Looking Beyond Test Score Gains: State Accountability's Effect on the Differential Black-White Levels of Education Attainment and Labor Market Outcomes

Kathleen N. Wong
Department of Economics
University of California, Irvine
knwong@uci.edu

April 24, 2008

Abstract

Many papers in the accountability literature have attempted to document whether accountability programs have played a role in producing higher test scores. However, given that test scores are not the appropriate measure to gauge accountability programs' success for a number of reasons, this paper seeks to determine whether they may have had larger impacts on more significant measures of student outcomes. In particular, the analysis will focus on students' levels of educational and labor market outcomes. Treatment effects are estimated for separately for white and black individuals, as blacks often have lower levels of education, employment and earnings relative to whites. The analysis allows one to learn if such state programs were successful in bridging the black-white achievement gap. While the achievement gap in test scores is of noteworthy concern, it is the socioeconomic achievement gap that policy makers would ultimately wish to narrow. The overall results suggest that accountability programs were successful in increasing the level of educational attainment and employment for blacks, but that these positive impacts do not have long-run impacts on earnings. Although increased high school graduation rates by blacks may have helped in narrowing part of the educational gap, the evidence indicates that accountability programs have not been able to successfully narrow the socioeconomic gap between whites and blacks.

1. Introduction

Starting in the early 1990's, states began to introduce school accountability programs, demonstrating the high priority being placed on education. Program rules varied across states, as states had authority to create their own incentive schemes, but the goals of all programs were similar. Accountability sought to hold schools responsible for achieving high levels of educational aptitude for all students, measured by student performance on standardized exams. Despite the states' successes in equalizing student access to quality schools and increasing test scores for minorities, an article in the *Los Angeles Times* evaluating the test score gap concludes that while students' test scores have improved, they still have "a long way to go, especially to close the gap that separates whites and Asians from their Latino or black peers" (Blume, 2007).

Many papers in the accountability literature have attempted to document whether accountability programs have played a role in producing higher test scores. However, given that school funding is at stake, schools have an incentive to cheat and game the system. This behavioral response leads average test scores to artificially appear higher and the black-white test score gap to appear lower than they would otherwise. Although the achievement gap in test scores is of noteworthy concern, it is the socioeconomic achievement gap that policy makers would ideally wish to narrow; whatever is learned about the effect of accountability programs on test scores still leaves room for doubt as to how the socioeconomic gap is affected. This paper looks to establish how accountability measures adopted by states have affected racial inequalities across education attainment and labor market outcomes. In particular, the analysis will focus on the influence accountability had on blacks, who often have lower levels of education,

employment and earnings relative to whites, to learn if state programs were successful in bridging the black-white achievement gap[1].

From a policy-making perspective, it is important to study the effect of accountability programs on educational attainment or labor market outcomes after the student has left the public school system, since the ultimate goal of ensuring students' progression through school is to best prepare them for their future endeavors. Recent trends in the education literature have begun to shift their focus away from test scores and more towards other economic outcomes, such that of high school graduation and wages (Card and Krueger, 1992; Betts, 2001; Cawley, Heckman and Vytlacil, 1999). If the increases in test scores are only artificial, in the sense that students are only better prepared to take standardized tests, then the goals of accountability programs have not been achieved. Rather, their success in achieving higher test scores (if that even occurs) is temporary and may not translate into more promising, long-run accomplishments.

The results from the analysis regarding accountability programs' effect on student outcomes indicate that accountability programs were successful in encouraging blacks to increase their completed years of schooling. High school graduation rates increased for black students by 2.6 percentage points on average, while college enrollment rates for black students were on average unaffected. In comparison, the high school graduation rates for white students did not increase, while the college enrollment rates for white students actually decreased slightly by 0.8 percentage points. Analysis concerning the effect of accountability programs on labor

[1]An analysis concerning how accountability programs affected Hispanics' education attainment and labor market outcomes would also be of importance, particularly since they represent such a large proportion of the minority population. However, confidence cannot be placed in the results of such an analysis, as large proportions of this ethnic group are immigrants to the US and may not have completed their schooling domestically. Since treatment by accountability programs occurs through enrollment in the US education system, it calls into question whether or not one can accurately determine if treatment was received. As a result, analysis in this paper will exclude the discussions regarding the Hispanic-white achievement gap and rather, focus on accountability's effect on the black-white achievement gap.

market outcomes also indicates that exposure to accountability treatment led black employment rates to increase by approximately 1 percentage point, but did not have an overall impact on other labor market outcomes such as log earnings or labor force participation when measured in hours. The log earnings for a subset of blacks did however decrease, but after controlling for the number of hours worked, further analysis finds that the log hourly wages for this sample in fact increased. However, despite the statistical significance of the results, the magnitude of such an increase in log hourly wages is economically insignificant. The overall results suggest that accountability programs were successful in increasing the level of educational attainment and employment for blacks, but that these positive impacts do not have long-run impacts. Although increased high school graduation rates by blacks may have helped in narrowing part of the educational gap, the evidence indicates that accountability programs have not been able to successfully narrow the black-white socioeconomic gap.

