DOES SCHOOL ACCOUNTABILITY LEAD
TO IMPROVED STUDENT PERFORMANCE?

Eric A. Hanushek
Margaret E. Raymond

Does School Accountability Lead to Improved Student Performance
Eric A. Hanushek and Margaret E. Raymond
NBER Working Paper No. 10591
June 2004
JEL No. H7, I2, J1

ABSTRACT

The leading school reform policy in the United States revolves around strong accountability of schools with consequences for performance. The federal government's involvement through the No Child Left Behind Act of 2001 reinforces the prior movement of many states toward policies based on measured student achievement. Analysis of state achievement growth as measured by the National Assessment of Educational progress shows that accountability systems introduced during the 1990s had a clear positive impact on student achievement. This single policy instrument did not, however, also lead to any narrowing in the black-white achievement gap (though it did narrow the Hispanic-white achievement gap). Moreover, the black-white gap appears to have been harmed over the decade by increasing minority concentrations in the schools. An additional issue surrounding stronger accountability has been a concern about unintended consequences related to such things as higher exclusion rates from testing, increased drop-out rates, and the like. Our analysis of special education placement rates, a frequently identified area of concern, does not show any responsiveness to the introduction of accountability systems.

Eric A. Hanushek
Hoover Institution
Stanford University
Stanford, CA 94305-6010
and NBER
hanushek@stanford.edu

Margaret E. Raymond
CREDO
Hoover Institution
Stanford University
Stanford, CA 94305
macke@stanford.edu

The cornerstone of Federal educational policy has been expansion of school accountability based on measured student test performance. While many states had already installed accountability systems by 2000, a central campaign theme of George W. Bush was to expand this to all states, something that became a reality with the *No Child Left Behind Act of 2001 (NCLB)*. The policy has been controversial for a variety of reasons, leading to assertions that it has distorted schools in undesirable ways, that is has led to gaming and unintended consequences, and that it has not even accomplished its objectives of improving student achievement. This paper provides evidence on the expected effects of *NCLB* not only on student performance but also on other potential consequences.

The landmark *NCLB* codified a developing policy view that standards, testing, and accountability were the path to improved performance. It is nonetheless not possible to investigate the impact of *NCLB* directly. First, and most importantly, the majority of states had already instituted some sort of accountability system by the time the federal law took effect. While only 12 states had accountability systems at the school level in 1996, 39 states did so by 2000. Thus, there is no ready comparison group that can indicate what might have happened without any law. Second, the law has many facets but there is no obvious way to identify and measure the different components that are coming into play at a common pace across the states.

Isolating the impact of state accountability policies is inherently difficult. Because accountability invariably applies to entire states at an instant in time, variation across schools within a state provides no information about the impacts of accountability, and it is necessary to rely on state-level variation in student outcomes. Yet, states differ not only in their accountability policies but also in a variety other ways involving both population characteristics and other school policies. If these are not accounted for, they are likely to contaminate the estimates of the states' accountability system.

Our approach uses information about state differences in mathematics and reading performance as identified by the National Assessment of Educational Progress (NAEP). We pursue a number of strategies designed to isolate the effects of school accountability on performance. First, we look at growth in performance between 4th and 8th grades to eliminate fixed differences in circumstances and policies of each state. Second, we include explicit measures for major categories of time varying inputs: parental education, school spending, and racial exposure in the schools. Third, we estimate the growth models with state fixed effects to eliminate any other policies that lead to trends up or down in student performance in each state. Finally, to identify differences by race or ethnicity, we disaggregate the state results for whites, blacks, and Hispanics.

We find that the introduction of accountability systems into a state tends to lead to larger achievement growth than would have occurred without accountability. The analysis, however, indicates that just reporting results has minimal impact on student performance and that the force of accountability comes from attaching consequences to school performance. This finding supports the contested provisions of NCLB that impose sanctions on failing schools.

Much of the explicit interest in accountability and the federal legislation, however, focuses on low achievers. And, given the generally lower achievement by minority groups, an implicit assumption is that accountability – as revealed through mandatory disaggregation of performance for racial and ethnic groups – will simultaneously close the large achievement racial/ethnic gaps along with improving all performance. When we look specifically at the performance of subgroups, we find that Hispanic students gain most from accountability while blacks gain least.

Since the widespread introduction of accountability, a parallel interest has been whether more rigorous and consequential accountability also leads to other, less desirable impacts. For example, does accountability lead to increased cheating, more classifications of students as special education, or undesirable narrowing of teaching? To address a subset of these issues, we

analyze the rate of placement into special education across states but find no evidence of reaction in this dimension.

The next section briefly highlights the various lines of research that motivate the approach to estimating the determinants of state differences in achievement. This is followed by a more formal development of the statistical specification and by a description of the primary data employed. Two section then report the results of accountability systems on achievement growth and on special education placement.

Relevant Strands of Literature

Any consideration of state accountability systems must recognize the multitude of potential influences on student outcomes. The secret is separating the influence of accountability from these other factors.

The vast production function literature on variations in student performance provides a general backdrop for the analysis of achievement. This literature, dating from the Coleman Report (Coleman (1966)) and still being developed today, suggests significant differences in student achievement based on both family background and on schools (Hanushek (2002)).[1] A variety of controversies exists, particularly about the impact of various school resources (see Hanushek (2003)), but without going into detail about these it is sufficient to conclude that there is a lack of consensus about any specific measures of schools that adequately capture the relevant factors determining student performance. Some similar ambiguities exist when considering the measurement of family influences, even if there is strong consensus that families are very important in determining achievement. This lack of consensus on the appropriate specification of the determinants of student achievement motivates the analytical approach described below.

---

[1] Much of this literature is reviewed elsewhere. Here we simply identify sources both of basic analysis and of extended bibliographies on the relevant issues.

Throughout the study of schools and achievement, considerable attention has gone to the distribution of outcomes, and especially racial aspects of schooling. As famously highlighted more than 50 years ago by *Brown v. Board of Education*, the racial composition of schools may be relevant to achievement. The Coleman Report itself was legislatively mandated in the Civil Rights Act of 1964, and spawned attention to the racial composition of schools (U.S. Commission on Civil Rights (1967)). While most of the subsequent analysis flowing from *Brown* has related directly to the desegregation of schools (e.g., Armor (1995), Rossell, Armor, and Walberg (2002)), recent attention has turned more to issues related to the composition of schools.

Separating the effects of the racial composition of schools from other factors is clearly difficult, in large part because measurement errors for other school and family factors are likely to be correlated with racial composition. The analysis of Hanushek, Kain, and Rivkin (2002) approaches this through a generalized peer analysis that controls for family, school, and neighborhood effects through exploiting the rich longitudinal data from stacked panel data on student performance in Texas. That analysis suggests that increased black concentration of schools has a detrimental effect on black achievement, although racial composition does not seem to affect either whites or Hispanics. This consideration is particularly important given recent concern that racial concentration in the schools has been rising. Partly because court supervision over school racial patterns is ending but more importantly because white attendance in large urban systems has decreased, minority concentration has grown throughout the 1990s (Orfield and Eaton (1996), Clotfelter (2004)).

Each of these influences is embedded within school systems across the states that are pursuing a variety of policy reforms. The difficulty is that these other reforms are neither well specified nor readily measured, leading to considerable difficulty in adequately differentiating the relevant components (Hanushek (2002)). Moreover, as we look forward to an analysis of state level data, we know the potential damage of missing key ingredients to performances is amplified with aggregate data (Hanushek, Rivkin, and Taylor (1996)).

