending a little more time with students outside of normal school hours. In particular, increases in teacher accountability and NCLB accountability were associated with teachers spending 15 more minutes per week with students—or about nine more hours during a 36-week school year.[18] Table 4 shows that greater accountability was also associated with teachers spending more time outside normal school hours preparing lessons, grading, or in meetings.

We found no evidence that the effects of accountability on time spent with students differed between teachers in poor and non-poor schools. The evidence on differential effects of accountability on teachers' preparation time is more mixed. We found some evidence that increases in particular accountability polices, specifically the extent of testing and student retention, increased teachers' preparation time largely in non-poor and white schools. Yet we

[18] The variable measuring time spent with students outside of normal school hours is very skewed, with over 60% of teachers reporting that they spent less than one hour a week on such activities (teachers reported zero hours in such cases because they were not allowed to report fractional hours). Consequently, we also estimated these models by recoding time spent with students to a dummy variable, where 1 indicated that the teacher spent at least 1 hour a week with students outside normal hours. Those models indicated that having a statewide teacher accountability policy increased the odds that teachers in poor schools spent at least 1 hour with students outside of school hours by 27%. We found a larger, statistically significant association in poor, white schools but a smaller, statistically insignificant association in poor, minority schools. And, in these models, none of the other types of accountability policies had any statistically significant association with time spent with students in any of our samples.

also found evidence (albeit weaker) that increases in other accountability policies, specifically school accountability and accountability strength, increased teachers' preparation time in poor, minority schools.

Professional development

Accountability proponents also hope that accountability policies will stimulate school administrators to provide teachers with additional professional development resources that will help them raise student achievement. We examined effects for three different types of professional development: in-depth study of content, study of teaching methods, and assessment.

We found little evidence that accountability policies changed the amount of time teachers spent in professional development focused on in-depth study of their content areas, and the evidence we found was contradictory (see table 5). We found slightly more evidence that teachers received additional professional development in "methods of teaching" as a result of accountability policies, but there, too, the evidence varied inconsistently across types of school and in one instance was negative (see table 6).

In contrast, we found consistent evidence that additional state mandated testing was associated with additional professional development focused on student assessment (i.e., teaching teachers about assessment), especially in poor and minority schools (see table 7). For example, when a state added testing in one additional grade, the odds that a teacher in a poor school received nine or more hours of "assessment" professional development increased by 21%.[19]

Class size

[19] Table 7 also shows that the adoption of a student retention policy may have been associated with increases in the amount of "assessment" professional development that teachers in non-poor schools received.

Administrators might also respond to accountability pressures by reducing class sizes. Parents and teachers alike appreciate smaller classes, and research suggests that smaller classes (at least if they are small enough) raise achievement and reduce achievement gaps in the early elementary grades (see, e.g., Ferguson 1998; Ehrenberg et al. 2001). Table 8 suggests that state mandated student retention policies and teacher accountability policies were associated with reductions in class sizes in both poor and non-poor schools. Testing policies and accountability strength may also have been associated with class size reduction. Because some states implemented class size reduction policies during the same period that they increased accountability, our models control for changes in states' class size policies. Consequently, a concurrent within-state trend in class size reduction policies probably cannot explain our results, but a different, unmeasured within-state trend might.

Undesirable Consequences

Critics of accountability reforms worry that these policies will negatively impact teachers' working conditions and persuade good teachers to transfer to other schools, leave teaching altogether, or not enter the profession in the first place.

Instructional autonomy

Our results provide some evidence to support critics' concerns that accountability policies negatively affect teachers' sense of autonomy, at least in poor schools and in poor, minority schools. But the negative effects we find range from very small to moderate. For example, in high minority, poor schools, the expansion of state mandated testing to one more grade was associated with a .06 standard deviation decline in teachers' perceived autonomy. The adoption of NCLB-like accountability was associated with a quarter of a standard deviation

decline in those schools.[20] Our results also suggest that some types of accountability policies, such as teacher accountability, may *increase* teachers' perceived autonomy in non-poor schools. These results imply that certain types of accountability policies may give teachers in advantaged schools more leeway with respect to strategy and curriculum as long as they meet achievement goals.

Teacher turnover

Critics of accountability reform suspect that it will increase teacher turnover at the most disadvantaged schools. Turnover probably has a negative impact on academic success if students suffer when their teachers are less familiar with a particular school's policies and practices. Increased turnover is probably also associated with an increase in the percentage of teachers at poor, minority schools who are inexperienced (because vacancies at schools tend to

[20] Because concerns about teacher autonomy permeate discussions about accountability, we investigated these findings in more depth. First, we examined whether accountability policies had different effects on teachers' sense of control over the content of their instruction relative to their sense of control over teaching methods. A case can be made that high-quality teachers find it less offensive to be told *what* they should teach than *how* they should teach it. When we compared the effects of accountability policies on two items in our scale, teachers' perceptions of how much control they felt over "selecting content, topics, and skills to be taught" and how much control they felt over "selecting teaching techniques," the results for the content item mirrored those for our overall accountability scale. The effects on "teaching strategies" varied somewhat, however. We did not find negative effects of increased state testing on teachers' perceptions of control over their teaching methods among teachers in poor schools. However, teachers in poor, minority schools *did* feel statistically significantly less control over their teaching strategies when states added more testing requirements (though those effects were about half the size of the already small effects for content). What's more, teachers in non-poor, white schools felt statistically significantly *more* control over their teaching strategies when states added more testing requirements (at the $p<.10$), and the difference between the results for teachers in poor, minority schools and teachers in non-poor, white schools were statistically significant. While these results provide some reassurance that accountability policies negatively affect teachers' control over content more than they negatively affect control over methods, they also suggest that increased testing causes teachers in the most disadvantaged schools to have less autonomy over their teaching methods than do their peers in other types of schools, providing yet another reason for teachers to want to migrate to more advantaged settings. Second, because critics fear that accountability policies will drive "good" teachers out of the profession, we tried to investigate whether accountability policies affected the perceived autonomy of experienced teachers, or teachers from highly-selective colleges, the most. Our results showed that, in poor schools, increased state mandated testing was associated with a larger decline in autonomy among the most inexperienced teachers (those who had taught for four or fewer years) relative to their more experienced counterparts. We also found suggestive evidence that teachers from less selective colleges experienced greater declines in autonomy than did their counterparts from more selective colleges, although those differences were not statistically significant. The differentials by teacher experience may indicate that inexperienced teachers are more likely than their experienced counterparts to be monitored to ensure that they are using mandated methods, that inexperienced teachers feel less authority to ignore curricular impositions, or that inexperienced teachers are simply less adept (because of their inexperience) at integrating their own ideas and methods with the mandated curriculum.

be filled based on seniority, and experienced teachers tend to prefer to teach at more advantaged schools). On the other hand, increased turnover may improve student achievement if the weakest teachers respond to accountability pressures by exiting the profession.

Table 10 suggests that, if anything, accountability policies caused more teacher instability in non-poor, white schools than in minority schools. For example, when a state increased its accountability strength by one unit on our index, the odds that a teacher in a non-poor, white school reported being new to her school increased by 31%. The effects of teacher accountability and NCLB accountability were even larger. Unfortunately, our data cannot tell us why accountability policies seem to have caused instability in non-poor, white schools. It is possible that those schools reduced class sizes more than other schools did, and thus added more teachers (our estimates in table 8--which show the effects of accountability policies on class size reduction—have large standard errors that make it impossible to rule out this theory). It is also possible that teachers in those schools reacted especially negatively to accountability policies and left their jobs, creating vacancies for new teachers. Because we did not find any indication that turnover increased in minority (poor or non-poor) schools, however, our data do not indicate that teachers responded to accountability policies by leaving disadvantaged schools to teach in advantaged schools.

Teacher quality

Turnover is largely a concern to the extent that it results in lower teacher quality. Because we cannot measure high quality instruction or student learning with our data, we rely instead on common proxies for teacher quality, including teachers' experience, the selectivity of their undergraduate institution (as a proxy for their academic skills), whether a teacher has a regular or advanced credential, and whether a teacher has an MA degree.

Using these teacher quality measures, we found inconclusive results about the impact of accountability policies on teacher quality. Table 11 shows hardly any statistically significant impact of accountability policies on teacher experience, with the exception that increases in teacher and school accountability may have led to increases in teacher experience in poor, white schools, and NCLB accountability may have led to increases in teacher experience in non-poor minority schools. Note, however, that the standard errors on our coefficients are large. This means that we do not have enough statistical power to detect small to moderate changes in teacher experience that may have occurred as a result of accountability policies.[21]

Table 12 suggests that greater accountability caused a small (a twentieth to a tenth of a standard deviation) reduction in teachers' academic skills. Student retention policies were associated with a decline in teachers' skills that seems to have affected teachers in all types of schools. In contrast, increases in school accountability and accountability strength seem to have caused a decline in teachers' skills in non-poor, white schools but not in poor, minority schools.[22]

Overall, accountability policies seem to have increased teacher credentialing. Yet table 13 provides somewhat mixed evidence on how accountability policies affected credentialing in different types of schools. In non-poor, white schools, most of the effects of accountability policies on credentialing were positive. In poor, minority schools, both the existence of a student retention policy and increases in accountability strength were associated with increases in

[21] Because students seem to learn less when assigned to novice teachers, we also investigated whether accountability policies affected the likelihood that teachers had four or fewer years of experience (results available from the authors). Those models indicated that teacher accountability and NCLB accountability reduced the odds that teachers in non-poor schools and teachers in non-poor, minority schools, in particular, were inexperienced. But the results for our accountability strength measure told the opposite story.