The next section of the paper will provide a brief summary of the history of accountability programs. Section 3 reviews the existing literature, summing up what has been learned about these state programs thus far and why test scores may not be the appropriate measure for gauging the effectiveness of accountability programs. Section 4 discusses the identification strategy and econometric models employed in this paper, while section 5 provides a description of the data and variables used. The results are presented in section 6, and followed by a discussion of the robustness checks that were carried out in section 7. The paper then wraps up with a summary and the conclusions that can be drawn from the results.

2. Accountability Programs

In the early 1990's, states began to recognize the importance of education, leading them to place more emphasis on student learning by holding schools responsible for ensuring that

students were being properly educated. Connecticut, North Carolina and Wisconsin became the first three states to adopt their own state accountability programs in 1993. From that time on, more and more states followed suit. While only 14 states had accountability programs in 1996, 46 did so by 2001, prior to the federal mandate signed by President Bush in 2002 titled the No Child Left Behind Act. Table 1 provides details on the year each state adopted their accountability program. Although numerous states implemented programs in 1998 and 1999, figure 1 plots the relatively smooth adoption trend over the years from 1993 to 2003.

Program rules and incentives varied across states because each had authority over the design of their own state programs, but they all had a single unifying goal: to hold schools responsible for ensuring that students were being properly educated by gaining proficiency in core subjects and to encourage yearly performance progress. States used the school's average test score changes from year to year as a measure to determine whether or not adequate yearly progress was accomplished. Depending on the state, schools faced a number of incentives which were meant to encourage schools to achieve the state's established levels of proficiency. Schools that met their yearly progress goals were rewarded with additional funding from the government, while those that failed to improve faced the threat of having their funding rescinded. Program designers hoped that either the funding incentive schemes or threats of funding withdrawal would result in higher student achievement.

3. Existing Literature

Thus far, there have been a number of papers that have analyzed the effects of accountability programs on student performance through test scores[2]. However, the literature

[2] See Jacob (2005), Carnoy & Loeb (2002), Figlio & Rouse (2006) and Hanushek & Raymond (2004) for a sample of discussions in the literature.

does not provide a general consensus as to whether or not accountability programs are truly responsible for student test score increases. Interestingly enough, there is also a large literature which also supports the claim that test scores are an inappropriate measure of accountability's success. Since financial resources were at stake for schools that were at-risk of failing to meet their goal, it created a distortion in the incentives faced by the schools. Many papers document how school administrators and teachers are able to manipulate test scores by participating in cheating and gaming activities. All of their efforts were aimed at altering the pool of students who are tested, with the hopes of producing artificially higher test score averages[3].

Even in the absence of gaming and cheating, research suggests that test scores fail to accurately describe the changes in the level of student proficiency. Kane and Staiger, (2002, 2003) have noted that a school's mean test scores provide a noisy measure of school performance. This is mainly due to the fact that a large percentage of the variance in test score changes from year to year is found to be transitory and may only be the result of transitory good or bad luck. Figlio and Rouse (2006) find that test score gains in Florida can be explained by the spurious changes in student characteristics, rather than by changes in the effectiveness of schools to educate, as prescribed by accountability programs.

This implies that an analysis on short-term gains does not measure the true effect of accountability, since test scores may only temporarily increase, while later test scores may return to the original mean. If such testing noise is transitory, failure to account for mean reversion in exam performance may substantially bias the estimates of the true impact accountability

[3] Figlio and Getzler, 2002; Figlio and Rouse, 2006; Jacob, 2005; Jacob and Levitt, 2002; Cullen and Reback, 2002; Chakrabarti, 2005; Lemke, Hoerandner & McMahon, 2006 provide a sample of work which documents evidence of teacher cheating and gaming efforts made by teachers and school administrators. Burgess et al (2005) and Chakrabarti (2005) both also find evidence that the distorted incentives cause teachers to give more attention to low-performing students, since increasing their test performance will increase the school's test score average.

programs may have on test scores. Chay, McEwan and Urquiola (2005) test this hypothesis using school test scores from Chile and find that transitory noise and the resulting mean reversion lead the positive estimates of a government-sponsored school program to be greatly overstated.

Given the multitude of problems associated with the use of test scores, this paper will try to avoid some of these difficulties by evaluating the effect of accountability programs on student outcomes as measured by changes to the levels of students' educational attainment and their success in the labor market. Much less work has been done in this area of school accountability. Only a handful of other papers have studied how other student outcomes may have been affected by accountability treatment, but all with mixed results[4]. For example, Ladd (1999) reports potentially positive effects of accountability on Dallas' dropout rate, while Monk, Hussain and Miles (2000) find that increased district participation in a New York state-wide testing program did not serve to increase graduation rates. Rather, their results indicate that participation increases in the state-wide testing program led to an increase in the number of students that failed to complete high school. Focusing on only state-level accountability programs, these studies rely on small samples and produce state-specific results that cannot be generalized to populations in other states. In addition, their results do not allow for readers to draw any inference as to whether accountability programs were effective in narrowing the educational attainment gap between blacks and whites.

If accountability truly had lasting impacts on students, outcomes should include more than just increased test scores. Increased rates of high school graduation and college attainment should be observed as better educated students will be better prepared to complete their

[4] An interesting outcome studied by Donovan, Figlio and Rush (2006) studies the impact of accountability programs on high-performing students and the behavioral responses displayed by them in college as a result of accountability treatment.

secondary school education and pursue a college education. Accountability should also have increased employment outcomes and led to higher wages, since better educated students will appear more desirable in the labor market and will be able to demand higher wages. Given the discrepancies in educational attainment and labor market outcomes between blacks and whites, the effects of accountability will be evaluated separately to assess whether exposure to accountability programs have differential impacts on black and white outcomes.