The final strand of relevant literature pertains to accountability itself. Although a recent policy effort, policies related to accountability have already become quite controversial – rising to the level of front page stories in the *New York Times* (Winter (2002)),   Much of the work is very new and has not appeared journals yet. The available studies generally support the view that accountability has had a positive effect on student outcomes, although the limited observations introduce some uncertainty (Carnoy and Loeb (2002); Hanushek and Raymond (2003b); Jacob (2003); Peterson and West (2003)).[2]

A larger body of work has concentrated on whether or not accountability has produced gaming and subsequent unintended consequences. This available work, reviewed in Hanushek and Raymond (2003b), tends to suggest some immediate reactions to accountability in terms of focusing teaching on relevant subjects or even relevant students near performance cutoffs; of increased exclusions from tests; of explicit cheating on tests; and of like attempts to improve scores in ways other than improving student learning. Nonetheless, as we return to below, little analysis provides information on the longer run outcomes of this nature.

Strategies for Dealing with the Analytical Difficulties

Analyzing the effects of accountability on student performance is difficult. Because accountability systems are introduced across entire states, all local school districts in a state face a common incentive structure. Thus, the only possible variation comes from interstate differences in accountability, but, as noted above, states also differ in ways other than accountability and ways in which past research has not been very informative. The difficulty is that, with little progress having been made in describing explicitly the different policies, regulations, and incentives that might be important in determining student performance, statistical estimates of accountability will been to be biased.

---

[2] Some variation also comes from analytical methods; see Amrein and Berliner (2002) and the analysis in Raymond and Hanushek (2003).

Fundamental educational policy is made at the state level and involves a wide range of factors including financial structure, collective bargaining rules and laws, explicit regulations on educational processes, curricular specification, and so forth. The analytical complications are immediately apparent.

Consider a simple model of achievement such as:

$$(1) \qquad O_{st} = f(X_{st}, R_{st}, \rho_s)$$

where O is the level of student outcomes in state s at time t, X is a vector of family and nonschool inputs, R is a vector of school resources, and ρ captures the policies of the state.[3] It is not possible to understand the impact of newly introduced accountability systems without considering the range of other factors influencing achievement.

A linearized version of this model is simply:

$$(2) \qquad O_{st} = \beta_0 + \beta_X X_{st} + \beta_R R_{st} + (\rho_s + \varepsilon_{st})$$

where the β's are unknown parameters of the educational process.[4] If, however, ρ is not observed and the β's are estimated with just information on X and R, correlations with ρ obviously lead to bias in the estimation. When background factors (X) and/or school resources (R) are correlated with state policies (ρ), these variables will partially proxy for the other policies – leading to incorrect inferences about what would happen if just X or R changed.

Now consider just adding A, a measure of whether or not accountability affects incentives and thus student performance.

$$(3) \qquad O_{st} = \beta_0 + \beta_X X_{st} + \beta_R R_{st} + \gamma A_{st} + (\rho_s + \varepsilon_{st})$$

---

[3] It does not matter for this discussion that we begin with aggregate outcomes for a state instead of building up from the individual student level (where the outcomes are presumably generated). The more general situation is discussed and developed in Hanushek, Rivkin, and Taylor (1996). Where the aggregation is important, we discuss the implications.

[4] The linear form is not particularly crucial but simply makes the exposition easier. An alternative model where policies act as an efficiency parameter affecting the impact of resources is developed in Hanushek and Somers (2001). Within the limited data for this study, however, it is virtually impossible to distinguish between the alternative models. The results of estimating the alternative form, discussed below, are qualitatively very close to the included estimates.

The objective is to understand γ, but under almost all circumstances γ will also be biased by omission of relevant other state policies, through either their direct correlation with accountability or with the other inputs into achievement.

Moreover, Hanushek, Rivkin, and Taylor (1996) demonstrate that the bias in any estimation will generally increase with the level of aggregation in situations like this. Specifically, when the omitted variable is relevant at the state level, estimation of the model across states will have the most bias. Note that this does not say anything about the direction of any bias, only that aggregation worsens the bias. In the case of measures of school resources, all evidence indicates that there is an upward bias from omitting state policies (Hanushek (2003; Hanushek, Rivkin, and Taylor (1996)). It does not, however, give much indication of how any estimation of partial models of accountability would bias analyses of γ.

If, however, the relevant state policies other than accountability are constant over our observation period, a variety of estimation approaches becomes possible. In the simplest form, simply looking at outcome changes over time eliminates any state differences that are constant over the period t to t*:

$$(4) \qquad \mathop{\Delta}_{t,t^*} O_s = \beta_X \Delta X_s + \beta_R \Delta R_s + \gamma \Delta A_s + \Delta \varepsilon_s$$

The key element is that effects of accountability systems are identified from changes in accountability across states over the sample period. Specifically, if all states introduced new accountability systems at the same time, ΔA would be constant, and γ would not be separately identified. This estimation relies on the variation in introduction of accountability systems over the period during which student achievement gains are observed.

But states do a variety of things to try to improve their schools – not just relying on accountability (or the absence of accountability). In order to allow for other policies that are occurring over time, we add a state fixed effect (δ_s) to the estimation as in equation 5:

(5) $\quad \underset{t,t^*}{\Delta} O_s = \beta_X \Delta X_s + \beta_R \Delta R_s + \gamma \Delta A_s + \delta_s + \Delta \varepsilon_s$

Such a model can be estimated when there are multiple observations of achievement growth for each state. With multiple observations for states, achievement growth during periods of accountability can be contrasted with achievement growth when the state had no accountability.

This formulation provides much better control for other factors influencing performance growth, because the formulation effectively adds a trend in performance that is relevant for each state. The growth formulation itself incorporates any state differences in policies, student and family characteristics, or other things that exert a constant influence on states performance over the relevant observation period. Adding the state fixed effect to this permits states to have policies that lead to trend differences in their student performance. (And, of course, the other policies of each state may or may not be effective in raising achievement, and no presumption is made about how they influence achievement). Now estimates of the effects of accountability are identified and estimated entirely on the basis of the introduction of accountability systems within each state. In essence, the estimation relies on a state-specific prediction of performance gains and then considers how the addition of an accountability system affects outcomes.

One final issue is relevant for the estimation. The objective is to generalize about what would happen when accountability is introduced to all states. But, the analysis relies on observed student performance data, and the sample of students may not be representative of the entire population.

A school can respond to disappointing assessments in two ways. First, it can adjust teachers, curriculum, and programs in an attempt to improve the teaching that occurs. This is, however, a difficult long-run proposition, made even more difficult in schools with high rates of staff turnover. A second, shorter-run strategy may result: to become more selective about the student scores that are incorporated into the school scores. The second approach could

supplement or possibly replace the first. By weeding out students who are poor performers, the school score can appear to be improving even if nothing different is being done.

The formal version of this, selection bias through testing rules, can be considered simply by looking in more detail at equation 5. The estimation of the effects of accountability (γ) depends on ΔA being uncorrelated with $\Delta \varepsilon$. If in fact states purposefully select who will take the tests that enter into the performance calculation, this condition will be violated.

The main issue, which we explicitly consider below, is that individual states have some control over the exclusion of students for reason of language or special education. Two approaches are suggested. First, in the spirit of Heckman (1979), one can simply estimate:

$$(6) \qquad \underset{t,t*}{\Delta} O_s = \beta_X \Delta X_s + \beta_R \Delta R_s + \gamma \Delta A_s + \Delta p(t) + \delta_s + \Delta \varepsilon_s$$

where $\Delta p(t)$ is the observed change in probability of taking the test over the observation period.[5]

Second, it is possible to estimate directly the exclusion probabilities:

$$(7) \qquad p(t) = f(X, R, A)$$

This second approach, which we follow in a secondary analysis, provides direct information about the unintended consequences of accountability systems.