[22] We ran these models for teachers with different levels of experience, as well, in order to investigate whether accountability reforms dissuaded recent graduates from selective colleges from becoming teachers, or caused academically-skilled experienced teachers to exit the profession. We found that the decline in academic skills occurred only among experienced teachers (i.e., those with more than four years of experience).

credentialing. But the existence of an NCLB-like accountability policy may have been associated with a reduction in credentialing.

The evidence for teacher education is less mixed than that for credentialing. Table 14 shows that only a few accountability policies had any effect on teacher education. Those that did (student retention and teacher accountability policies) increased the odds of having a Masters degree by between 12 and 25%. These types of policies seem to have had their largest effects on teachers in non-poor, minority schools.

Ambiguous Consequences

Skeptics worry that teachers and administrators will respond to accountability pressures with marginal tinkering rather than important instructional changes. Previous research has found that accountability pressures motivate teachers to devote more time to teaching tested subjects (Hamilton, Stecher, and Klein 2002; Jacob 2004; Koretz 2002). The results in table 15 provide additional evidence of such changes. In both non-poor and poor schools, teachers devoted more classroom time to a particular subject when their state required more testing in that subject.[23] But these effects were quite small. For example, when states mandated testing in an additional grade in science or history, teachers increased the percentage of time they spent teaching science or history by less than one percent (or about a tenth of a standard deviation).

We also investigated whether administrators responded to accountability policies by reducing class sizes in tested grades or allocating their best teachers to tested grades, especially in states that had "strong" accountability policies (see table 16). Administrators seem to have

[23] In our models, increased math testing was unrelated to the time teachers spent teaching math. A likely explanation for this result is that our models cannot distinguish changes in math testing from changes in reading testing because states tended to require tests in both subjects when they decide to test an additional grade. In other words, changes in math and reading testing were very highly correlated. Our results may suggest, however, that when states mandate testing in both subjects in a grade, elementary school teachers are more likely to respond by teaching more reading than by teaching more math.

assigned teachers they perceived as more highly qualified to tested grades. For example, teachers who taught tested grades were more likely than their counterparts who taught untested grades to have an MA, and this effect did not differ by school type or accountability strength. Similarly, in low-income schools located in "strong accountability" states, teachers who taught tested grades had over two more years of experience than did their counterparts in untested grades and were also slightly more likely to be credentialed. In addition, the data hint that administrators in low income schools were more likely to assign teachers from more selective colleges to tested grades and to reduce class sizes in tested grades when their schools were located in "strong accountability" states.[24]

Conclusion

Our results offer support to both proponents and critics of accountability reform. On the one hand, accountability policies seem to have motivated teachers to spend more time working on school-related activities outside of school hours. Accountability policies may have also encouraged administrators to offer more professional development and to reduce class sizes. On the other hand, accountability policies seem to have reduced teachers' sense of autonomy. Skeptics would probably add, however, that most of these effects, whether positive or negative, were quite small. They would probably also argue that the clearest changes we found in response to accountability policies seem to have involved strategic decisions intended to temporarily improve students' scores rather than permanently improve their skills.

One of our purposes in writing this paper was to examine the effect of accountability policies on school inequality. We expected that teachers in low-income, minority schools would

[24] Table 16 shows that class sizes were larger in tested grades in weak accountability states, presumably because tested grades tend to be the upper elementary grades and upper grades tend to have larger class sizes. Yet the difference between strong and weak accountability states suggests that administrators responded to accountability pressures by reducing class sizes in the upper grades.

bear the brunt of the effects of accountability reform (Clotfelter et al. 2004). And we worried that policies intended to raise achievement and reduce achievement gaps might have led, instead, to dramatic declines in instructional quality in schools that already faced the most challenges. But we did not find strong negative effects in disadvantaged schools. We did find that teachers in poor or minority schools seemed to receive more professional development about "assessment" after increases in state-mandated testing than did their counterparts in non-poor, white schools. And teachers in poor, minority schools felt less autonomy over their curriculum and methods after increases in state-mandated testing than did their counterparts in non-poor, white schools. But we found no consistent evidence that accountability policies differentially affected teacher quality in disadvantaged schools. And we also found some positive changes from which we would expect students to benefit, such as increases in the time that teachers spent doing school-related work outside of school hours and decreases in class sizes. These mixed results raise the crucial question of what will transpire as states comply with NCLB.

Although we cannot answer that question with certainty, both because NCLB is a federal rather than state policy and because none of the accountability policies adopted by states during the 1990s had the exact same components as NCLB, this paper informs the debate in at least two ways. First, our results strongly suggest that accountability policies do, in fact, change educators' behavior. Consequently, policymakers at the federal, state, and district levels need to ensure that the incentives they create for administrators and teachers are aligned with the educational outcomes they desire. To take one example, our results suggest that administrators try to concentrate resources on grades that are tested. NCLB does not mandate testing until third grade and, as a consequence, administrators may try to concentrate educational resources in the third grade or above. Yet it is very hard for students to catch up academically after they have

already fallen behind in the early elementary grades (see, e.g., Juel 1988). Consequently, policymakers would be wise to add additional incentives at the federal, state, or district levels that encourage schools to measure and remediate students' skill gaps during kindergarten, first, and second grade.

Our results also show that teachers spend more time teaching tested subjects, providing further support for the adage "You get what you test" (see Linn 2000). Consequently, policymakers need to ensure that testing requirements cover the subjects, knowledge, and skills that they want children to master. For example, because NCLB currently excludes history testing, teachers will probably teach less history unless policymakers either mandate history testing as part of NCLB or create an additional policy that mandates history testing.

Second, our results suggest that different types of accountability policies have different types of effects on teachers and administrators. During the 1990s at least, accountability policies that imposed consequences on schools seem to have been less effective in changing educators' behavior than were accountability policies that imposed consequences on teachers or students. Although these results are tentative, because only a small number of states adopted teacher or student accountability policies over the 1990s, they nonetheless suggest that policymakers should consider adding teacher, and perhaps student, accountability components to NCLB when it comes up for reauthorization.[25] Such decisions should be made, however, only after assessing whether the direct effects of teacher and student accountability policies on student academic success are positive or negative.

[25] States' use of teacher accountability policies during the 1990s did not vary enough for us to distinguish the effects of negative teacher sanctions from the effects of positive teacher sanctions. We do know, however, that the states that adopted teacher accountability policies during the 1990s—which are the states that drive our results—typically had policies that included either only positive incentives (such as recognizing teachers for students' score gains or giving teachers monetary rewards) or both positive and negative incentives (such as both recognizing and evaluating teachers based on students' scores).

When NCLB comes up for reauthorization, legislators will have another opportunity to weigh in on the most significant federal education policy in decades. Despite the significance of this policy, the results in this paper indicate that accountability reform will not be a panacea for our nation's educational problems. Significantly raising achievement or eliminating the achievement gap will require more than systemic school reform. Nonetheless, our results suggest that accountability policies do have the potential to change educators' attitudes and behaviors. Yet the changes we found during the 1990s were quite small. These small effects probably reflect considerable heterogeneity in educators' responses to the incentives created by accountability reform. If these results signal that educators try to respond to accountability policies but make widely varying choices about how to respond, and choices that are not always equally effective, it may make the most sense to couple strong accountability policies with the research-based guidance and support that educators need to improve instruction.

References

American Association of Colleges for Teacher Education. 1993. *Teacher Education Policy in the States: A 50-State Survey of Legislative and Administrative Actions.* Washington, DC.

Angrist, Joshua D. and Victor Lavy. 2001. "Does Teacher Training Affect Pupil Learning? Evidence from Matched Comparisons in Jerusalem Public Schools. *Journal of Labor Economics* 19: 343-69.

Asimov, Nanette. 2005. "37 Schools Forced to Make Changes; Improving or Not, They Failed U.S. Goals." *San Francisco Chronicle,* B1, September 20.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan. 2004. "How Much Should We Trust Differences-in-Differences Estimates?" *Quarterly Journal of Economics* 119:249-275.

Boyd, Donald, Hamilton Lankford, Susanna Loeb, and James Wyckoff. 2005. "The Impact of Assessment and Accountability on Teacher Recruitment and Retention: Are There Unintended Consequences?" Working paper.

Bureau of Labor Statistics. Department of Labor. Current Population Survey. October 1993-October 2000. "Civilian labor force and unemployment by state, seasonally adjusted." Accessed January 15, 2002 at http://www.bls.gov/schedule/archives/laus_nr.htm

Carnoy, Martin and Susanna Loeb. 2002. "Does External Accountability Affect Student Outcomes? A Cross-State Analysis." *Educational Evaluation and Policy Analysis* 24:305-331.

Chenfeld, Mimi B. June 2006. "Handcuff me, Too!" *Phi Delta Kappan* 87 (10): 745-747.

Chin, Tiffani. 2002. "The Accountable Classroom: How assessment and evaluation influence everyday classroom practices." Paper presented at the annual meetings of the American Sociological Association, Chicago, Illinois.

Clotfelter, Charles, Helen Ladd, Jacob Vigdor and Roger Aliaga. 2004. "Do School Accountability Systems Make It More Difficult to Attract and Retain High Quality Teachers?" *Journal of Policy Analysis and Management* 23(2): 251-271.

Cohen, David K. and H.C. Hill. 2000. "Instructional Policy and Classroom Performance: The Mathematics Reform in California. *Teachers College Record* 102: 294-343.

Council of Chief State School Officers & North Central Regional Educational Laboratory. 1995. *State Student Assessment Programs Database, June 1995.* Washington, DC: North Central Regional Education Laboratory.

Council of Chief State School Officers & North Central Regional Educational Laboratory. 1996. *State Student Assessment Programs Database, May 1996.* Washington, DC: North Central Regional Education Laboratory.