4. Regression Framework and Discussion of Data

Identification of the Treatment Effect

Typically, when states choose to implement programs in different years, models evaluating the effects of such programs are identified because treatment varies at the state-year level, allowing researchers to compare the outcomes of treated individuals in states that chose to implement programs with the outcomes of non-treated individuals in states that did not adopt similar state programs. In these cases, all individuals that belong in the same state-year cell are either considered treated or non-treated. However, in the case of accountability, although state adoption varied at the state-year level, treatment does not apply to everyone within the same state-year cell.

Exposure to school accountability only occurred if the individual was a student in primary or secondary school during the year or years after accountability mandates were implemented. In this sense, rather than being identified on the state-year level, the effect of treatment is identified at the state-year-'year seventeen' level. Typically, individuals are 17 years of age when they begin their senior year of high school. Thus, in order to determine whether a student was exposed to school accountability, the year in which the state adopted accountability is compared to the year in which the student reached her 17^{th} birthday. Those who reached 17

years of age before states adopted school accountability are counted as 'not treated', since they would have completed their last year of high school before the programs were implemented in the state. On the other hand, those who celebrated their 17th birthday in years during or after the year in which accountability programs were adopted by the state are said to be 'treated', since they were still likely to be enrolled in school when states implemented their programs. In other words:

treatment = 1 if the year state adopt accountability $\leq$ the year student is 17 years old
treatment = 0 if the year state adopt accountability $>$ the year student is 17 years old

In addition to the variation of treatment assignment within each state-year cell, there is also variation amongst treated individuals, measured by the length of exposure to accountability programs. Similar to before, the length of exposure experienced depends on the year in which the individual was 17 years of age and when accountability was first implemented in the state. The length of treatment is calculated as the difference between these two dates. To summarize,

length of accountability treatment =
[year student is 17 years old – max(year state adopt accountability, year student is 6 years old) + 1]

Thus, the length of treatment for treated individuals in the same state-year cell can range from one year of exposure (if accountability was introduced when the individual is a high school senior and age 17) to twelve years of exposure (if accountability was introduced when the individual is in the first grade and age 6). Individuals who were age 17 before the adoption of accountability programs will have a negative length of accountability treatment; their length of treatment is replaced by 0, to indicate that they are not treated.

Two assumptions must be made when defining treatment in this manner. First, treatment is assigned according to the method described above, where the state mentioned is the individual's birth state. This requires the assumption that respondents receive and complete their education in their reported state of birth. The Bureau's Historical Geographical Mobility Reports provides data regarding geographical mobility rates, broken down according to the type of movement and by three age groups: 5-9, 10-14 and 15-19. Based on data from the March Current Population Surveys administered between 1967 and 2005 (the years respondents in the sample were between the ages of 5 and 17) their annual publication documents that only 2.6%-3.4% of children between the ages of 5 and 19 are cross-state movers. As this percentage of children that report moving to different states before their 19th birthday is low, this assumption appears reasonable and should not greatly influence the estimates.

The second assumption relates to the method used to assign treatment and length of treatment to the individual. Supposing that the individual made timely progress throughout her primary and secondary school education, data describing the individual's highest grade completed and year of birth allows one to calculate the last year the individual was enrolled in school. Using this year in which the individual was last in school and comparing it to the year her state adopted their state accountability program, it is possible to determine if the individual was exposed to accountability programs. In the same way, it is possible to calculate the number of years the individual received treatment.

However, assigning treatment in this manner is problematic in the sense that being in school for a longer length of time is, by construction, positively associated with receiving longer lengths of treatment, as well as a greater probability of completing high school. This correlation is problematic when determining the effect of accountability on graduation outcomes because

completing more grade levels in school is also associated with a higher probability of graduating from high school. When the graduation outcomes are evaluated, the direction of causality will be unclear; staying in school longer may have led an individual to experience accountability, or accountability exposure may have led the individual to complete high school because of the program's treatment effect. Due to this positive association, treatment and accountability cannot be assigned based on the year in which the individual actually completes her highest grade level in school. Rather, treatment and the length of treatment received must be based on the year in which the individual would have potentially entered her senior year in high school (the year in which she turns 17 years old).

Econometric Models

To estimate the effect accountability programs, the analysis will begin with the basic model specification which evaluates how exposure to accountability treatments affected the outcome of treated individuals, relative to those who never experienced treatment.

$$outcome_{ist} = \alpha_0 + \alpha_1 Account_{ist} + \alpha_2 B_i + \alpha_3 Account_{ist} \cdot B_i + X'_{ist}\alpha_4 + \gamma_a + \varepsilon_{ist} \qquad (1)$$

Here, *outcome* is the educational attainment or labor market outcome variable, which will be discussed below in more detail. *Account* is a dummy variable that is equal to 1 if the individual received accountability treatment and 0 otherwise. *B* is a dummy variable equal to 1 if the individual is black, and γ_a is a vector of age dummies. *X* is a vector of individual characteristics that controls for the individual's sex, as well as school characteristics, such as the level of per-pupil spending and the school's ethnic diversity[5], both measured at the state level for the year the

[5] The school's ethnic diversity is measured as the percent of minority children (less than age 17) in each state, for each year. The data come from the March Current Population Survey for the years 1977 to 2005.

individual was seventeen years of age. To allow for a more flexible model, all variables included in the vector *X* and age vector are interacted with *B*, abbreviated for black, to permit the control variables to have differential effects across races.