Our estimation of the direct effects of accountability relies on variants of Equation 6. The essential question throughout is whether the introduction of accountability into a state alters the achievement that would be expected due to parents, school characteristics, and other policies that have also been put in place. Below we return to the estimation of whether accountability also leads to changes in the tested population in addition to any potential impacts on student performance.

---

[5] Note that, if the probability of exclusion from the testing is constant, this term will drop out from the growth calculations. Only changes in test taking rates will be relevant.

Data on State Accountability

The primary assessment of student performance for our analysis is the National Assessment of Educational Progress. This testing, often referred to as the "Nation's report card," provides a consistent measure of student performance that allows comparisons of students across time and across states. The focus throughout the NAEP testing (which began in 1969) has been developing assessment information for a representative sample of students at different age and grade levels.

The estimation of accountability effects uses two elements of the NAEP testing information. First, since the introduction of state level testing in 1990, NAEP has tracked performance over time for participating states. This testing provides directly useful data for two tests (mathematics and reading). The sampling/testing design of NAEP is particularly helpful because it has a basic four- year testing cycle that involves testing fourth and eighth graders. Thus, for example, fourth grade tests in math in 1992 can be paired with eighth grade math tests in 1996. While not the same students, this approach allows tracking the same cohort in each state, and thus holds constant common experiences for the cohort. Two cohort observations for math growth (1992-96 and 1996-2000) and two for reading growth (1994-98 and 1998-2002) makes is possible to create a panel of achievement growth in each subject – thus permitting estimation that removes individual state fixed effects.[6]

Second, throughout this analysis we also disaggregate by race and ethnicity. The consistent performance data separated by whites, blacks, and Hispanics permits a direct investigation of relative performance gains. Note, however, that the availability of disaggregated data for blacks or Hispanics within a state depends upon having a sufficiently large population to

---

[6] Note that the analysis relies on state aggregates and not individual level scores. While NAEP provides some disaggregated data, the testing scheme does not permit analysis of individual level performance. Pooling the data presumes that other state factors equally affect growth in both math and reading.

support separate reporting of test information. Thus, there are fewer state observations of black and Hispanic achievement than of white achievement.

The sample of student performance for the estimation thus depends both on the availability of disaggregated achievement data and on participation of the state in testing during both of the relevant testing years (e.g., 8th grade math testing in 1996 and 4th grade math testing in 1992). The relevant testing and racial/ethnic breakdowns for each of the sample periods for the separate tests is shown in Appendix Table A1. A total of 348 observations of state gains on the tests is available.[7] This sample is somewhat more heavily weighted toward whites, with fewer black observations and even fewer Hispanic observations. Note, however, that there are more distinct states (42) than appear for any of the time period-test breakdowns; a varying group of states participates in each of the tested grades and subjects for the different years.

Measured attributes of state education inputs include three primary factors: demographics, school resources, and school racial and ethnic composition. The key demographic factor is the education of the adult population. While we have various measures of the education distribution, we concentrate on the percentage of the population 25 years old or more that has at least a high school education. We calculate this separately for each population subgroup and for the relevant years of testing.[8] Not surprisingly, there are significant differences in average attainment for each of the groups: whites, 82 percent; blacks, 74 percent; and Hispanics, 60 percent. Substantial differences in these aggregate patterns also arise across states.

School resources are measured by the average state expenditure per pupil in real terms over the relevant time period. This measure cumulates the spending over the growth period being studied (i.e., each relevant four year period on which achievement growth is defined) and varies by state and time but not by subgroup.

---

[7] Because of missing data on exclusions from testing, the analytical samples are reduced to 348 observations from the 351 state observations with matched 4th and 8th testing for specific cohorts.
[8] The analysis interpolates data from the decennial censuses in 1990 and 2000 for each state and race/ethnic group to get the appropriate annual data for each state. We use the percentage of high school or more adults at the midpoint for each testing period.

To investigate the impact of racial concentration and trends over time, we include summary data on the racial and ethnic composition across the schools in each state. Specifically, for whites, blacks, and Hispanics, we calculate exposure to minority students in each school of the state (using the Common Core of Data of the U.S. Department of Education). The exposure measure indicates the proportion of school mates who are minority for the average white, black, and Hispanic student in the state in each year. These exposure rates are again averaged over the relevant test growth periods. The pattern of concentration of minorities by school yields disparate results for the amount of minority exposure for each group. Whites attended schools that on average over the period have 16 percent minority students, while the comparable percentages for blacks and Hispanics are 48 and 38 percent, respectively.

Data on accountability come from a survey and analysis of all states by CREDO (Fletcher and Raymond (2002)). For each state, information was collected on when a state introduced an accountability system for schools. For these purposes, an accountability system was defined as publishing outcome information on standardized tests for each school along with providing a way to aggregate and interpret the school performance.[9] States were classified by whether or not they attached consequences to school performance or simply provided a public report. Additionally, data were also collected on when a state began disaggregating test information by subgroups of the population. Note that these accountability measures pertain just to accountability for schools and do not mix in accountability for students that may have been introduced at a different time.[10]

The estimation relies on the varying timing of introduction of accountability systems into the different states. Figure 1 displays the overall cumulative pattern of accountability across the states. The data are broken up into states that attach consequences to their systems and states that simply report on school achievement. To understand the estimation strategy better, the set of

---

[9] The survey further collected information on the method by which schools aggregated scores. The alternative approaches are discussed in Hanushek and Raymond (2003b).

[10] Carnoy and Loeb (2002) employ an index of intensity of accountability that covers both school and student accountability measures but do not consider differential times of introduction.

Figure 1. State Accountability over Time (with NAEP Testing Dates)
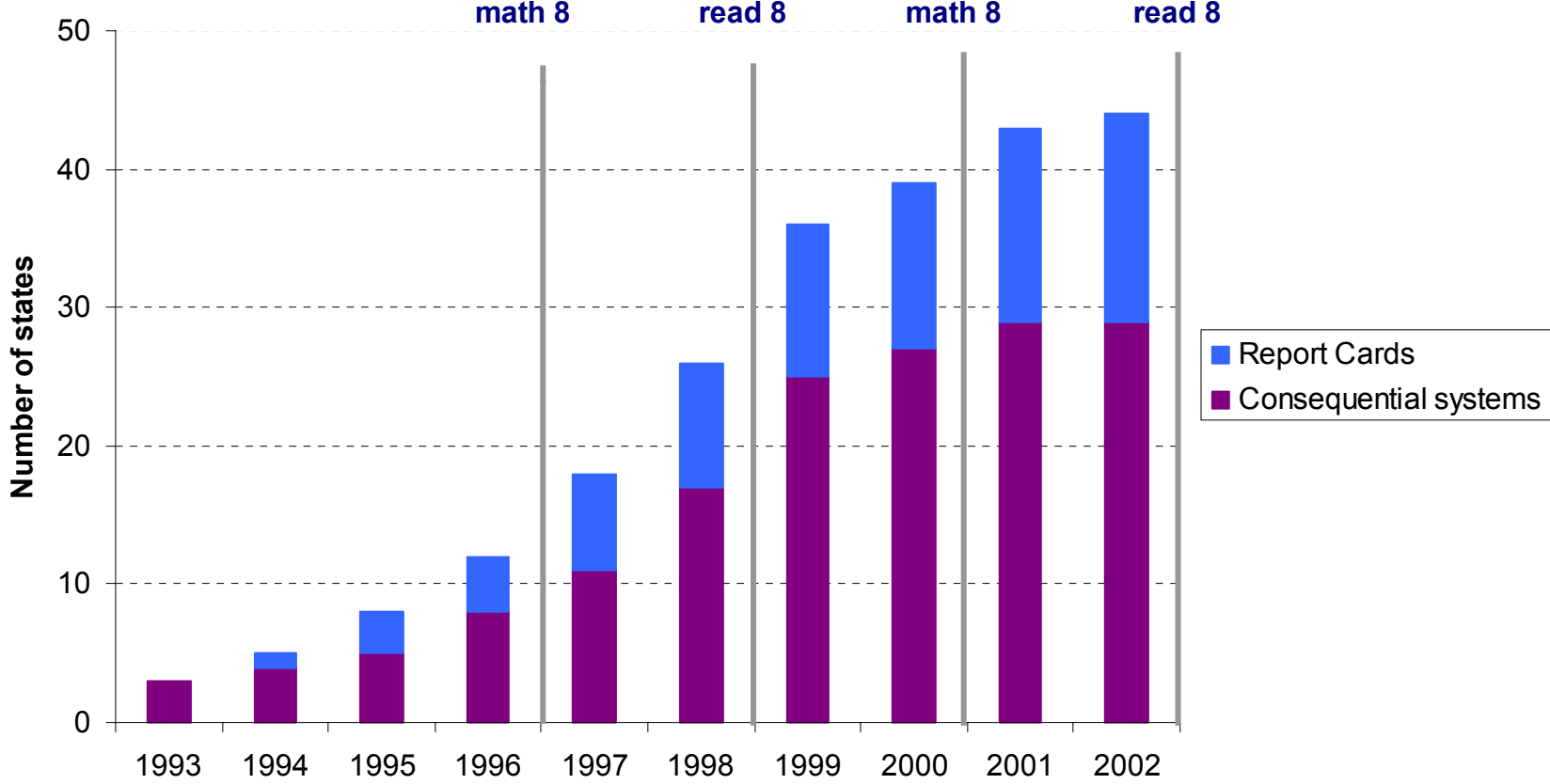
NAEP testing dates for 8th grade math and reading performance is superimposed on the pattern of accountability. The varied introduction across time and across the different testing periods permits disentangling the impact of accountability.