Council of Chief State School Officers. 1997. *Annual Survey of State Student Assessment Programs, Fall 1996*. Washington, DC: Council of Chief State School Officers.

Council of Chief State School Officers. 1998. *Annual Survey of State Student Assessment Programs, Fall 1997*. Washington, DC: Council of Chief State School Officers.

Council of Chief State School Officers. 1999. *Data from the Annual Survey of State Student Assessment Programs, Fall 1999*. Washington, DC: Council of Chief State School Officers.

Council of Chief State School Officers. 2001. *State Student Assessment Programs Annual Survey, Fall 2001*. Washington, DC: Council of Chief State School Officers.

Council of Chief State School Officers. 2001. *State Student Assessment Programs Annual Survey, Spring 2001*. Washington, DC: Council of Chief State School Officers.

Cullen, J.B. and R. Reback. 2002. "Tinkering Toward Accolades: School Gaming under a Performance Based Accountability System." Working paper, Department of Economics, University of Michigan.

Deere, Donald and Wayne Strayer. 2001. "Putting Schools to the Test: School Accountability, Incentives, and Behavior." Working Paper 113, Private Enterprise Research Center, Texas A&M University.

Dobbs, Michael. 2005. "New Rules for 'No Child' Law Planned." *The Washington Post*. Washington, A13, April 7.

Eberts, Randall, Kevin Hollenbeck, and Joe Stone. 2002. Teacher Performance Incentives and Student Outcomes. *Journal of Human Resources* 37(4): 913-927.

Education Week. 2000. *Quality Counts 2000: Who Should Teach?* Accessed January 25, 2006 at http://counts.edweek.org/sreports/qc00/ .

Ehrenberg, Ronald G. and Dominic J. Brewer. 1995. "Did Teachers' Race and Verbal Ability matter in the 1960's? Coleman Revisited." *Economics of Education Review* 14: 1-21.

Ehrenberg, Ronald G., Dominic J. Brewer, Adam Gamoran, and J. Douglas Willms. 2001. "Class Size and Student Achievement." *Scientific American* 285 (5, November 1): 78-85.

Ferguson, Ronald F. and Helen F. Ladd. 1996. "How and Why Money Matters: An Analysis of Alabama Schools." Pp. 265-298 in *Holding Schools Accountable*, edited by Helen Ladd. Washington, DC: Brookings Institution.

Ferguson, Ronald F. 1998. "Can Schools Narrow the Black-White Test Score Gap?" Pp. 318-374 in the *Black-White Test Score Gap*, edited by Christopher Jencks and Meredith Phillips. Washington, DC: Brookings Institution.

Figlio, David N. and L.S. Getzler. 2002. "Accountability, Ability, and Disability: Gaming the System?" Working Paper No. 9307, National Bureau of Economic Research.

Gallagher, Chris W. January 2004. "Turning the Accountability Tables: Ten Progressive Lessons from One 'Backward' State. *Phi Delta Kappan* 85(5): 352-360.

Goertz, M.E. and M.C. Duffy. 2001. "Assessment and Accountability Systems in 50 States:1999-2000." in *CEPRE Research Report No. RR-046*. Philadelphia, PA: Consortium for Policy Research in Education.

Goldhaber, Dan and Dominic Brewer. 2000. "Does Teacher Certification Matter? High School Teacher Certification Status and Student Achievement." *Educational Evaluation and Policy Analysis* 22:129-145.

Gordon, Robert. 2005. "What Democrats need to say about education." *The New Republic,* June 6-13.

Greenwald, Rob, Larry Hedges, and Richard Laine. 1996. "The Effect of School Resources on Student Achievement." *Review of Educational Research* 66:361-396.

Hamilton, L.S., B.M. Stecher, and S.P. Klein. 2002. *Making Sense of Test-Based Accountability in Education*. Santa Monica, CA: RAND Corporation.

Hanushek, Eric and Margaret Raymond. 2005. "Does School Accountability Lead to Improved Student Performance?" *Journal of Policy Analysis and Management* 24:297-327.

Jacob, Brian. 2005. "Accountability, Incentives and Behavior: The Impact of High Stakes Testing in the Chicago Public Schools." *Journal of Public Economics* 89:761-796.

Jacob, Brian A. and Lars Lefgren. 2002. "The Impact of Teacher Training on Student Achievement: Quasi-Experimental Evidence from School Reform Efforts in Chicago." Working Paper. Cambridge, MA: National Bureau of Economic Research.

Juel, Connie. 1988. "Learning to Read and Write: A Longitudinal Study of Fifty-Four Children from First through Fourth Grade," *Journal of Educational Psychology* 80: 437-47.

Kennedy, Mary. 1999. "Form and Substance in Mathematics and Science Professional Development." Arlington, VA: National Science Foundation.

Koretz, Daniel M. 2002. "Limitations in the Use of Achievement Tests as Measures of Educators' Productivity." *Journal of Human Resources* 37(4): 752-777.
Lee, Jaekyung and Kenneth Wong. 2004. "The Impact of Accountability on Racial and Socioeconomic Equity: Considering Both School Resources and Achievement Outcomes." *American Educational Research Journal* 41:797-832.
Linn, R. L. (2000). "Assessments and Accountability." *Educational Researcher*, 23(9), 4-14.
Loeb, Susanna and Felicia Estrada. 2005. "Have Assessment-Based School Accountability Reforms Affected the Career Decisions of Teachers?" in *Measurement and Research Issues in the Accountability Era,* edited by Carol Anne Dwyer. Mahwah, NJ: Lawrence Erlbaum Assoc.
Marshak, David. November 2003. "No Child Left Behind: A Foolish Race into the Past." *Phi Delta Kappan* 85(3): 229-231.
Mathews, Jay. 2006. "Let's Teach to the Test." *Washington Post*, p. A21, February 20.
McNeil, Linda M. 2000. *Contradictions of School Reform: Educational Costs of Standardized Testing*. New York: Routledge.
National Association of State Directors of Teacher Education and Certification. 1999. *The NASDTEC Manual on the Preparation and Certification of Educational Personnel*. Dubuque, IA:Kendall/Hunt Publishing Company.
Pate-Bane, Helen. 2004. "Reduce Class Size Now: State of the States." Accessed January 25, 2006 at http://www.reduceclasssizenow.org/state_of_the_states.htm .
Perie, M., W. Grigg, and G. Dion. 2005. *The Nation's Report Card: Mathematics 2005*. (NCES 2006– 451) Washington, DC: U.S. Government Printing Office.
Perie, M., W. Grigg, and P. Donahue. 2005. *The Nation's Report Card: Reading 2005*. (NCES 2006– 453) Washington, DC: U.S. Government Printing Office.
Public Law 103-382. 103rd Cong., 20 October 1994. *Improving America's Schools Act of 1994*.
Public Law 107–110. 107th Cong., 8 January 2002. *No Child Left Behind Act of 2001*.
Rivkin, S.G., E.A. Hanushek, and J.F. Kain. 1998. "Teachers, Schools, and Academic Achievement." Working Paper No. 6691, National Bureau of Economic Research.
Schaffer, Bob. 2005. "U.S. Education Reform and the Role of the States." *Denver Post*, E-05, December 25.
Schiller, Kathryn and Chandra Muller. 2003. "Raising the Bar and Equity? Effects of High School Graduation Requirements and Accountability Policies on Students' Mathematics Course Taking." *Educational Evaluation and Policy Analysis* 25:299-318.
U.S. Department of Education. 2000. *State Regulation of Private Schools*. Washington, D.C.: Office of Non-Public Education.
U.S. Department of Education. 2003. National Center for Education Statistics. *Digest of Education Statistics, 2002*. NCES 2003–060, by Thomas D. Snyder. Production Manager, Charlene M. Hoffman. Washington, DC: 2003.
Walsh, Kate. 2001. *Teacher Certification Reconsidered: Stumbling for Quality*. Abell Foundation, Baltimore, MD.
Wilms, Wellford W. April 2003. "Altering the Structure and Culture of American Public Schools." *Phi Delta Kappan* 84 (8): 606-615.

Data Appendix

We constructed two accountability measures that require further explanation:

Accountability Strength

We tried to replicate the index that Carnoy and Loeb (2002; hereafter C&L) used in their paper on the effects of accountability on student achievement. Because we wanted to measure accountability in both 1993-1994 and 1999-2000, and because the C&L data source (CPRE) was not available for 1993-1994, we tried to follow their classification rules yet make an index based on the CCSSO data. When we applied their schema to the CCSSO data, and then correlated our 1999-2000 index with their 1999-2000 index, the two indices were highly, but not perfectly, correlated (r=.81). Most states received either the same index score across the two data sources or received a score that differed by only one point. Six states differed by more than one point, however. Because these discrepancies concerned us, we tried to verify the CCSSO data for these states by consulting a third data source: *Quality Counts 2000* (Education Week 2000). We assigned each state a new accountability index score based on the Quality Counts data and then compared the results with our CCSSO results. With the exception of Pennsylvania and Tennessee, the Quality Counts data largely confirmed our state accountability classifications.