Econometric theory suggests that analyzing the treatment effects of a macro-level policy on micro-level individual outcomes will cause standard errors to be underestimated. To correct for this, the errors are clustered on the state-of-birth level to account for the serial correlation of residuals by state. This allows the standard errors for individuals within each birth-state cluster to be correlated, but assumes independence across clusters (Moulton, 1990; Bertrand, Duflo and Mullainathan, 2004).

Regression equation (1) implies that the probability an individual experiences an outcome is dependent on two main factors: race and accountability treatment, making the variables of interest in this model *Account* [α_1] and (*Account* + *Account·B)* [$\alpha_1 + \alpha_3$]. However, it would be naïve to assume that student outcomes are affected only by school-level characteristics, age and gender. One way to improve the model would be to include other factors that may be affecting the way student outcomes are changing. Rather than limit the determination of outcomes to time-varying state variables such as per-pupil spending and ethnic diversity, equation (2) includes a full set of state-year fixed effects (γ_{st}) to account for any other time-varying state specific characteristics that may have impacted student outcomes, but cannot be fully controlled for.

$$outcome_{ist} = \alpha_0 + \alpha_1 Account_{ist} + \alpha_2 B_i + \alpha_3 Account_{ist} \cdot B_i + X'_{ist}\alpha_4 + \gamma_a + \gamma_{st} + \varepsilon_{ist} \qquad (2)$$

Similar to the interpretation of model specification (1), α_1 is the effect of accountability on the outcomes of white students, while ($\alpha_1 + \alpha_3$) is the treatment effect experienced by blacks.

Thus far, the model specifications of equation (1) and (2) have not made use of the treatment variation across treated individuals. Rather than consider treatment as a binary

variable, the adoption of accountability at the school-level allows for one to test whether variation in the length of treatment may also have affected student outcomes. Replacing *Account* and *Account·B* with two vectors of treatment dummies, the variables *YrsAcc* and *YrsAcc·B* account for the variation in the length of treatment each individual experienced.

$$outcome_{ist} = \alpha_0 + YrsAcc'_{ist}\alpha_1 + \alpha_2 B_i + YrsAcc'_{ist} \cdot B_i\alpha_3 + X'_i\alpha_4 + \gamma_a + \gamma_{st} + \varepsilon_{ist} \quad (3)$$

YrsAcc describes the number of years the individual was exposed to accountability while in school and ranges from 1 to 12. The excluded dummy variable here is 0 years of treatment, which is equal to 1 for individual who were not exposed to accountability programs. Thus, the coefficients in *YrsAcc* and *YrsAcc·B* are interpreted as the treatment effect, since each coefficient estimates the effect of accountability programs for individuals that received specific years of treatment, relative to the outcomes of individuals who received zero years of treatment. If one believes that the treatment effect is a function of the length of accountability exposure, the results should reveal this relationship, providing stronger estimates as the length of accountability exposure increases. ε is an error term, uncorrelated with the covariates and clustered on the birth-state level. As in the earlier model specification, all variables are interacted with *B* to allow the control variables to have differential effects on blacks and whites.

Accounting for Differences in State Accountability Programs

Even after the passage of NCLB, states maintained the freedom to create and design their own program rules, allowing state accountability programs to differ from one another on a multitude of dimensions. Fletcher and Raymond (2002) document the many levels of cross-state differences and compare these differences to California's accountability measure, the Academic

Performance Index (API), using a 'minimum', 'better' and 'best' ranking for each level of program variation.

Incorporating such differences into a model will not only explain how accountability programs may have affected educational attainment and labor market outcomes for students who were treated by accountability, but may also offer some insight as to if and how varying levels of program strength may have affected students differently. Although state programs differ from one another on a number of dimensions, the analysis in this paper will focus on the number of times students are required to be tested by state programs throughout their primary and secondary school education. Each state's rating is listed in the third column of Table 1. The number of grade levels assessed varied dramatically from state to state. According to Fletcher and Raymond, programs that assessed their students less than 5 times received a 'minimum' rating. In contrast, state programs which required their students be assessed 9 or more times received a 'best' rating.

Focusing on the number of times students are assessed is ideal as it provides an objective way to measure the strength of accountability programs. As the names 'better' and 'best' suggest, other measures of accountability strength may be subjective as it requires the task of subjectively ranking the different types and methods of accountability used by states. In addition, use of testing frequency is the most relevant in determining how accountability affected the black-white achievement gap since more frequent testing directly affect the students, as opposed to other measures of accountability strength, which may have more of a direct effect on schools and teachers (such as how schools report results or the rewards and sanctions faced by schools when they succeed or fail to achieve yearly progress). Accounting for the number of times students are assessed throughout their education is also important as it may highlight the

different effects testing may have had on black students, who are less likely to perform as well on exams as their white classmates.