Finally, while the NAEP testing provides a consistent sample of performance for the states, some variations might arise simply because of differences in the test taking procedures in the states. Specifically, over the period a variety of students could be excluded from the testing because of special conditions including most importantly being identified as either a special education or Limited English Proficient student. The common presumption is that, since these students usually fall near the bottom of the achievement distribution, excluding them will artificially raise average scores of the tested population. Fortunately, NAEP provides information on test exclusions by test and year. Over the relevant time period, special education placements rose for the nation as a whole and for the separate states – going from 11.4 percent in 1990 to 13.3 percent in 2001. Over that same time period, test exclusions also rose, but by amounts that exceed the overall growth in the special education population. The pattern, however, differs dramatically by state with some states actually reducing the NAEP exclusion rate while others saw very large increases. These data on NAEP exclusions permit us to adjust for whether exclusion rates increased or decreased across separate testing periods in each state (which we do in a regression framework).

State Accountability and Student Achievement

We begin with the overall effect of state accountability on NAEP performance in the 8th grade across the three race/ethnic groups: whites, blacks, and Hispanics. The basic estimation pools the different time periods and tests but includes indicator variables for time period and test. The regression estimates predict 8th grade performance based on 4th grade performance of students in the state four years prior. Table 1 provides a summary of the key results for the performance for all students across the states that appear at least once in the testing. (Specific

Table 1. Determinants of State Growth in NAEP Performance (4th to 8th Grade)

	(1)	(2)	(3)
Consequential accountability	3.324	3.349	3.433
	(2.64)***	(2.66)***	(2.74)***
Report card system	0.734	0.633	0.760
	(0.44)	(0.38)	(0.46)
%pop(age 25+)≥high school	0.075	0.040	0.030
	(1.70)*	(0.81)	(0.60)
School spending, $/ADM	-0.001	-0.001	-0.001
	(0.53)	(0.39)	(0.43)
Change in exclusion rates	0.525	0.528	0.522
	(3.98)***	(4.01)***	(3.98)***
Black	-10.815	-10.137	-7.362
	(7.82)***	(7.02)***	(3.80)***
Hispanic	-9.707	-9.934	-10.062
	(6.92)***	(7.06)***	(6.53)***
Minority exposure rate		-4.875	
		(1.59)	
Minority exposure x white			1.783
			(0.35)
Minority exposure x Hispanic			-3.278
			(0.98)
Minority exposure x black			-8.468
			(2.35)**
Observations	348	348	348
Number of states	42	42	42
R-squared	0.94	0.94	0.94

* significant at 10%; ** significant at 5%; *** significant at 1%

Notes: All models estimated with state fixed effects. Models include NAEP 4th grade scores for reading and for math (lagged four years) and indicator variables for test and period. Absolute value of t statistics in parentheses

variable definitions along with descriptive statistics are found in Appendix Table A2). All estimates include individual state fixed effects, but the results remain qualitatively the same if simply estimated with random effect models across the states.

The accountability measure indicates the share of the period of study when a state had accountability (i.e., it ranges from 0.25 for accountability being in place for one year of the growth period for performance to 1.0 for accountability being in place for all four years. The data collection was designed to measure when the accountability system became effective, not when it was legislatively passed (Fletcher and Raymond (2002)).[11]

From Table 1 we find consistent evidence that introduction of state accountability had a positive impact on student math performance during the 1990s. Specifically, states that introduced consequential accountability systems early, tended to show more rapid gains in NAEP performance, holding other inputs and policies constant. This is consistent with our prior estimates of the effects of accountability for aggregations of all students in each state (Hanushek and Raymond (2003a, (2003b)).[12]

Interestingly, we find that report cards do not have a significant influence on performance. The point estimates, while positive, are not significantly different from zero. Thus, it seems important that policies include direct incentives rather than rely on indirect forces operating through just information.

The large differences in spending per pupil never influence scores. Consistent with past evidence on the impacts of resources (Hanushek (2003)), the pattern of NAEP scores across states

---

[11] Nonetheless, potential state-to-state differences in the phase in of accountability systems could effectively introduce measurement error into the accountability variable. An alternative approach is simply to measure whether or not the accountability system was in effective during the period, i.e., taking on the values 0 or 1. Pursuing this estimation yields qualitatively similar results, although a variety of the effects are not as precisely estimated (Hanushek and Raymond (2004)).

[12] The prior aggregate estimates, however, did not find a statistically different impact of report card systems versus consequential systems. In the estimates here, equality of consequences and reporting is rejected at the 10 percent level or better.

is not explained by spending. The impact of aggregate state spending is consistently small and statistically insignificant.

Test exclusions always have the expected effect on tests: more exclusions from a test for special education or language increase the average growth in test score. The introduction of exclusions, however, does not impact the estimates of accountability – chiefly because the introduction of accountability was not associated with large increases in exclusions. In fact, when states introduce accountability measures, they tend simultaneously to reduce on average their exclusion rates by a small amount.

In terms of general effects, the simplest model (column 1) indicates that higher parental education positively (and significantly at the 10 percent level) affects scores. However, as we refine the estimates in the more detailed specifications, any effects of parental education are imprecisely estimated and the effects are insignificant.

The remainder of Table 1 concentrates on the basic differences in performance by race. With disaggregation of performance by race (compared to aggregate state effects presented in Hanushek and Raymond (2003a, (2003b)), we see distinct differences in gains by blacks and Hispanics. These subgroups show growth that is 6-10 points lower than whites on NAEP between 4th and 8th grade. This spread overshadows the 3.5 point gain that came with accountability. This finding of lower black and Hispanic growth is particularly interesting in light of the narrowing of the achievement gap that occurred in the 1980s and the subsequent explanations for this improvement (Jencks and Phillips (1998)). The analysis of state details here that controls for state policy, family backgrounds, and testing exclusions shows a clear reversal of the prior decade.