NCLB Accountability

We intended to make a variable that reflected the requirements of the NCLB legislation. We did not have any items in both years that measured school transfer options or access to supplemental educational services, so we could not include those aspects of the legislation in our measure. NCLB requires testing in math and reading in grades 3 through 8, requires score disaggregation by race and social class at the school level, puts schools on probation if they do not meet their Adequate Yearly Progress targets, and then removes resources from them, takes them over, or dissolves them entirely if they continue to fail to meet their performance targets. The CCSSO data did not allow us to measure whether states required test score disaggregation specifically at the school level. So, we began by creating a variable that required a state to have *all* of the following policies in order to be classified as an NCLB state:

- Mandatory math and reading tests in 3 elementary grades;
- Disaggregated test scores by race and SES (reported at any level);
- Schools receive a warning *or* probation based on test scores; and
- Schools receive a monetary penalty *or* are taken over *or* are dissolved based on test scores

However, very few states met these requirements in 1994 or 2000. Relaxing the test requirement, by requiring only two elementary grades to have mandatory reading and math tests, did not improve the numbers much. Our final NCLB approximation requires *all* of the following:

- Mandatory math and reading tests in 2 elementary grades;
- Disaggregated test scores by race and SES (reported at any level); and
- Schools receive a warning *or* probation *or* a monetary penalty *or* are taken over *or* are dissolved based on test scores.

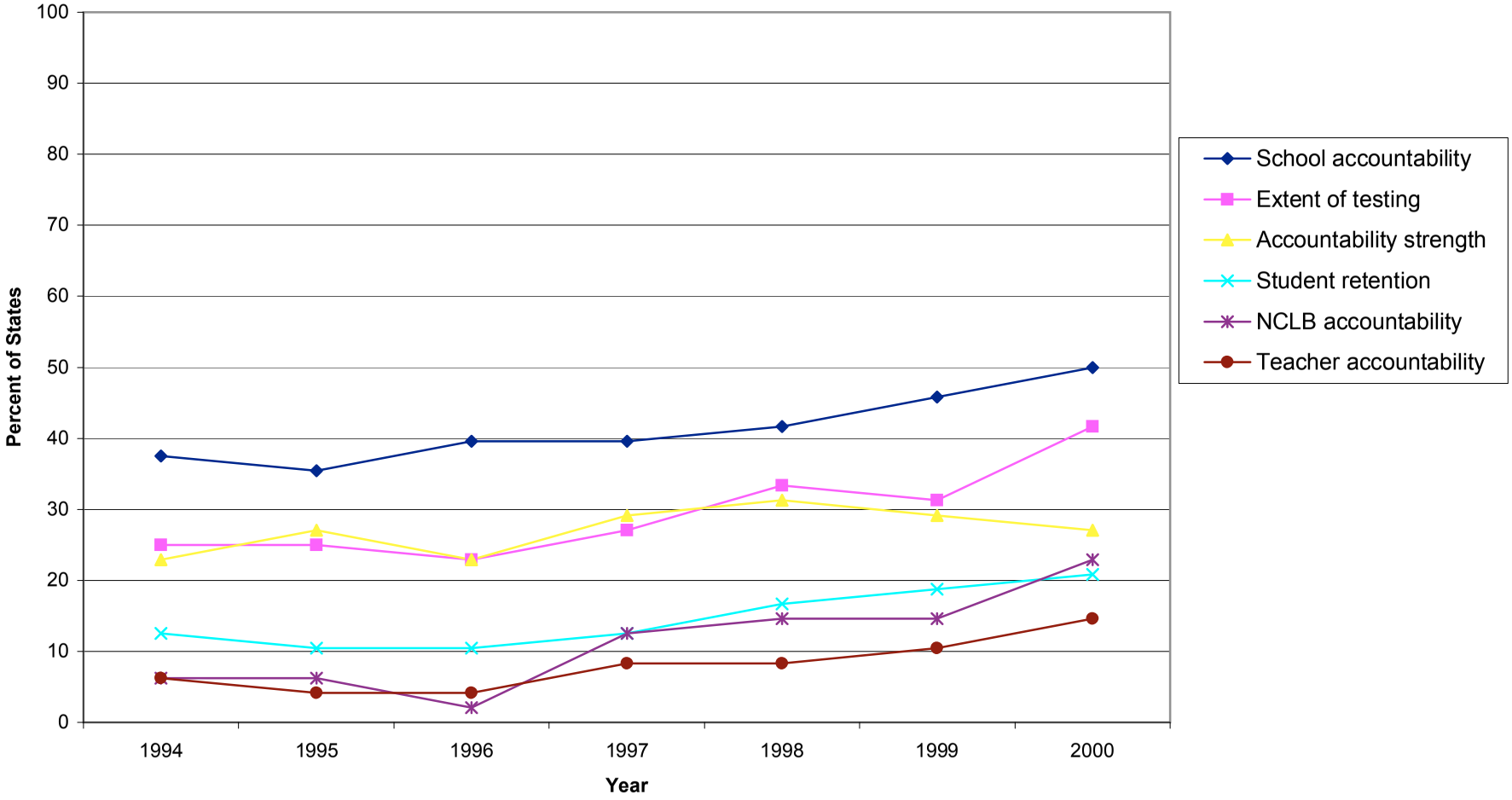
This measure classifies 3 states as NCLB states in 1994 and 11 states as NCLB states in 2000.

Table 1. Changes in State Accountability Policies: 1994-2000

	1993-1994		1999-2000		Change: 2000-1994	
	Mean	SD	Mean	SD	Mean	SD
Accountability Measures						
Extent of Testing	1.83	1.40	2.65	1.45	0.81	1.44
Student retention	0.13	0.33	0.19	0.39	0.06	0.43
Teacher accountability	0.04	0.20	0.13	0.33	0.08	0.35
School accountability	2.27	1.94	2.54	2.06	0.27	1.92
Accountability strength	2.05	1.47	2.39	1.46	0.33	1.35
NCLB accountability	0.06	0.24	0.23	0.42	0.17	0.38
State Covariates						
Per-pupil expenditures	6070.99	1450.52	6829.90	1349.22	758.91	378.13
Unemployment rate	5.38	1.22	3.74	1.00	-1.65	1.03
School enrollment (ln)	13.21	0.97	13.26	1.00	0.05	0.07
Class size legislation	0.15	0.36	0.40	0.49	0.25	0.44
Required teacher cert. tests	0.73	0.45	0.81	0.39	0.08	0.35

Note: Sample includes the 48 states for which we have data in both years. Florida did not respond to the 1999-2000 survey and data from Nebraska are missing for 1993-1994. See Appendix table 1 for descriptions of variables and sources.

Figure 1. Assessment and Accountability Trends, 1994-2000



Source: Authors' tabulations of data from the Council of Chief State School Officers (1995-2001). See text for more details.

Table 2. Correlations among Changes in State Accountability Policies and Covariates: 1994-2000

	1	2	3	4	5	6	7	8	9	10	11
Change in:											
1 Extent of Testing	1.00										
2 Student retention	0.09	1.00									
3 Teacher accountability	-0.10	0.25	1.00								
4 School accountability	0.08	-0.02	0.12	1.00							
5 Accountability strength	0.08	0.04	0.10	0.78	1.00						
6 NCLB accountability	-0.10	-0.20	0.38	0.17	0.10	1.00					
7 Per-pupil expenditures	-0.02	-0.14	0.12	-0.02	0.02	0.13	1.00				
8 Unemployment rate	0.21	0.07	-0.17	0.02	0.04	0.09	-0.03	1.00			
9 School enrollment	0.17	0.01	0.27	0.23	0.17	0.01	-0.27	-0.18	1.00		
10 Class size legislation	0.11	0.03	-0.14	0.17	0.22	0.00	0.06	0.17	-0.15	1.00	
11 Required teacher cert. tests	-0.01	0.11	-0.06	0.06	0.08	-0.11	-0.22	-0.10	0.10	0.14	1.00

Note: Sample includes the 48 states for which we have data in both years. Florida did not respond to the 1999-2000 survey and data from Nebraska are missing for 1993-1994. See Appendix table 1 and Data Appendix for more details on variables.

Table 3. Impact of State Accountability Policies on Time with Students

	Teachers in:						
	Public Schools	Non-Poor Schools	Poor Schools	Non-Poor White Schools	Non-Poor Minority Schools	Poor White Schools	Poor Minority Schools
Extent of Testing	0.026	-0.038	0.032	-0.026	-0.040	-0.019	0.053
	(0.031)	(0.047)	(0.039)	(0.068)	(0.089)	(0.088)	(0.037)
Student Retention	0.083	0.079	0.024	-0.005	0.431 +	0.084	-0.040
	(0.067)	(0.111)	(0.146)	(0.174)	(0.219)	(0.330)	(0.125)
Teacher Accountability	0.248 *	0.083	0.279 *	-0.067	0.530	0.345	0.197
	(0.096)	(0.128)	(0.116)	(0.182)	(0.329)	(0.226)	(0.146)
School Accountability	-0.004	-0.017	-0.011	-0.016	-0.088	0.054	-0.031
	(0.023)	(0.052)	(0.030)	(0.061)	(0.083)	(0.076)	(0.032)
Accountability Strength	-0.010	-0.019	-0.054	-0.035	-0.044	-0.031	-0.060
	(0.030)	(0.058)	(0.040)	(0.074)	(0.064)	(0.100)	(0.039)
NCLB Accountability	0.251 +	0.302 *	0.142	0.273	0.364	0.235	0.249
	(0.141)	(0.123)	(0.233)	(0.167)	(0.482)	(0.352)	(0.279)

Note: Columns reflect different samples. Rows show regression coefficients and robust standard errors for each accountability variable estimated in a model that includes no other accountability variables but does include teacher controls, school controls, district controls, a year dummy, state fixed effects, and state covariates. All models are weighted with the teacher weight. All standard errors are corrected for clustering at the state level. N in overall public school model for this outcome is 16,517. See text for details.