Schiller and Muller's (2000) work is similar as it looks specifically at how cross-state policy differences affect the relationship between earning a diploma and three predictors of academic success: students' social background, their own motivation and their teachers' expectations. One of the between-state policy variations they study is the extensiveness of states' testing programs. Based on such policy variation, their analysis compares the high school graduation outcomes for students in states that (as of 1993) had adopted some type of performance accountability at the high school level. Their results indicate that a more extensive state testing policy is positively related to the log-odds that students earn a diploma. Specifically, they find that increasing the extensiveness of testing by one standard deviation will lead graduation rates to increase by 9.2%. However, given that very few states had adopted accountability programs by 1993, their results are unlikely to be very reliable, despite the fact that their study and methods are well conceived.

In this paper, a fourth specification uses this variation across state programs and will help determine if varying levels of testing intensity had different effects on student outcomes. Assigning testing intensity dummies that correspond to Fletcher and Raymond's (2002) rating of 'minimum, better and best', these 3 testing intensity dummies are interacted with *Account* to produce treatment variables *low*, *med* and *high*. Since *Account* is equal to zero for individuals in the control group, the interaction of *Account* and the intensity dummies will lead *low*, *med* and *high* to all be equal to zero. In the equation below, the treatment variables *low*, *med* and *high* are summarized by the vector *Intensity*.

$$outcome_{ist} = \alpha_0 + Intensity_{ist}\alpha_1 + \alpha_2 B_i + Intensity_{ist} \cdot B_i\alpha_3 + X_i\alpha_4 + \gamma_a + \gamma_{st} + \varepsilon_{ist} \qquad (4)$$

Each of the student outcomes of interest will be analyzed using the four model specifications described above. Six student outcomes are considered in this paper, thee that relate to the individual's level of educational attainment and three that relate to measures of labor market outcomes. Educational attainment outcomes are evaluated for the full sample of survey respondents between the ages of 18 and 30. Based on the individual's self report, the highest grade completed is a discrete variable that measures the number of completed years of education[6]. In contrast, whether or not the individual earned a high school diploma or ever enrolled in a college is a binary variable. Earning a high school degree is equal to 1 for individuals who reported earning at least a high school degree or GED. Unfortunately, due to the nature of the survey, there is no way to distinguish between individuals who earned a high school diploma, and those who earned a general educational development (GED)[7]. Individuals who earned an associate degree or have completed some college or completed at least a bachelor's degree are considered having enrolled in college.

The analysis on labor market outcomes includes studying both employment and two measures of wage. Employment is a dichotomous variables, where individuals are considered employed if they reported earning positive wages within the last year, and unemployed otherwise. The paper also evaluates how accountability treatment affected the individual's log earnings and log hourly wages. Here, log hourly wages are calculated by dividing log earnings by the product of "usual hours worked in a week" and the "number of weeks worked in the past

[6] In this paper, a high school diploma equals 12 years of completed education, while individuals who reported not receiving a high school degree report the highest grade level they completed before they dropped out. Those that reported earning an associate degree or completing some college are reported as finishing 2 additional years of schooling after high school. A Bachelor's degree is equal to 16 years of schooling, while a Master's degree is equal to 18 years of completed education. Professional and doctoral degrees are counted as being completed in 20 years.
[7] Evidence presented by Cameron and Heckman (1993) suggest that there are large differences between individuals who earn a high school degree and those that earn a GED. They find that the labor market outcomes of GED recipients are more similar to those of high school dropouts, than compared to the outcomes of individuals earning a high school diploma.

year". The log hourly wage is a rough measure, but attempts to allow for more flexibility in the model by allowing individuals to have varying hours of employment across individuals, whether due to different tastes for work or spells of unemployment.

5. Data

Data Sources

Multiple data sources were used to evaluate the ability of state programs to improve student outcomes and narrow the black-white achievement gap. The main source of data comes from the 1% unweighted state samples of the US Census and the American Community Surveys (ACS), available to the public at the Integrated Public Use Microdata Series (IPUMS) website. The Census is conducted on a decennial basis, while the ACS is performed annually. Both surveys provide data on a nationally representative sample of respondents, including their individual characteristics (sex, race, age, marital status, family size and the number of children present in the household), their highest level of education, their employment status and the wages earned.

Data regarding each state's accountability program are found through the Consortium for Policy Research in Education (CPRE) (Goertz, Duffy and Le Floch, 2001) and state education websites. The CPRE provides profiles for each state, with in depth details regarding their accountability programs. Information on each state program included the date accountability programs were adopted and the performance standards adopted by the state. In addition, it provides information regarding how performance data is reported and lays out the procedure for identifying and assisting low-performing schools and districts. The year in which states adopted accountability programs is defined as the school year in which schools began reporting test scores as a way to gauge school performance and quality. Data collected from the CPRE

regarding the year of state adoption was also compared to the years of state adoption used by Hanushek and Raymond (2004) for confirmation of accuracy.

Time-varying state level variables describing school characteristics, such as per-pupil spending and the percentage of minority youth in the state, are also included to control for any differences in schooling environments. The US Census Bureau's Public Education Finances Report annually publishes per-pupil spending dollar figures on the state level, drawn from data provided by the National Center for Education Statistics (NCES). The state's proportion of minority students in primary and secondary school comes from CPS data collected between the years 1978 and 2004, which provides the weighted proportion of minority individuals between the ages 5 and 18 sampled in each state-year. Survey respondents are matched with the values of these two state-level variables for the year they were seventeen years of age.