The separate columns of Table 1 relate directly to the other major policy movement that had potential racial aspects: the influence of changing concentrations of minorities.[13] In these, we introduce measures of exposure rates of white, Hispanics, and blacks to minorities (Hispanics and blacks) across the schools in each state.[14] Total minority concentration has a negative but statistically insignificant impact in column 2. But the story changes when the impact of minority concentration is permitted to vary by subgroup in column 3, instead of being constrained to have the same effect. Higher minority concentrations have a statistically significant negative impact on blacks but do not significantly affect either whites or Hispanics. This finding is generally consistent with the analysis of racial composition in Texas by Hanushek, Kain, and Rivkin (2002). In that work, blacks were quite sensitive to school composition – specifically the proportion of blacks in the school negatively affected blacks, but whites and Hispanics were unaffected by student body composition.

To test the effect of intra-race influences, a further refinement of these models (not shown) considered black exposures to blacks instead of to minorities (blacks plus Hispanics). It is very difficult within these data to distinguish between the two alternative specifications. Using black exposure for blacks produced slight less precise estimates (t=2.0) but did not alter the other conclusions.

The models discussed so far (and represented in Table 1) consider the effects of accountability to be equivalent across the separate groups. For a variety of reasons, the effects may not be uniform. Thus, we estimate the same basic models but permit the effects of accountability to differ by race and ethnicity. Table 2 presents the results for these models. The first column is directly comparable to the previous table, but it now indicates distinct differences

---

[13] Earlier discussion of the lack of progress in closing the black-white gap in the 1990s speculated that changing patterns in school composition due to school desegregation patterns influenced the aggregate time series pattern of scores (Hanushek (2001)).

[14] These exposure rates are calculated on an individual school basis within each state. The variable for minority exposure in column 2 calculates exposure relative to each subgroup in the pooled sample; i.e., the variable is the exposure of white students to minorities for the white subset of the sample and the exposure of blacks to minorities for the black subset.

Table 2. Determinants of State Growth in NAEP Performance (4th to 8th Grade) with Disaggregation by Race and Ethnicity

	(1)	(2)
Consequential Accountability	3.446	3.590
	(2.53)***	(2.81)***
Report card system	0.827	0.820
	(0.52)	(0.51)
Consequential Accountability x Hispanic	3.084	
	(2.59)***	
Consequential Accountability x black	-2.081	
	(1.84)*	
Disaggregated x Hispanic		2.969
		(2.46)**
Disaggregated x black		-2.410
		(2.21)**
%pop(age 25+)≥high school	0.052	0.059
	(1.06)	(1.21)
School spending, $/ADM	-0.001	-0.001
	(0.44)	(0.40)
Change in exclusion rates	0.514	0.517
	(4.03)***	(4.04)***
Black	-6.234	-6.670
	(3.21)***	(3.50)***
Hispanic	-10.034	-9.694
	(6.58)***	(6.40)***
Minority exposure x white	1.288	0.924
	(0.26)	(0.19)
Minority exposure x Hispanic	-4.426	-4.604
	(1.33)	(1.39)
Minority exposure x black	-8.169	-7.827
	(2.32)**	(2.22)**
Observations	348	348
Number of states	42	42
R-squared	0.90	0.90

* significant at 10%; ** significant at 5%; *** significant at 1%
  Notes: All models estimated with state fixed effects. Models include NAEP 4th grade scores for reading and math (lagged four years) and indicator variables for test and period. Absolute value of t statistics in parentheses

by subgroup. Specifically, we see in column 1 that Hispanics seem significantly more affected than whites by having consequential accountability, while blacks appear less affected than whites.

When states introduce accountability systems they may or may not disaggregate the test results by racial group (as now required by *NCLB*). In the second columns we look at the differential impact of accountability for systems with subgroup disaggregation. When disaggregated accountability information is provided, Hispanics gain significantly and now show a greater reaction to accountability than whites. On the other hand, blacks do significantly worse than both whites and Hispanics.

In these more detailed models, we again find the strong indications that the racial composition of the schools is important for blacks. With the substantial negative impact of increased minority exposure, blacks appear hurt when attending less integrated schools.

It is useful to understand the magnitudes of both the accountability effects and the racial differences. Figure 2 displays the expected gains for states without consequential accountability and for states with consequential accountability. These gains are based on the disaggregations in column 2 of Table 2. As can be seen, the introduction of consequential accountability leads to improved growth in NAEP performance for each of the groups. To put the gains in perspective, on average, the white improvement is 0.22 standard deviations.[15]

At the same time, the subgroup patterns both in performance gains and in the impacts of accountability clearly differ. The differences are most easily seen in Figure 3, which translates the data into the black-white and Hispanic-white gaps in NAEP performance gains. The Hispanic-white gap in gains falls from 0.6 standard deviations to 0.42 standard deviations when consequential accountability (with disaggregated scores) is introduced in a state. But, the black-white gap in performance actually increases with accountability (from 0.43 to 0.57 standard deviations).
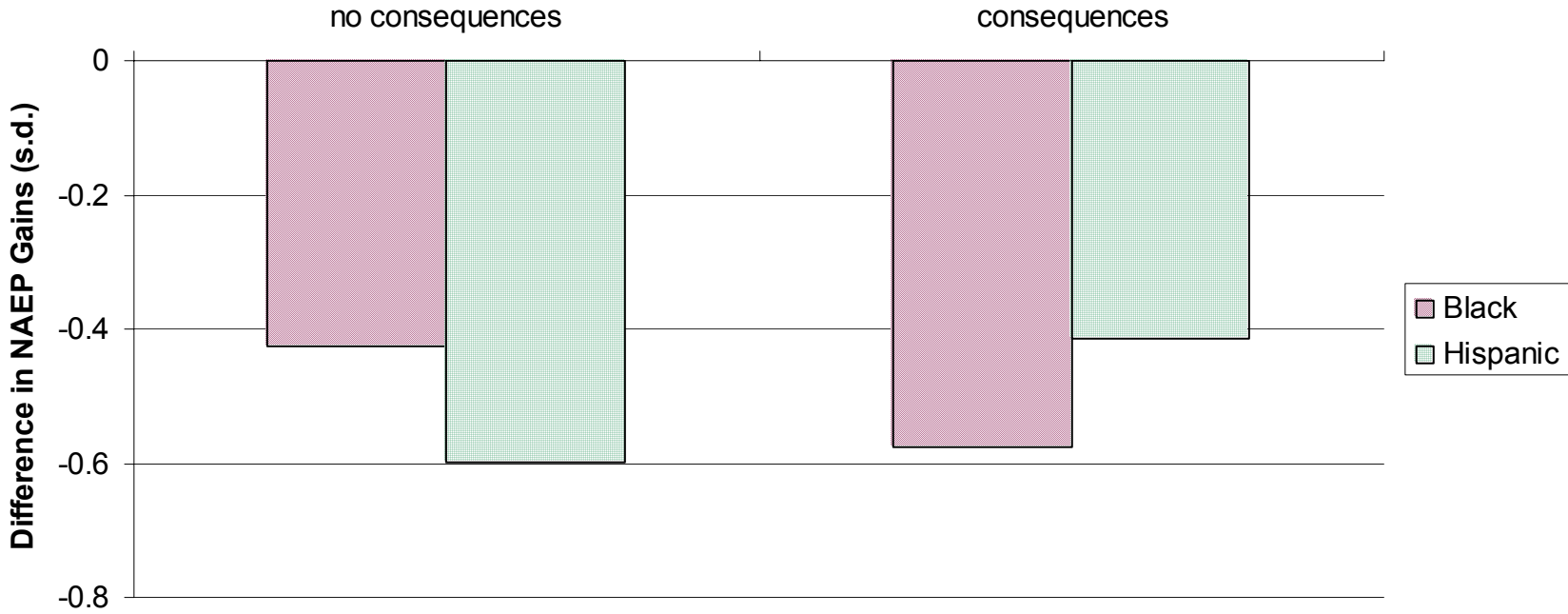
---

[15] These calculations rely on the standard deviation of average scores across states and subgroups for the eighth grade performance, which equals 16.2 scale score points.