Table 4. Impact of State Accountability Policies on Preparation/Meeting Time

	Teachers in:											
	Public Schools		Non-Poor Schools		Poor Schools	Non-Poor White Schools		Non-Poor Minority Schools		Poor White Schools		Poor Minority Schools	
Extent of Testing	0.033		0.243	*	-0.186	0.114		0.669	**	-0.351		-0.148	
	(0.098)		(0.113)		(0.139)	(0.108)		(0.239)		(0.285)		(0.107)	
Student Retention	0.335		0.693	*	0.028	0.745	*	0.788		1.956	+	-0.435	
	(0.208)		(0.276)		(0.447)	(0.326)		(0.548)		(1.029)		(0.395)	
Teacher Accountability	0.260		0.724		-0.009	0.799		1.274		1.160		-0.340	
	(0.338)		(0.577)		(0.313)	(0.584)		(1.225)		(0.752)		(0.325)	
School Accountability	0.082		0.079		0.117	0.105		0.110		-0.010		0.155	+
	(0.064)		(0.099)		(0.108)	(0.092)		(0.284)		(0.257)		(0.079)	
Accountability Strength	0.140		0.155		0.231	0.087		0.478		0.351		0.210	+
	(0.086)		(0.137)		(0.145)	(0.147)		(0.297)		(0.379)		(0.124)	
NCLB Accountability	0.008		-0.264		0.129	-0.148		-0.850		-0.608		0.301	
	(0.355)		(0.511)		(0.415)	(0.546)		(1.468)		(0.595)		(0.517)	

Note: Columns reflect different samples. Rows show regression coefficients and robust standard errors for each accountability variable estimated in a model that includes no other accountability variables but does include teacher controls, school controls, district controls, a yea dummy, state fixed effects, and state covariates. All models are weighted with the teacher weight. All standard errors are corrected for clustering at the state level. N in overall public school model for this outcome is 16,517. See text for details.

Table 5. Impact of State Accountability Policies on "Content" Professional Development

	Teachers in:						
	Public Schools	Non-Poor Schools	Poor Schools	Non-Poor White Schools	Non-Poor Minority Schools	Poor White Schools	Poor Minority Schools
Extent of Testing	0.016	-0.023	0.040	-0.015	-0.055	0.194 *	-0.008
	(0.024)	(0.037)	(0.031)	(0.058)	(0.079)	(0.081)	(0.031)
Student Retention	-0.073	-0.130	-0.026	-0.002	-0.355	0.528	-0.112
	(0.068)	(0.126)	(0.151)	(0.181)	(0.219)	(0.345)	(0.185)
Teacher Accountability	-0.023	-0.283 **	0.151	-0.118	-0.456	-0.042	0.193
	(0.075)	(0.107)	(0.167)	(0.172)	(0.282)	(0.347)	(0.226)
School Accountability	-0.005	0.004	-0.027	0.000	0.048	0.046	-0.047
	(0.027)	(0.044)	(0.041)	(0.066)	(0.100)	(0.075)	(0.048)
Accountability Strength	-0.009	-0.007	-0.005	0.019	-0.004	0.110	-0.020
	(0.028)	(0.052)	(0.045)	(0.074)	(0.105)	(0.130)	(0.054)
NCLB Accountability	0.058	-0.234	0.253	-0.263	-0.123	-0.453 +	0.608 +
	(0.119)	(0.173)	(0.209)	(0.267)	(0.335)	(0.252)	(0.364)

Note: Columns reflect different samples. Rows show logit coefficients and robust standard errors for each accountability variable estimated in a model that includes no other accountability variables but does include teacher controls, school controls, district controls, a year dummy, state fixed effects, and state covariates. All models are weighted with the teacher weight. All standard errors are corrected for clustering at the state level. N in overall public school model for this outcome is 16,517. See text for details.

Table 6. Impact of State Accountability Policies on "Methods" Professional Development

	Teachers in:											
	Public Schools		Non-Poor Schools		Poor Schools		Non-Poor White Schools		Non-Poor Minority Schools		Poor White Schools		Poor Minority Schools
Extent of Testing	0.023		0.054		-0.025		-0.047		0.287	**	-0.025		-0.021
	(0.034)		(0.043)		(0.045)		(0.058)		(0.098)		(0.091)		(0.033)
Student Retention	0.121		0.076		0.097		0.086		0.042		0.465	+	0.070
	(0.076)		(0.113)		(0.106)		(0.190)		(0.247)		(0.259)		(0.111)
Teacher Accountability	0.283	+	0.333	+	0.213		0.010		0.502		0.274		0.167
	(0.168)		(0.191)		(0.223)		(0.253)		(0.402)		(0.364)		(0.187)
School Accountability	0.002		-0.018		0.011		-0.009		0.011		0.066		-0.003
	(0.040)		(0.055)		(0.041)		(0.069)		(0.089)		(0.050)		(0.035)
Accountability Strength	0.017		0.014		0.012		0.018		-0.020		0.051		0.016
	(0.039)		(0.060)		(0.045)		(0.083)		(0.099)		(0.083)		(0.033)
NCLB Accountability	-0.162		-0.181		-0.109		-0.524	*	0.637		-0.274		-0.040
	(0.175)		(0.237)		(0.241)		(0.229)		(0.566)		(0.310)		(0.300)

Note: Columns reflect different samples. Rows show logit coefficients and robust standard errors for each accountability variable estimated in a model that includes no other accountability variables but does include teacher controls, school controls, district controls, a year dummy, state fixed effects, and state covariates. All models are weighted with the teacher weight. All standard errors are corrected for clustering at the state level. N in overall public school model for this outcome is 16,517. See text for details.

Table 7. Impact of State Accountability Policies on "Assessment" Professional Development

	Teachers in:														
	Public Schools		Non-Poor Schools		Poor Schools		Non-Poor White Schools	Non-Poor Minority Schools		Poor White Schools		Poor Minority Schools			
Extent of Testing	0.138	*	0.057		0.193	***	-0.016	0.297	**	0.134	+	0.246	***		
	(0.058)		(0.072)		(0.046)		(0.068)	(0.090)		(0.078)		(0.042)			
Student Retention	0.180		0.275	*	-0.088		0.230	0.336		-0.192		-0.075			
	(0.163)		(0.135)		(0.231)		(0.156)	(0.243)		(0.288)		(0.245)			
Teacher Accountability	0.296		0.147		0.308		0.056	0.300		0.181		0.378			
	(0.292)		(0.312)		(0.282)		(0.293)	(0.429)		(0.334)		(0.311)			
School Accountability	0.016		0.061		-0.009		0.077	0.062		-0.046		0.022			
	(0.053)		(0.066)		(0.056)		(0.063)	(0.091)		(0.065)		(0.067)			
Accountability Strength	0.040		0.014		0.044		-0.009	0.126		-0.068		0.104			
	(0.071)		(0.081)		(0.074)		(0.088)	(0.102)		(0.090)		(0.083)			
NCLB Accountability	-0.070		-0.334		0.203		-0.228	-0.529		0.270		0.257			
	(0.246)		(0.310)		(0.255)		(0.294)	(0.420)		(0.351)		(0.311)			

Note: Columns reflect different samples. Rows show logit coefficients and robust standard errors for each accountability variable estimated in a model that includes no other accountability variables but does include teacher controls, school controls, district controls, a year dummy, state fixed effects, and state covariates. All models are weighted with the teacher weight. All standard errors are corrected for clustering at the state level. N in overall public school model for this outcome is 16,517. See text for details.

Table 8. Impact of State Accountability Policies on Class Size

	Teachers in:													
	Public Schools		Non-Poor Schools		Poor Schools		Non-Poor White Schools		Non-Poor Minority Schools		Poor White Schools		Poor Minority Schools	
Extent of Testing	-0.396		-0.499		-0.376		-0.278		-0.756	+	-0.043		-0.530	+
	(0.255)		(0.302)		(0.245)		(0.221)		(0.384)		(0.130)		(0.294)	
Student Retention	-1.304	*	-1.559	+	-1.084	+	-0.865		-2.387	*	-1.359	**	-1.004	
	(0.623)		(0.853)		(0.538)		(0.639)		(1.141)		(0.430)		(0.606)	
Teacher Accountability	-1.973	*	-2.562	*	-1.500	+	-1.997	*	-2.906	*	-0.369		-1.796	+
	(0.908)		(1.260)		(0.779)		(0.897)		(1.380)		(0.565)		(0.920)	
School Accountability	0.028		0.036		-0.062		-0.100		0.110		-0.099		-0.073	
	(0.119)		(0.135)		(0.137)		(0.129)		(0.205)		(0.084)		(0.181)	
Accountability Strength	-0.250		-0.284		-0.391	*	-0.187		-0.442		-0.184		-0.462	*
	(0.230)		(0.300)		(0.189)		(0.240)		(0.366)		(0.134)		(0.220)	
NCLB Accountability	0.054		-0.306		0.441		-0.273		-1.040		1.403	*	-0.226	
	(0.546)		(0.572)		(0.641)		(0.452)		(1.019)		(0.558)		(0.933)	

Note: Columns reflect different samples. Rows show regression coefficients and robust standard errors for each accountability variable estimated in a model that includes no other accountability variables but does include teacher controls, school controls, district controls, a year dummy, state fixed effects, and state covariates. All models are weighted with the teacher weight. All standard errors are corrected for clustering at the state level. N in overall public school model for this outcome is 13,925. See text for details.