Accounting for per-pupil spending is important as it may put to rest any concerns relating to policy endogeneity. Since states chose to adopt accountability programs at different times, this may lead some to question whether states chose to adopt accountability programs in a non-random fashion. States that place a high priority on education may perhaps adopt accountability programs earlier, relative to states with a lower priority on education. Accounting for per-pupil spending will control for this if high per-pupil spending is indeed correlated with a high valuation of education.

Age and Survey Year Selection

When deciding on the survey years and the age range to include in the analysis, careful consideration was given to studying the variation in treatment. Table 2 describes the treatment variation by age and year of state adoption, where the first column lists the survey years to be included in the analysis and the first row lists a range of ages for young adults between 18 and

30. For each survey year row, the cells in the table describes the year in which the individual was 17 years old, providing the information needed to identify individuals who received at least one year of accountability treatment. According to table 2, younger individuals in the sample were high school seniors in years relatively close to the survey year, whereas older surveyed individuals were high school seniors in years much earlier than the survey year[8]. The shaded cells indicate the survey years and ages in which treated individuals are present. Since adoption of accountability programs varied at the state-year level, states with early adoption years will include both young and older treated individuals in the sample, while states with late adoption years will have only young treated individuals[9]. On the other hand, treated individuals in states with late adoption years such as 2003 will have relatively young treated individuals[10]. Individuals who are over the age of 30 would not have been exposed to accountability programs, regardless of adoption year. Thus, the analysis conducted will focus on individuals between the ages of 18 and 30.

Unfortunately, since the Census is only conducted on a decennial basis, only the 1990 and 2000 surveys are available for analysis in this study. All survey respondents between the ages of 18 and 30 in the 1990 Census serve as a large proportion of individuals in the control group, since states only began to introduce accountability programs starting in 1993. As demonstrated in table 1, the presence of shaded cells under the survey year 2000 is the first year that includes respondents that experienced accountability. Data for the years 2001 to 2006 come

[8] For example, in the first row, the table tells us that an individual who was 18 years of age in the 1990 survey was recently a high school senior in 1989, while a 29 year old individual surveyed in 1990 was enrolled as a high school senior back in 1978.

[9] For example, the age range of treated individuals in states that adopted accountability in 1993 varies from 18 years age (those who were treated in 1993 and surveyed in 1994) to individuals who are 29 years of age (those who were treated in 1993 and surveyed in 2005).

[10] In this case, the oldest individual who was treated in a state with a 2003 adoption year is 19 years of age (those who were treated in 2003 and graduated in 2005).

from the ACS. The full sample of individuals between the ages of 18 and 30 are included in the sample to document whether accountability treatment affected individuals' level of educational attainment.

However, the full sample of individuals cannot be included in analysis relating to labor market outcomes, particularly if many are still pursuing post-secondary schooling degrees. To determine whether accountability programs can indeed be accredited with more promising labor market outcomes, the analysis will focus on a narrower age range; namely, it will be limited to individuals between the ages of 25 and 30. Early in their careers, young workers often experience short job tenures as they attempt to pinpoint what they are looking for in a career. Concentrating on the labor market outcomes older individuals allows for workers to find good employment matches, where employment status is steadier and wages are able to increase over time. Individuals who are between this age range, but report still attending school are excluded from the labor market sample. The employment decisions and wages earned by these older students are distorted by their student status and may not reflect their true earnings potential, had they not been in school.

Sample Restrictions and Summary Statistics

Across the 1990 and 2000 Census, as well as the ACS samples from 2001 to 2006, nearly 16 million individuals were surveyed by the Bureau of Labor Statistics. However, only a portion of individuals in this sample are of interest. Table 3 describes the way each sample restriction decreased the original group of individuals surveyed, both for the entire sample and by race. Individuals who were under the age of 18 or who were still currently enrolled in high school at the time of the survey were excluded since the outcome of their high school education had not yet been determined. Survey respondents over the age of 30 were also excluded as persons over

this age threshold when sampled would not have had the opportunity to experience accountability. Individuals who were living in group quarter or reported their ethnicity to be non-Hispanic white or non-Hispanic black were also dropped from the sample. Since accountability treatment was assigned according to the individual's state of birth, individuals who were born outside of the 50 states (and DC) were also excluded from the sample. After considering all the sample restrictions, a total of 1,704,753 individuals remain in the entire sample. Of that, 1,492,630 are white and 212,123 are black.

Table 4 provides some summary statistics for this restricted sample, both as a whole and disaggregated by race. Blacks were slightly more likely to have been exposed to accountability programs. However, of those treated by accountability, blacks and whites received similar lengths of average treatment (~4 years). Average per-pupil spending dollar values were also slightly higher for whites relative to blacks, while blacks were more likely to live in more ethnically diverse states. Whites were more than twice as likely as blacks to be married, and had smaller families, namely because black families had more children than white families. The two groups are relatively similar in age, with an average age of 24, but black respondents are more likely to be female, relative to the proportion of women in the white sample. In terms of educational attainment, white individuals were more likely to be better educated. On average, whites finished over half a year of schooling more than blacks, were nearly 10 percentage points more likely to graduate from high school and 15 percentage points more likely to enroll in college.