Figure 2. Effect of Consequential Accountability on Achievement by Race/ethinicity
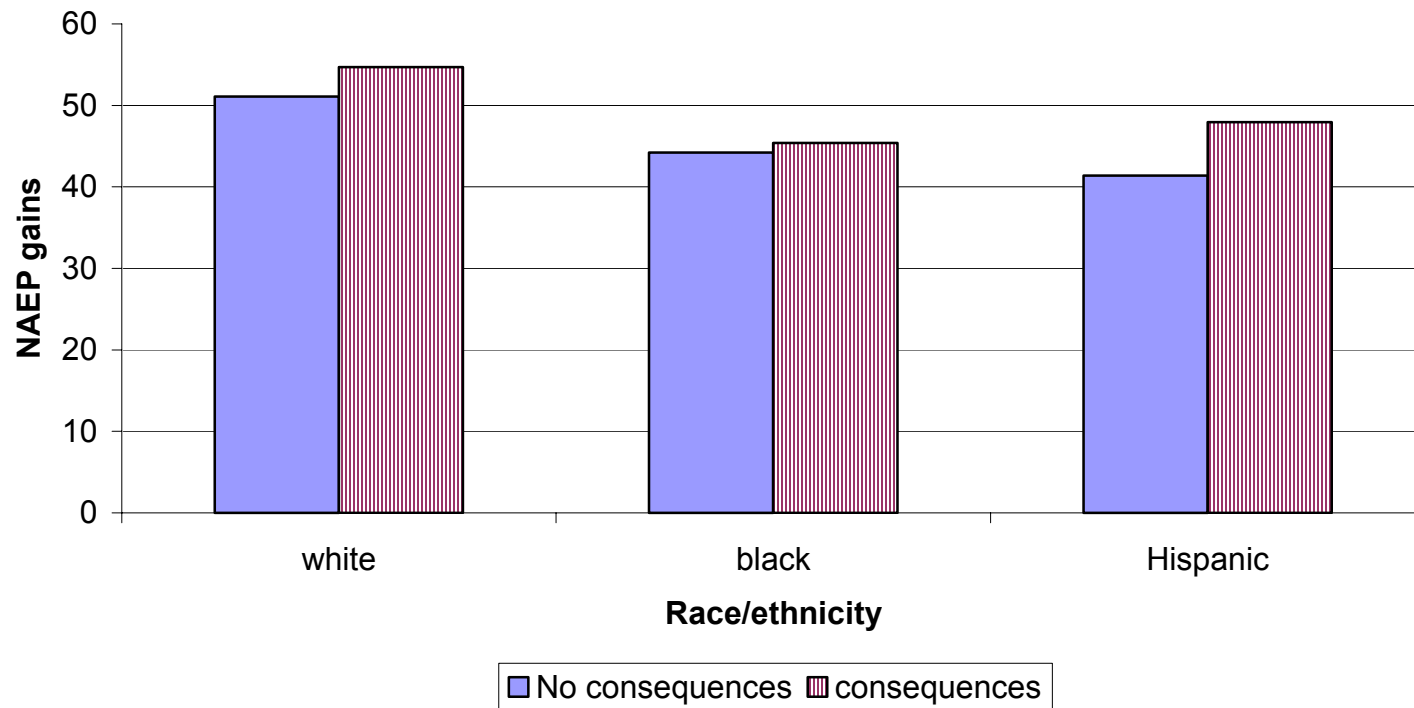
Figure 3. Racial/Ethnic Gaps
by Consequential Accountability Status
(NAEP gains relative to whites)

Accountability systems thus lead to overall improvements in student performance on NAEP mathematics and reading tests, but they do not uniformly meet the objective of also closing achievement gaps. This finding appears to be a simple demonstration of the well known principle that achieving multiple objectives with a single policy instrument is not generally feasible. We return to this below.

Other Consequences: Special-Education Placement

As many people have suggested, there is an immediate incentive in most existing accountability systems to exclude students who might be expected to have low achievement. A method often discussed is to place students into special education and thereby exclude them from testing and from subsequent inclusion in the accountability system. The previous analysis of the impact of accountability on achievement explicitly controlled for alterations in exclusions from NAEP testing, but the exclusion behavior is interesting in its own right.

Several studies have investigated whether schools appear to react to accountability through exclusions. Jacob (2002) considers the introduction of test-based accountability for Chicago public schools. He finds that the large increases in test scores after accountability went into effect were also accompanied by increases in special education placement and by increased grade retentions. Deere and Strayer (2001a, (2001b) and Cullen and Reback (2002) also find apparent increases in special education placement with the introduction of accountability in Texas. Prior work in Kentucky by Koretz and Barron (1998) suggested no strategic use of grade retentions. Haney (2000) suggests that both grade retention and increased dropouts were key to improvements in Texas tests, although both Carnoy, Loeb, and Smith (2001) and Toenjes and Dworkin (2002) seriously question this after reanalysis of the data.[16] Any grade retentions are, however, short run effects that do not provide lasting "accountability" value except if the

---

[16] Carnoy, Loeb, and Smith (2001) also find that at least in larger urban areas lower dropout rates are associated with higher student achievement.

placement is educationally valuable. Figlio and Getzler (2002) concentrate on special education placement after the introduction of a state accountability system in Florida. The most persuasive evidence is that placement rates increase relatively over time in grades that enter into the accountability system as opposed to those grades that do not.

In each case, the analysis considers changes that occur around the time of introduction of an accountability system. In fact, the key element of most of this research is using the change in accountability to identify the effects on special education placement rates and the like through finding breaks in the patterns of prior placement. Three things are important. First, there is very little relevant data for these analyses – breaks in trends, perhaps compared to trends of other schools (such as schools outside of Chicago and its accountability system). The validity of the interpretation depends crucially on whether or not other things are changing over time that could also affect the patterns of observed changes. Second, since later periods are always compared with earlier periods, there is concern about the general trend in special education placement that has been going on for two decades. Third, each of these analyses provides information just on the short run immediate effects. Since the incentives change over time, it is important to understand what happens as these systems continue.[17] Because of the recentness of introduction of accountability systems, little is know about the long run dynamics.

In order to test the importance of this incentive, we study the responsiveness of special-education placement rates to the introduction of an accountability system. We concentrate on the period 1995-2000, the period of large growth in state accountability systems as described in

---

[17] Hanushek and Raymond (2003b) consider the incentives that are set up by the design of different accountability systems. While the method of aggregating student performance and of judging change over time has an impact, the main conclusion is that incentives to exclude are generally largest in the first year of an accountability system and then decline if not reverse in subsequent years. This change in incentives results from the fact that exclusions in one year are generally built into the base for the next year, so that exclusions in any year must be maintained in subsequent year or they will lead to potential reductions in scores. Moreover, getting added gains from exclusions over time requires continual increases in the exclusion *rates*.

Figure 1. As with the achievement analysis, our basic strategy is to relate special-education

placement rates to accountability and other factors that might affect placement.