Table 9. Impact of State Accountability Policies on Instructional Autonomy

	Teachers in:													
	Public Schools		Non-Poor Schools		Poor Schools		Non-Poor White Schools	Non-Poor Minority Schools	Poor White Schools	Poor Minority Schools				
Extent of Testing	-0.016		0.010		-0.042	*	0.024	-0.012	-0.020	-0.061	**			
	(0.020)		(0.021)		(0.018)		(0.033)	(0.029)	(0.025)	(0.019)				
Student Retention	-0.002		0.023		-0.013		-0.024	0.087	-0.067	-0.009				
	(0.051)		(0.066)		(0.074)		(0.107)	(0.115)	(0.099)	(0.089)				
Teacher Accountability	-0.006		0.123	**	-0.105		0.116	0.182	0.024	-0.138				
	(0.068)		(0.044)		(0.092)		(0.110)	(0.126)	(0.118)	(0.120)				
School Accountability	0.021		0.014		0.027		0.034	-0.043	0.050	0.023				
	(0.013)		(0.014)		(0.017)		(0.022)	(0.031)	(0.031)	(0.020)				
Accountability Strength	0.000		0.013		-0.002		0.007	0.014	0.043	-0.012				
	(0.017)		(0.019)		(0.024)		(0.030)	(0.032)	(0.036)	(0.029)				
NCLB Accountability	-0.128	+	-0.081		-0.148	+	-0.113	0.020	0.006	-0.240	*			
	(0.066)		(0.076)		(0.086)		(0.081)	(0.227)	(0.085)	(0.113)				

Note: Columns reflect different samples. Rows show regression coefficients and robust standard errors for each accountability variable estimated in a model that includes no other accountability variables but does include teacher controls, school controls, district controls, a year dummy, state fixed effects, and state covariates. All models are weighted with the teacher weight. All standard errors are corrected for clustering at the state level. N in overall public school model for this outcome is 16,517. See text for details.

Table 10. Impact of State Accountability Policies on Whether Teacher is New to the School

	Teachers in:											
	Public Schools	Non-Poor Schools	Poor Schools		Non-Poor White Schools		Non-Poor Minority Schools		Poor White Schools		Poor Minority Schools	
Extent of Testing	-0.015	0.006	0.005		-0.047		0.010		0.426	**	-0.120	+
	(0.038)	(0.103)	(0.062)		(0.134)		(0.107)		(0.124)		(0.068)	
Student Retention	0.048	0.031	0.099		0.045		-0.488	+	1.266		-0.114	
	(0.164)	(0.322)	(0.174)		(0.409)		(0.252)		(0.802)		(0.275)	
Teacher Accountability	-0.026	0.328	-0.327	+	0.947	*	-0.777	**	0.090		-0.447	+
	(0.154)	(0.311)	(0.189)		(0.457)		(0.290)		(0.762)		(0.248)	
School Accountability	-0.020	0.006	-0.051		0.079		-0.191		-0.081		-0.045	
	(0.031)	(0.059)	(0.041)		(0.080)		(0.135)		(0.078)		(0.055)	
Accountability Strength	0.068	0.180	+ 0.038		0.273	*	-0.141		0.261		-0.021	
	(0.043)	(0.093)	(0.058)		(0.127)		(0.113)		(0.161)		(0.085)	
NCLB Accountability	0.091	0.385	-0.169		0.876	**	-0.146		-0.999	+	-0.134	
	(0.195)	(0.275)	(0.212)		(0.313)		(0.431)		(0.527)		(0.310)	

Note: Columns reflect different samples. Rows show logit coefficients and robust standard errors for each accountability variable estimated in a model that includes no other accountability variables but does include teacher controls, school controls, district controls, a year dummy, state fixed effects, and state covariates. All models are weighted with the teacher weight. All standard errors are corrected for clustering at the state level. N in overall public school model for this outcome is 16,517. See text for details.

Table 11. Impact of State Accountability Policies on Teacher Experience

	Teachers in:						
	Public Schools	Non-Poor Schools	Poor Schools	Non-Poor White Schools	Non-Poor Minority Schools	Poor White Schools	Poor Minority Schools
Extent of Testing	-0.154	-0.162	-0.200	-0.142	-0.105	-0.323	-0.179
	(0.193)	(0.268)	(0.173)	(0.328)	(0.603)	(0.269)	(0.216)
Student Retention	0.080	-0.650	0.403	-1.105	1.435	1.471	0.189
	(0.637)	(0.905)	(0.433)	(1.178)	(1.068)	(1.472)	(0.505)
Teacher Accountability	-0.047	-0.502	0.302	-1.283	1.613	2.800 +	-0.480
	(0.714)	(1.088)	(0.812)	(1.939)	(1.019)	(1.476)	(0.968)
School Accountability	0.054	-0.017	0.188	-0.043	0.026	0.352 +	0.135
	(0.162)	(0.173)	(0.175)	(0.288)	(0.286)	(0.201)	(0.227)
Accountability Strength	0.034	-0.227	0.069	-0.372	0.275	-0.240	0.102
	(0.214)	(0.260)	(0.213)	(0.411)	(0.358)	(0.428)	(0.273)
NCLB Accountability	0.379	0.818	0.215	0.104	2.590 *	1.080	-0.297
	(0.701)	(0.788)	(0.865)	(1.129)	(1.225)	(1.235)	(1.105)

Note: Columns reflect different samples. Rows show regression coefficients and robust standard errors for each accountability variable estimated in a model that includes no other accountability variables but does include teacher controls, school controls, district controls, a year dummy, state fixed effects, and state covariates. All models are weighted with the teacher weight. All standard errors are corrected for clustering at the state level. N in overall public school model for this outcome is 16,517. See text for details.

Table 12. Impact of State Accountability Policies on Teachers' College Quality

	Teachers in:						
	Public Schools	Non-Poor Schools	Poor Schools	Non-Poor White Schools	Non-Poor Minority Schools	Poor White Schools	Poor Minority Schools
Extent of Testing	-0.834	-0.993	-0.612	-0.580	0.843	0.589	-1.666
	(1.760)	(2.757)	(1.903)	(3.166)	(3.551)	(1.891)	(2.520)
Student Retention	-8.485 *	-11.261	-10.048 *	-12.744	-2.460	-3.832	-11.507 *
	(3.923)	(6.728)	(4.418)	(7.942)	(9.639)	(8.029)	(4.446)
Teacher Accountability	-5.648	-6.866	-2.746	-6.985	1.605	17.843 **	-9.350
	(5.708)	(10.290)	(5.781)	(12.293)	(14.649)	(5.190)	(6.219)
School Accountability	-1.103	-3.664 +	1.421	-5.755 *	-0.529	2.289	0.517
	(1.052)	(1.878)	(1.204)	(2.396)	(3.248)	(2.251)	(1.704)
Accountability Strength	-4.287 ***	-8.332 ***	-0.445	-10.712 ***	-3.491	0.838	-1.263
	(1.139)	(2.163)	(1.721)	(2.667)	(3.529)	(2.559)	(2.250)
NCLB Accountability	6.649	4.781	10.199	2.035	0.605	14.570 *	3.699
	(5.476)	(8.203)	(6.410)	(7.597)	(19.595)	(6.879)	(10.165)

Note: Columns reflect different samples. Rows show regression coefficients and robust standard errors for each accountability variable estimated in a model that includes no other accountability variables but does include teacher controls, school controls, district controls, a year dummy, state fixed effects, and state covariates. All models are weighted with the teacher weight. All standard errors are corrected for clustering at the state level. N in overall public school model for this outcome is 15,294. See text for details.

Table 13. Impact of State Accountability Policies on Teacher Credentials

	Teachers in:														
	Public Schools		Non-Poor Schools		Poor Schools		Non-Poor White Schools		Non-Poor Minority Schools		Poor White Schools		Poor Minority Schools		
Extent of Testing	-0.097		-0.120		-0.083		-0.226	+	0.124		-0.100		-0.045		
	(0.079)		(0.102)		(0.053)		(0.132)		(0.229)		(0.189)		(0.072)		
Student Retention	0.592	**	0.297		0.850	***	0.416		0.847	*	-0.663		1.221		***
	(0.204)		(0.304)		(0.194)		(0.467)		(0.387)		(0.925)		(0.274)		
Teacher Accountability	0.332	*	0.524	*	0.309		0.849	*	0.482		1.349	*	0.315		
	(0.160)		(0.215)		(0.239)		(0.341)		(0.563)		(0.542)		(0.341)		
School Accountability	0.076	+	0.083		0.042		0.111		0.100		0.319	*	-0.041		
	(0.040)		(0.066)		(0.063)		(0.083)		(0.203)		(0.161)		(0.092)		
Accountability Strength	0.189	**	0.130		0.228	**	0.227	+	0.170		0.351		0.236		+
	(0.068)		(0.092)		(0.079)		(0.118)		(0.211)		(0.229)		(0.120)		
NCLB Accountability	-0.224		0.411		-0.743	*	0.575	+	0.080		0.243		-0.911		+
	(0.280)		(0.312)		(0.328)		(0.321)		(0.798)		(0.799)		(0.477)		

Note: Columns reflect different samples. Rows show logit coefficients and robust standard errors for each accountability variable estimated in a model that includes no other accountability variables but does include teacher controls, school controls, district controls, a year dummy, state fixed effects, and state covariates. All models are weighted with the teacher weight. All standard errors are corrected for clustering at the state level. N in overall public school model for this outcome is 16,517. See text for details.