When considering labor market outcomes, the sample is further restricted such that only individuals between the ages of 25 and 30 are included in the labor market sample. With this additional age restriction, the sample becomes smaller, now containing a total of 750,135

individuals, comprised of 662,494 whites and 87,641 blacks. The bottom panel of table 4 indicates that on average, blacks were 5 percentage points less likely to be employed, worked fewer hours in the year and reported log earnings that were $0.36 lower than whites.

Table 5 provides more information on the variation in treatment length experienced across individuals in the treatment group. The top panel of the table is in reference to the full sample of individuals who are between the ages of 18 and 30. The length of exposure ranges from 1 to 12 years, with the number of treated individuals reported in each cell decreasing as the length of exposure to accountability increases. The last row of the panel reports that the average length of exposure to accountability programs for the full sample is approximately 4 years. On the other hand, the bottom panel of table 5 reports the number of individuals in each treatment length category for the labor market sample. As this sample contains older individuals, the length of treatment they experienced only ranges from 1 to 6 years. Similar to the pattern observed in the top panel, the number of individuals in each treatment group decreases as the length of exposure increases. The average individual in this group received slightly more than two years of accountability treatment.

6. Results

The tables summarizing the regression results all have the same set-up. The results for each student outcome are presented in its own separate table. Each table consists of two panels. Within the top panel, column (i) describes the results for the first specification, the second column (ii) is associated with the second specification, and the third column (iii) is associated with the fourth specification. The results from the third specification are reported in the bottom panel of the table, as treatment in this model is distinguished by the length of treatment the individuals experienced.

Only the variables of interest are presented in the tables, namely the separate treatment effect of accountability for whites and blacks. All other variables included as regressors in the analysis possess the expected sign and magnitude. In all 4 specifications, α_3 represents only the marginal treatment effect for blacks, compared to $(\alpha_1+\alpha_3)$, which represents the total treatment effect. To aid in the interpretation of the results, the total treatment effect for blacks will be presented in the tables, calculated as the linear combination of $(\alpha_1+\alpha_3)$. The marginal effect, α_3, will also be presented in the table with square brackets for reference purposes.

Completed Years of Education

Table 6 presents the regression results of accountability program's effect on the completed years of education. Estimates from the first specification indicate that treatment caused whites to slightly decrease their completed years of education by 0.12 years, but did not affect the level of schooling for blacks. After controlling for state-year fixed effects, the decline experienced by whites falls to -0.07 years, while blacks were found to increase their schooling by 0.07 years. Results in the bottom panel of table 6 confirm the hypothesis that lengthier exposure to accountability programs magnifies the treatment effect. White individuals in the treatment group experienced declines in their completed schooling by a range of -0.06 years (for individuals with one year of exposure) to -0.11 years (for individuals with 6 years of treatment). Blacks, on the other hand, experienced increases in their level of schooling between 0.09 years (for those with two years of exposure) and 0.32 years (for individuals with 10 years of accountability treatment).

The results of the final specification (presented in column iii), which categorizes treatment according to the intensity of testing prescribed by the state, suggests that higher levels of testing had positive effects on both white and black students. Whites in low-intensity testing

states decreased their schooling by 0.11 years, which is a larger fall compared to the decline of 0.06 years for whites in states with medium levels of testing. Blacks treated in high-intensity testing states experienced an increase in schooling by 0.15 years, a larger effect when compared to the average treatment effect of 0.07 years.

High School Graduation Rates

The effects observed in Table 6 are modest, but the increase in schooling experienced by blacks may have been enough to motivate a proportion of these students to finish their final year of high school and graduate with a diploma. Research suggests that graduating with a high school diploma has large implications on the high school graduate's labor market outcomes[11], despite the fact that completing the last few months of their senior year in high school may not have had a large affect on their level of human capital.

The results presented in Table 7 appear to corroborate such a story. Columns (i) and (ii) both suggest that accountability programs induced black students to increase the probability of graduating from high school by approximately 3 percentage points. Despite the decline in completed years of education observed by whites in the previous table, the results in Table 7 suggest that this effect did not affect white graduation rates. As before, the effect of accountability programs increases as the length of treatment increases. The graduation rates of white students do not initially respond to accountability programs. However, once exposure increases beyond year 6, treatment increases high school completion rates between 1 and 5 percentage points. In contrast, black graduation rates respond immediately to accountability programs with increasing levels as treatment lengthens.

[11] Clark and Jaeger (2002), Tyler (2004), Park (1999), and Stern et al (1989) all find that students graduating with a high school diploma fare better relative to their peers who never finish high school.

The results in column (iii) suggest that graduation rates for students also respond to the intensity of testing experienced. Blacks in low- and medium-intensity testing states experienced an average increase in completion rates by approximately 2 and 2.5 percentage points respectively, compared to the 3.5 percentage point increase experienced by blacks who were treated in states with high-intensity testing programs. Whites in low- and medium-intensity testing states did not experience any increases to their graduation rates, but those who were tested the most rigorously experienced nearly a 1 percentage point increase in high school completion.

All four model specifications indicate that accountability programs were successful at narrowing the black-white difference in high school completion rates. Table 7 suggests that, on average, white graduation rates were unaffected by accountability programs, while blacks experienced increases in graduation rates between 2.2 and 3.5 percentage points. The results also suggest that the increases in graduation outcomes are the largest for students who were treated in states with high levels of testing, implying that increased levels of testing led to greater treatment effects for blacks in these states.