For this analysis, we do not rely on the sample of states participating in NAEP but instead

turn to annual information on all states. (Note, however, that it is not possible to disaggregate

these data by race and ethnicity, so we concentrate on overall state behavior). The basic

modeling considers special education placement rates across all 50 states plus the District of

Columbia. All estimation includes state-specific fixed effects. The "standard approach," that

found in the existing literature, simply considers comparisons of placement rates before and after

the introduction of accountability systems or how placement rates differ with time since the

introduction of accountability systems. (The difference between consequential and report card

systems was never significant in the estimation, so the analysis relies entirely on the combination

of the two). The standard model is then compared with a simple analysis that allows for national

time trends in placement rates.

The standard approach results in Table 3 show that the introduction of an accountability

or report-card system is associated with roughly 1.5 percentage point higher special-education

placement rates in a state. These estimates are essentially generalizations of difference-in-

difference estimators that allow for comparisons across all of the states. The second column

indicates that the reaction to accountability occurs over time, with a 1.1 percentage point higher

placement rate with accountability or report cards, and with an increase of 0.4 percentage point

increase each year that the system is in place. Thus, the state estimates appear to confirm the

estimates from individual states and districts.

The final two columns that place special education placement into the national picture,

however, show a markedly different picture. The final columns introduce a time trend and its

square to allow for the strong and ubiquitous increases in special-education placement. Columns

3 and 4 show that both the effect of having a consequential or report card system and the effect of

Table 3. Effect of Accountability on Special-Education Placement Rate, 1995 through 2000

	Standard Approach		Allowance for Placement Trend	
Accountability or report card system	1.45 (10.1)	1.09 (7.9)	.11 (1.0)	.10 (.9)
Time in place		.38 (7.9)		-.02 (-.5)
Time trend			.86 (12.4)	.87 (14.4)
Time trend squared			-.08 (-6.3)	-.08 (-6.0)

Note: Estimation employs a panel of special education placement rates for all states and the District of Columbia over the period 1995-2000. Estimation includes a fixed effect for each state. The t-statistics appear below each estimate. Time trend =1 in 1995; =2 in 1996; etc.

how long such a system has been in effect have an insignificant impact on placement rates (in terms of magnitude and of statistical significance) once the overall trends are considered.

These estimates suggest caution in interpreting analyses of the gaming of accountability systems. If such gaming were generally important, it should show up in the national data—but it does not. Moreover, the national trends in special-education placement offer a ready explanation for the divergent results.

Some Conclusions

Considerable public attention has focused on school accountability. While many states were pursuing their own versions of accountability, the discussion was elevated to new heights during the 2000 Presidential campaign when George W. Bush made school accountability a centerpiece of his domestic policy platform. Indeed the first year of his presidency involved significant pressure on Congress to enact accountability legislation, which it did with the *No Child Left Behind Act of 2001*.

NCLB mandated that all states introduce accountability systems that included annual testing of all students in grades 3-8 by 2006 and disaggregated data on student performance for all schools. This complex law also considered the development of state performance goals along with a variety of sanctions if schools failed to meet those goals.

NCLB has yet to be fully implemented, thus precluding a direct analysis of it.[18] Nonetheless, because *NCLB* calls for each state to design its own system and because most states have keyed off of their existing systems, the analysis here of the impacts of state systems enacted prior to *NCLB* provides information about what can be expected with full implementation.

---

[18] Somewhat ironically, when implemented, *NCLB* essentially precludes analysis of further impacts of overall accountability systems, because it eliminates any comparison group of states without accountability systems. Since, however, individual states will still follow their own locally developed schemes, it will still be possible to contrast the impacts of alternative types of accountability systems and alternative rewards and sanctions.

The most important result is that accountability is important for students in the United States (and in a variety of other countries that are pushing for better performance measurement). Despite design flaws in the existing systems (Hanushek and Raymond (2003b)), we find that they have a positive impact on achievement. This significantly positive effect of accountability holds across the alternative specifications of the basic achievement model.

However, the impact holds just for states attaching consequences to performance. States that simply provide better information through report cards without attaching consequences to performance do not get significantly larger impacts over no accountability. Thus, the *NCLB* move toward adding consequences to accountability systems is supported by looking at the historic introduction of consequential accountability systems.

It is useful to put the detailed subgroup impacts into perspective. Accountability significantly increases state achievement gains, particularly for Hispanics. However, because both blacks and Hispanics generally show lower gains relative to whites on each of the tests, accountability by itself is insufficient to close the gap in learning.

We also find that the effect varies by subgroup, with Hispanics gaining most and blacks gaining least. Because whites gain more than blacks after accountability is introduced, the racial achievement gap actually widens with the introduction of accountability.

In addition to accountability, the analysis looks into other determinants of student performance. Most relevant for consideration of where we stand fifty years after *Brown v. Board of Education*, black students are hurt by greater minority concentration in the schools. This compositional effect has no significant influence on white or Hispanic scores, making the effects very similar to those found in Hanushek, Kain, and Rivkin (2002).

These findings, taken together, underscore the fact that there is no one answer that will lead to all of the improvements that we desire. The introduction of consequential accountability systems has a clearly beneficial impact on overall performance. But other forces are simultaneously pushing the distribution of performance – particularly as observed in the black-

white achievement gap – in less desirable ways. First, accountability as seen during the 1990s tended to help white achievement more than black achievement. Second, the observed movement toward higher minority concentrations in schools has a detrimental effect on black achievement, again pushing toward a wider distribution of achievement.

The finding of differential effects of accountability raises a clear policy dilemma. A prime reason for the U.S. federal government to require each state to develop a test based accountability system involved raising the achievement of all students, particularly those at the bottom. It has done that, but not at the same rate across groups. We conclude from this that additional policies are needed to deal with the multiple objectives. Again, as is frequently the case, a single policy cannot effectively work for two different objectives – raising overall student performance and providing more equal outcomes across groups.

The movement toward stronger accountability in schools has also suggested to many that there would be adverse consequences – more exclusions, higher dropout rates, a narrowing of the curriculum, and the like. While some existing research supports these presumptions, we conclude that the negative impacts are likely to be considerably overstated (Hanushek and Raymond (2003b)). Importantly, many of the adverse effects that involve "gaming" the system come from short run incentives that are unlikely to be strongly present over time. Moreover, our own analysis of special education placement rates indicates clearly that accountability has not had an overall impact through this form of exclusions.

Finally, while we have not dwelled on it, the pattern of currently available accountability systems is not particularly strong. A majority of the systems concentrates on overall achievement levels (with highly variable passing scores across states). Such systems do not generally provide clear signals about the value-added of schools. Instead they combine a variety of effects including those resulting from family background differences and neighborhood effects. As such, they cannot provide truly clear and strong incentives. Yet, in the face of the rather blunt incentives from existing systems, the introduction of an accountability systems leads to

achievement improvements on the order of 0.2 standard deviations. Such improvements, while not revolutionary, are notable when compared to the failure to find alternative reforms that yield such impacts on a broad and sustained basis. As accountability systems evolve, they are likely to have considerably stronger impacts if they can be moved in the direction of more precise incentives for individual schools.