Table 14. Impact of State Accountability Policies on Teacher Education

	Teachers in:													
	Public Schools		Non-Poor Schools		Poor Schools		Non-Poor White Schools		Non-Poor Minority Schools		Poor White Schools		Poor Minority Schools	
Extent of Testing	0.041		0.024		0.028		0.012		0.102		-0.020		0.019	
	(0.037)		(0.064)		(0.038)		(0.062)		(0.097)		(0.094)		(0.029)	
Student Retention	0.117	+	0.039		0.196	*	-0.116		0.672	*	0.142		0.165	
	(0.060)		(0.163)		(0.083)		(0.207)		(0.293)		(0.537)		(0.120)	
Teacher Accountability	0.222	**	0.230	+	0.186		0.121		0.769	***	0.664		-0.029	
	(0.083)		(0.132)		(0.113)		(0.195)		(0.209)		(0.458)		(0.111)	
School Accountability	-0.011		-0.016		-0.009		-0.005		-0.006		0.023		-0.017	
	(0.028)		(0.043)		(0.040)		(0.037)		(0.094)		(0.073)		(0.030)	
Accountability Strength	-0.009		-0.041		-0.006		-0.041		0.024		-0.184		0.018	
	(0.033)		(0.055)		(0.044)		(0.051)		(0.098)		(0.120)		(0.034)	
NCLB Accountability	0.025		-0.019		0.028		-0.110		0.429		0.429		-0.069	
	(0.126)		(0.217)		(0.215)		(0.270)		(0.336)		(0.309)		(0.254)	

Note: Columns reflect different samples. Rows show logit coefficients and robust standard errors for each accountability variable estimated in a model that includes no other accountability variables but does include teacher controls, school controls, district controls, a year dummy, state fixed effects, and state covariates. All models are weighted with the teacher weight. All standard errors are corrected for clustering at the state level. N in overall public school model for this outcome is 16,517. See text for details.

Table 15. Impact of Testing Specific Subjects on Time Spent Teaching Those Subjects

	Teachers in:														
	Public Schools		Non-Poor Schools		Poor Schools		Non-Poor White Schools		Non-Poor Minority Schools		Poor White Schools		Poor Minority Schools	
Percent of Classroom Time Spent on Reading															
Number of Grades w/ Reading Tests	1.156	*	0.984	+	1.014	+	0.428		2.434	*	0.528		1.211	*
	(0.511)		(0.562)		(0.545)		(0.509)		(1.191)		(0.754)		(0.550)	
Percent of Classroom Time Spent on Math															
Number of Grades w/ Math Tests	0.053		0.395	+	-0.138		0.941	**	-0.475		-0.081		-0.243	
	(0.185)		(0.234)		(0.275)		(0.348)		(0.340)		(0.447)		(0.311)	
Percent of Classroom Time Spent on Science															
Number of Grades w/ Science Tests	0.732	**	0.849	*	0.495	*	0.832	*	1.344		-0.131		0.659	**
	(0.246)		(0.378)		(0.188)		(0.406)		(0.874)		(0.452)		(0.211)	
Percent of Classroom Time Spent on Soc. Stud.															
Number of Grades w/ Soc. Stud. Tests	0.651	**	0.774	**	0.544	*	0.520	+	1.014		1.354	*	0.330	
	(0.205)		(0.269)		(0.219)		(0.286)		(0.627)		(0.522)		(0.257)	

Note: Columns reflect different samples. Rows show regression coefficients and robust standard errors for the sum of grades tested in a given subject, estimated in a model that also includes number of grades tested in the three other core subjects, teacher controls, school controls, district controls, a year dummy, state fixed effects, and state covariates. All models are weighted with the teacher weight. All standard errors are corrected for clustering at the state level. N in overall public school model for all outcomes in this table is 11,051. See text for details.

Table 16. Resource Allocations to Tested Grades, by Strength of State Accountability Policy and School Poverty

	Teachers in:													
	Public Schools				Non-Poor Public Schools				Poor Public Schools					
	Overall	Weak Account.	Strong Account.	Int. Stat. Sig?	Overall	Weak Account.	Strong Account.	Int. Stat. Sig?	Overall	Weak Account.	Strong Account.	Int. Stat. Sig?
Masters Degree														
Teach in tested grade	0.181 *	0.190 **	0.168	no	0.139	0.100	0.287	no	0.270 **	0.324 *	0.113	no
	(0.073)	(0.066)	(0.141)		(0.099)	(0.081)	(0.310)		(0.101)	(0.126)	(0.184)	
Teaching Experience (Yrs)														
Teach in tested grade	0.392	0.055	1.282 +	no	0.384	0.321	0.216	no	0.637	-0.084	2.437 *	yes
	(0.345)	(0.355)	(0.711)		(0.425)	(0.456)	(1.535)		(0.529)	(0.554)	(0.920)	
Credential														
Teach in tested grade	-0.049	-0.167	0.323	no	-0.153	-0.081	-0.305	no	0.018	-0.279	0.869 *	yes
	(0.171)	(0.148)	(0.319)		(0.159)	(0.217)	(0.621)		(0.215)	(0.196)	(0.382)	
College Quality														
Teach in tested grade	-0.536	-0.613	5.557	no	-2.876	0.409	-1.569	no	-2.399	-7.971 +	9.150	yes
	(2.967)	(3.407)	(6.625)		(3.908)	(4.275)	(7.237)		(4.259)	(4.046)	(8.888)	
Class Size														
Teach in tested grade	0.140	0.338 *	-0.653	yes	0.364	0.355	0.048	no	-0.025	0.425 +	-1.299	yes
	(0.200)	(0.161)	(0.544)		(0.225)	(0.262)	(0.573)		(0.294)	(0.241)	(0.852)	

Note: Columns reflect different samples. "Weak accountability" states are defined as those with an Accountability Strength Index of less than 4. "Strong accountability" states are those with Accountability Strength Index of 4 or 5. See appendix table 1 for more details on the construction of the Accountability Strength Index. Rows show regression or logit coefficients and robust standard errors for a dummy variable coded 1 if a teacher teaches in a tested grade and 0 otherwise. These coefficients are estimated from models that also include binary variables indicating whether or not each grade has a state mandated test, binary variables for the grade each teacher teaches, as well as the other teacher, school, district, and state controls. The "Int. Stat. Sig?" column indicates whether the interaction between "strong state" and "teach in tested grade" is statistically significant at least at the .10 level. All models are weighted with the teacher weight. All standard errors are corrected for clustering at the state level. Models that include state fixed effects produce very similar results.

Appendix Table 1. Description of Accountability Measures and State Covariates

Variable	Description
Testing	
Grades tested	Sum of grades 1-6 requiring at least one mandatory test in reading, math, science, or history
Consequences for Students, Teachers or Schools	
Student retention	1 if state retains students based on test scores; 0 otherwise
Teacher accountability	1 if state has any of the following teacher accountability policies: awards or recognition, monetary rewards, monetary penalties, evaluation or certification
School accountability	Sum of the following school accountability policies: performance reporting, warning, probation, monetary penalties, dissolution, takeover, recognition, monetary awards, exemption from regulations, accreditation
Overall Accountability	
Accountability strength	Replication of Carnoy and Loeb (2002) Index: Index ranges from 0-5 where 0=no state level testing; .5=state level testing, no performance reporting; 1=state level testing, performance reporting, no school accountability or exit exams; 1.5=state level testing, performance reporting, weak accountability, no exit exams; 2=state level testing, performance reporting, moderate accountability or exit exam; 2.5= state level testing, performance reporting, weak accountability, exit exam; 3=state level testing, performance reporting, moderate accountability, exit exam; 3.5=state level testing, performance reporting, moderate-strong accountability, no exit exam; 4=state level testing, performance reporting, strong accountability, no exit exam; 4.5=state level testing, performance reporting, moderate-strong accountability, exit exam; 5=state level testing, performance reporting, strong accountability, exit exam.
NCLB accountability	Dummy variable approximation of NCLB legislation, coded 1 if the state has mandatory reading and math tests in at least two elementary grades (out of grades 1-6), disaggregates test scores based on race and SES, and has at least one of the following test-based consequences: warning, probation, monetary penalty, school take over, or school dissolution.
Labor Market Characteristics	
Per-pupil expenditures	Current expenditures per pupil in fall enrollment in public elementary and secondary schools, by state, in 2000 dollars
Unemployment rate	Seasonally adjusted October unemployment estimate, by state
School enrollment	Natural log of public school enrollment, by state
Other State Education Policies	
Class size legislation	1 if state had implemented mandatory class size legislation in any grade at any level; 0 otherwise
Required teacher cert. tests	1 if state legislation required teachers to take either a subject-specific or a general test to obtain a regular credential; 0 otherwise

Note: All accountability data are from the State Student Assessments Annual Survey (Council of Chief State School Officers 2001; Council of Chief State School Officers/North Central Regional Education Laboratory 1995) for the 1993-1994 and 1999-2000 school years. See text for more details. Per-pupil expenditures and school enrollments are from U.S. Department of Education (2003). Unemployment rates come from the Bureau of Labor Statistics (1993-2000). Class size legislation data are from Pate-Bane (2004). Teacher certification testing data are from the American Association of Colleges for Teacher Education (1993) and the National Association of State Directors of Teacher Education and Certification (1999).

Appendix Table 2. Description of Teacher Outcomes

Variable	Description
Time on School-Related Activities	
Time with students outside normal school hours	Teacher reported. Hours per week. Teachers reporting 15 or more hours top-coded at 15 (approximately 3% of teachers); Question text in 1994: "During your most recent full week, how many hours did you spend AFTER school, BEFORE school, and ON THE WEEKEND on each of the following types of activities? School-related activities involving student interaction (e.g., coaching, field trips, tutoring, transporting students)?" The question text differs only slightly (and non-substantively) in 2000.
Preparation/ meeting time outside normal school hours	Teacher reported. Hours per week. Teachers reporting 30 or more hours top coded at 30 (less than 1% of teachers); Question text in 1994: "During your most recent full week, how many hours did you spend AFTER school, BEFORE school, and ON THE WEEKEND on each of the following types of activities? Other school-related activities (e.g., preparation, grading papers, parent conferences, attending meetings)?" The question text differs only slightly (and non-substantively) in 2000.
Professional Development	
"Content" professional development	Teacher reported. Nine or more hours spent in "content" professional development over past year (1=9 or more hours; 0 otherwise). Question text in 1994: "Since the end of the last school year, have you participated in any in-service or professional development programs which focused on the following topics? In-depth study in your subject field." Question text in 2000: "In the past 12 months, have you participated in any professional development activities that focused on in-depth study of the content in your MAIN teaching assignment field?" Note that categories for reporting hours other than 8 hours or less differ across years so we cannot measure hours in other categories.
"Methods" professional development	Teacher reported. Nine or more hours spent in "methods of teaching" professional development over past year (1=9 or more hours; 0 otherwise). Question text in 1994: "Since the end of the last school year, have you participated in any in-service or professional development programs which focused on the following topics? Methods of teaching your subject field." Question text in 2000: "In the past 12 months, have you participated in any professional development activities that focused on methods of teaching?" See caveats above.
"Assessment" professional development	Teacher reported. Nine or more hours spent in professional development on "assessment" over past year (1=9 or more hours; 0 otherwise). Question text in 1994: "Since the end of the last school year, have you participated in any in-service or professional development programs which focused on the following topics? Student assessment (e.g., methods of testing, evaluation, performance assessment). Question text in 2000: "In the past 12 months, have you participated in any professional development activities that focused on student assessment, such as methods of testing, evaluation, performance assessment, etc?" See caveats above.
Class Size	
Class size	Teacher reported. Total number of students enrolled in their class in the last full week of school for teachers teaching self-contained classes. Average number of students enrolled in one class period in the last full week of school for teachers teaching departmentalized classes. Restricted to teachers teaching self-contained and departmentalized classes. Teachers reporting classes with 65 or more students, or with fewer than 5 students, were excluded (less than 1% of the sample).

Appendix Table 2 Continued. Description of Teacher Outcomes

Perceived Instructional Autonomy	
Instructional autonomy	Average of four standardized, teacher-reported items about their perceptions of control over content and teaching methods. Question text for 1994: "At this school, how much control do you feel you have IN YOUR CLASSROOM over each of the following areas of your planning and teaching? Selecting textbooks and other instructional materials; Selecting content, topics, and skills to be taught; Selecting teaching techniques?" "At this school, how much actual influence do you think teachers have over school policy in each of the following areas? Establishing curriculum." The question text differs only slightly (and non-substantively) in 2000. However, the response set differs across years, with the 1994 SASS providing a scale from 0-5 and the 2000 SASS providing a scale from 1-5. Thus, we standardized the items and created the standardized index within years before pooling the data across years. Alpha=0.746
Teacher Stability	
Teacher is New to Current School	Teacher reported. Whether taught at same school last year. 1 if teacher is new to the school; 0 otherwise.
Teacher Quality	
Experience	Total years. Sum of years taught at public or private schools (full or part-time).
College quality	Mean SAT of undergraduate institution. Data from the Higher Education Research Institute (HERI) selectivity file, which is based on public reports of institutions' mean SAT and/or ACT scores. We inferred year of college entrance from teacher-reported year of college graduation. HERI data are available for 1973, 1977, 1982, and 1999. We attached the HERI data from the year closest to the year of college entrance but excluded the 1999 selectivity data because the 1999 scores reflect recentering of the SAT in 1995.
Credentials	Teacher reported. 1 if has regular, standard, or advanced teaching certificate in any state; 0 otherwise. Response options varied slightly between years. In 1994, respondents could report that they had a "certificate for persons who have completed an alternative program," but that option was not available in 2000. We coded respondents in that category as being certified.
Education	Teacher reported. 1 if has MA or more; 0 otherwise (less than 1% of sample has less than a BA)
Time Allocated to Subjects	
Percent time on reading	Teacher reported. Number of hours divided by the total number of hours for reading, math, history, and science combined. Restricted to teachers teaching self-contained classes. Teachers reporting fewer than 5 hours on all subjects or more than 40 hours on all subjects are excluded (approximately 1% of sample). Teachers who report teaching only one subject are excluded (N=18). Question text in 1994 and 2000: "During your most recent FULL WEEK of teaching, approximately how many hours did you spend teaching each of these subjects at THIS school? English/Reading/Language arts"
Percent time on math	Teacher reported. Number of hours divided by the total number of hours for reading, math, history, and science combined. Restricted to teachers teaching self-contained classes. Teachers reporting fewer than 5 hours on all subjects or more than 40 hours on all subjects excluded (approximately 1% of sample). Teachers who report teaching only one subject are excluded (N=18). Question text in 1994 and 2000: "During your most recent FULL WEEK of teaching, approximately how many hours did you spend teaching each of these subjects at THIS school? Arithmetic/Mathematics"
Percent time on history	Teacher reported. Number of hours divided by the total number of hours for reading, math, history, and science combined. Restricted to teachers teaching self-contained classes. Teachers reporting fewer than 5 hours on all subjects or more than 40 hours on all subjects excluded (approximately 1% of sample). Teachers who report teaching only one subject are excluded (N=18). Question text in 1994 and 2000: "During your most recent FULL WEEK of teaching, approximately how many hours did you spend teaching each of these subjects at THIS school? Social studies/History"

Appendix Table 2 Continued. Description of Teacher Outcomes

Percent time on science	Teacher reported. Number of hours divided by the total number of hours for reading, math, history, and science combined. Restricted to teachers teaching self-contained classes. Teachers reporting fewer than 5 hours on all subjects or more than 40 hours on all subjects excluded (approximately 1% of sample). Teachers who report teaching only one subject are excluded (N=18). Question text in 1994 and 2000: "During your most recent FULL WEEK of teaching, approximately how many hours did you spend teaching each of these subjects at THIS school? Science"

Note: All data are from the Schools and Staffing Survey (1993-1994 and 1999-2000). See text for more details.

Appendix Table 3. States that Experienced Changes in their Accountability Policies between 1994 and 2000

State	Extent of Testing	Student Retention	Teacher Accountability	School Accountability	Accountability Strength	NCLB Accountability
AK	X					
AL	X			X	X	
AR	X					X
AZ	X					
CA	X	X	X	X	X	
CO	X				X	
CT			X	X	X	X
DE				X	X	
FL	NA	NA	NA	NA	NA	NA
GA	X		X	X		X
HI				X	X	
IA				X	X	
ID	X			X	X	
IL		X	X	X	X	
IN	X	X			X	
KS	X					
KY	X		X	X		
LA				X	X	
MA	X			X	X	
MD				X		X
ME				X	X	
MI				X	X	
MN	X				X	X
MO	X			X	X	
MS	X			X	X	
MT				X	X	
NC				X	X	
ND		X				
NE	NA	NA	NA	NA	NA	NA
NH	X					
NJ	X			X	X	
NM	X	X		X	X	X
NV				X	X	
NY	X	X		X	X	
OH		X			X	
OK				X	X	
OR					X	
PA				X	X	
RI						
SC		X		X	X	
SD	X			X	X	
TN	X		X	X	X	X
TX				X		
UT	X	X				
VA	X					
VT	X					
WA	X			X	X	
WI	X			X	X	
WV	X			X	X	X
WY	X			X	X	

Note: The Xs indicate the states that changed their accountability policies between 1994 and 2000 (according to the CCSSO data). NA indicates that our dataset lacks information on FL and NE.

Appendix Table 4. Descriptive Statistics for Teacher Sample

	Mean	SD	N
Dependent Variables			
Time with students	1.54	2.96	16,517
Preparation & meeting time	10.01	6.36	16,517
Content professional development	0.38	0.49	16,517
Methods professional development	0.41	0.49	16,517
Assessment professional development	0.33	0.47	16,517
Class size	22.79	4.86	13,925
Instructional autonomy	0.00	1.00	16,517
New to current school	0.11	0.31	16,517
Years of experience	15.18	9.81	16,517
College quality	938.06	97.30	15,294
Credentialed	0.93	0.26	16,517
Education (MA+)	0.43	0.50	16,517
% time on reading	48.12	11.47	11,051
% time on math	24.93	7.02	11,051
% time on history	13.98	7.12	11,051
% time on science	12.98	7.33	11,051
Teacher, School, and District Controls			
Male	0.11	0.31	16,517
Age	43.27	10.29	16,517
American Indian	0.01	0.09	16,517
Asian American	0.02	0.13	16,517
African American	0.08	0.27	16,517
Latino	0.05	0.21	16,517
Teach in grade 1	0.17	0.38	16,517
Teach in grade 2	0.16	0.37	16,517
Teach in grade 3	0.18	0.38	16,517
Teach in grade 4	0.15	0.36	16,517
Teach in grade 6	0.16	0.36	16,517
Elementary school	0.93	0.26	16,517
Urban	0.27	0.45	16,517
Rural	0.31	0.46	16,517
Ln of school size	6.21	0.50	16,517
% eligible for subsidized meals	43.46	29.47	15,959
% minority	33.90	32.75	16,517
Charter	0.00	0.06	16,517
Magnet	0.06	0.24	16,517
Ln of district size	9.08	1.74	14,443
% in district eligible for subsidized meals	40.19	25.87	14,249
% minority in district	33.22	29.80	14,443
1999-00 school year	0.54	0.50	16,517

Note: All descriptive statistics are weighted with the teacher weight. Sample is limited to full-time teachers in grades 1 through 6 who teach academic (or general elementary) subjects and work in regular or "special program" emphasis schools. Note that in addition to the variables shown above, statistical models also include quadratics for % subsidized meals and % minority, as well as a missing data indicator for the school level subsidized meal variable, district level subsidized meal variable, % minority in district, and district size. Statistical models for some outcomes also include age and a quadratic in age. Most statistical models also include state fixed effects. See text for more details.