Appendix Table A1. Number of states for analysis by race/ethnicity, test, and sample period

	White	Black	Hispanic	Total
Mathematics				
1992-1996	35	29	32	96
1996-2000	34	26	32	92
Reading				
1994-19998	32	27	16	75
1998-2002	34	29	22	85
Total	135	111	102	348

Appendix Table A2. Variable Definitions and Sample Descriptive Statistics: Means and Standard Deviations (in Parentheses) by Race/ethnic Group

		All	White	Hispanic	Black
Consequential Accountability	Proportion of period with school accountability system having consequences for the school; Fletcher and Raymond (2002)	0.39 (0.45)	0.38 (0.44)	0.35 (0.44)	0.44 (0.45)
Report card system	Proportion of period with report card system; Fletcher and Raymond (2002)	0.14 (0.31)	0.14 (0.31)	0.14 (0.31)	0.14 (0.31)
Disaggregated	Proportion of period with school accountability system disaggregated by race/ethnic subgroups			0.37 (0.43)	0.4 (0.45)
%pop(age 25+)≥high school	% of population age 25 and older with a high school degree or greater; interpolation for period of decennial census data by race/ethnicity between 1990 and 2000	71.2 (11.8)	81.7 (5.1)	58.2 (8.3)	70.5 (7.3)
School spending, \$/ADM	Average expenditure per pupil in average daily membership for growth period (2000 \$)	6109 (1354)	6005 (1273)	6202 (1431)	6149 (1383)
Change in exclusion rates	NAEP exclusion rates: difference in 8th grade and 4th grade lagged four years by test	-0.16 (2.9)	-0.17 (2.8)	-0.25 (2.9)	-0.11 (3.0)
Minority exposure	Average exposure rate to minorities (black + Hispanic) by school averaged across growth period years	0.38 (0.24)	0.16 (0.17)	0.45 (0.21)	0.57 (0.17)
NAEP8	Average scale score, NAEP 8th grade test	257.4 (16.2)	274.8 (8.5)	249.0 (9.6)	244.0 (6.4)
NAEP4	Average scale score, NAEP 4th grade test	207.6 (16.0)	224.7 (6.2)	200.4 (10.5)	193.5 (7.3)
Special education placement rate	State placement rate for special education (percent)	12.63 (2.0)			
Accountability (report card or consequence)	=1 if either consequential or report card system in place; =0 otherwise	0.493 (0.5)			
Time in place	Years since introduction of accountability system	1.046 (1.7)			

References

Amrein, Audrey L., and David C. Berliner. 2002. *The impact of high-stakes tests on student academic performance: An analysis of NAEP results in states with high-stakes tests and ACT, SAT, and AP test results in states with high school graduation exams*. Tempe, Arizona: Educational Policy Research Unit, College of Education, Arizona State University (December).

Armor, David J. 1995. *Forced justice: School desegregation and the law*. New York: Oxford University Press.

Carnoy, Martin, and Susanna Loeb. 2002. "Does external accountability affect student outcomes? A cross-state analysis." *Educational Evaluation and Policy Analysis* 24,no.4 (Winter):305-331.

Carnoy, Martin, Susanna Loeb, and Tiffany L. Smith. 2001. "Do higher state test scores in Texas make for better high school outcomes?" Research Report Series, RR-047, Consortium for Policy Research in Education (November).

Clotfelter, Charles T. 2004. *After Brown: The rise and retreat of school desegregation*. Princeton, NJ: Princeton University Press.

Coleman, James S. 1966. *Equality of Educational Opportunity*. Washington, D.C.: U.S. Dept. of Health, Education, and Welfare

Cullen, Julie B., and Randall Reback. 2002. "Tinkering toward accolades: School gaming under a performance based accountability system." Department of Economics, University of Michigan (mimeo)

Deere, Donald, and Wayne Strayer. 2001a. "Closing the gap: School incentives and minority test scores in Texas." Department of Economics, Texas A&M University (mimeo) (September).

———. 2001b. "Putting schools to the test: School accountability, incentives, and behavior." Working Paper 113, Private Enterprise Research Center, Texas A&M University (March 2001).

Figlio, David N., and Lawrence S. Getzler. 2002. "Accountability, ability and disability: Gaming the system?" National Bureau of Economic Research, W9307, (November).

Fletcher, Stephen H, and Margaret E. Raymond. 2002. "The future of California's academic performance index." CREDO, Hoover Instiution, Stanford University (April).

Haney, Walter. 2000. "The myth of the Texas miracle in education." *Education Policy Analysis Archives* 8,no.41 (August).

Hanushek, Eric A. 2001. "Black-white achievement differences and governmental interventions."
 American Economic Review 91,no.2 (May):24-28.

———. 2002. "Publicly provided education." In *Handbook of Public Economics*, edited by Alan
 J. Auerbach and Martin Feldstein. Amsterdam: Elsevier:2045-2141.

———. 2003. "The failure of input-based schooling policies." *Economic Journal* 113,no.485
 (February):F64-F98.

Hanushek, Eric A., John F. Kain, and Steve G. Rivkin. 2002. "New evidence about *Brown v.
 Board of Education*: The Complex Effects of School Racial Composition on
 Achievement." Working Paper 8741, National Bureau of Economic Research (January).

Hanushek, Eric A., and Margaret E. Raymond. 2003a. "Improving Educational Quality: How
 Best to Evaluate Our Schools?" In *Education in the 21st Century: Meeting the
 Challenges of a Changing World*, edited by Yolanda Kodrzycki. Boston, MA: Federal
 Reserve Bank of Boston:193-224.

———. 2003b. " Lessons about the Design of State Accountability Systems." In *No Child Left
 Behind? The Politics and Practice of Accountability*, edited by Paul E. Peterson and
 Martin R. West. Washington, DC: Brookings:127-151.

———. 2004. "The effect of school accountability systems on the level and distribution of
 student achievement." *Journal of the European Economic Association* 2,no.2-3 (April-
 May):406-415.

Hanushek, Eric A., Steven G. Rivkin, and Lori L. Taylor. 1996. "Aggregation and the estimated
 effects of school resources." *Review of Economics and Statistics* 78,no.4
 (November):611-627.

Hanushek, Eric A., and Julie A. Somers. 2001. "Schooling, inequality, and the impact of
 government." In *The causes and consequences of increasing inequality*, edited by Finis
 Welch. Chicago: University of Chicago Press:169-199.

Heckman, James J. 1979. "Sample selection bias as a specification error." *Econometrica* 47:153-
 161.

Jacob, Brian. 2003. "High stakes in Chicago: Did Chicago's rising test scores reflect genuine
 academic improvement?" *Education Next* 3,no.1 (Winter):66-72.

Jacob, Brian A. 2002. "Making the grade: The impact of test-based accountability in schools."
 Kennedy School of Government, Harvard University (mimeo) (April).

Jencks, Christopher, and Meredith Phillips, eds. 1998. *The black-white test score gap*.
 Washington, DC: Brookings.

Koretz, Daniel M., and Sheila I. Barron. 1998. *The validity of gains in scores on the Kentucky
 Instructional Results Information System (KIRIS)*. Santa Monica, CA: RAND
 Corporation.

Orfield, Gary, and Susan E. Eaton. 1996. *Dismantling Desegregation: The Quiet Reversal of Brown v. Board of Education*. New York, NY: The New Press.

Peterson, Paul E., and Martin R. West, eds. 2003. *No child left behind? The politics and practice of accountability*. Washington, DC: Brookings.

Raymond, Margaret E., and Eric A. Hanushek. 2003. "High-Stakes Research." *Education Next* 3,no.3 (Summer):48-55.

Rossell, Christine H., David J. Armor, and Herbert J. Walberg, eds. 2002. *School desegregation in the 21st century*. Westport, CT: Praeger.

Toenjes, Laurence A., and A. Gary Dworkin. 2002. "Are increasing test scores in Texas really a myth, or is Haney's myth a myth?" *Education Policy Analysis Archives* 10,no.17 (March).

U.S. Commission on Civil Rights. 1967. *Racial isolation in the public schools*. Washington, DC: Government Printing Office.

Winter, Greg. 2002. More schools rely on tests, but big study raises doubts. *New York Times*, December 28, 1.