The reasons behind such a phenomenon, however, are not entirely clear. States that choose to adopt programs which engage their students in a rigorous testing schedule may be successful in encouraging their students to complete high school because they are able to use test scores as a method of quickly identifying students that are falling behind. In this sense, students are better monitored through high levels of testing, which allow teachers to easily identify and provide attention early on, before students have fallen very far behind. However, there is also reason to believe that the gains made were artificial and may have been a product of the unintended gaming responses practiced by schools.

Even prior to the passage of NCLB in 2002, which required that all high schools explicitly report racially disaggregated graduation rates, as well as test scores by race, a number of states had already begun to report high school graduation rates as a way of allowing parents and the state to measure how well schools were performing. The fact that schools choose to adopt programs which frequently tested their students may have been an indication that the state placed a high priority on ensuring that schools were held accountable for their students. Schools in these states may have faced harsher punishments if graduation rates for the school did not reflect well upon the school. With so much at stake, incidences of gaming may have been more prevalent. Students, particularly low-performing blacks, may have been socially promoted throughout high school and even given diplomas so as to improve the school's image, despite the fact that they would not have graduated under normal circumstances (Brasington, 1999).

As a final test of accountability's effect on education outcomes, an analysis on college enrollment will be conducted, to determine if and by how much accountability programs influenced the individuals' choice to attend college. Due to the fact that college enrollment is conditional on the student earning a high school diploma, the sample is restricted to individuals who earned a high school diploma.

College Attendance

The results presented in Table 8 indicate that accountability did not have significant effects on the decision of treated individuals to attend college. As reported in columns (i) and (ii), college enrollment rates did not increase for blacks, despite the significant increase in the number of black students earning their high school diploma. Similarly, the bottom panel of the table reports that enrollment rates were not affected by accountability treatment, regardless of the length of exposure. Although black college enrollment rates were insensitive to accountability

treatment, Table 8 suggests that whites experienced a decline in attendance rates. After controlling for start-year fixed effects, the results suggest that the number of individuals that chose to pursue post-secondary educations declined by an average 0.8 percentage points. Individuals treated in low-intensity testing states experienced even larger declines in college enrollment, by 1.2 percentage points. Fortunately, whites in states with medium and high levels of testing were not affected.

Despite the fact that accountability programs were able to encourage black students to increase the time they spent in school by a few months, which as a result allowed many to graduate from high school with a diploma, it did not have the same positive effects on students in terms of continued education. Black college enrollment rates were not affected, while the rates for white students declined slightly by approximately 1 percentage point.

Labor Market Outcomes

As explained earlier, analyses on labor market outcomes will be conducted on a smaller sub-sample of individuals; specifically, only individuals between the ages of 25 and 30 will be included since interest will focus on how accountability programs affected the labor market outcomes of individuals later in their careers. Before proceeding, it is important to note the characteristic differences when switching from the full sample to the labor market sample. First, by restricting individuals to be between the ages of 25 and 30, the overall sample size decreased by over 50 percent. The total sample now contains just over 750,000 individuals. Second, there are also fewer individuals that experienced accountability programs in the smaller sample. Compared to 33% of the full sample that belonged in the treatment group, there are now approximately only 10% of individuals in the labor market sample who belong in the treatment group. Of those that were treated, these older individuals also have on average fewer years of

exposure to accountability programs. Whereas the full sample of individuals had up to 12 years of exposure to accountability programs, the length of exposure for individuals in labor market sample ranged between 1 and 6 years, with an average length of exposure just over 2. The specific counts of individuals in each cell are provided in the bottom panel of Table 5. Also, personal characteristics, such as marital status, family size and the number of children present in the household, will now be included into the vector *X*, as these variables will control for much of the variance in employment status and wages, as well as avoiding omitted-variable bias.

Employment Status

The results for the employment status outcome are presented in table 9. Column (i) suggests that accountability programs did not have much of an effect on labor market outcomes for either whites or blacks. However, the results from column (ii) estimate that blacks increased their employment rates by approximately 1 percentage point after accounting for state-year fixed effects. Estimates from the bottom panel of Table 9 are estimated with less precision than before due to the small samples of individuals in each treatment cell, but report that all of the significant estimates for blacks are positive, supporting the results seen in earlier specifications. Results from the final specification, listed in column (iii), suggest that employment rates increased by 1 percentage point for blacks, but only for those treated in states which required medium levels of testing. Overall, the results from Table 9 suggest that accountability programs did not have much of an impact on the employment outcomes for white individuals, but were successful in increasing the likelihood of employment for blacks by approximately 1 percentage point.

Log Earnings and Log Hourly Wages

As a final test of accountability's effect on long-run student outcomes, the wages of individuals in the sample will be analyzed to determine if individuals in the treatment group experienced higher wages. This analysis is restricted to include only individuals who reported being employed and whose wages were greater than zero. Results from column (i) and (ii) in Table 10 suggest that, under both specifications, accountability programs did not affect individuals' log earnings. When treatment is disaggregated by length of accountability exposure, individuals in only one of the treatment years was found to have a significantly negative impact on their earnings; all other coefficients reported for both blacks and whites are insignificant. Finally, turning to the final specification in column (iii), the results again suggest that the varying levels of testing may have been an important component to how accountability programs affected long-run student outcomes. Whites that were treated in states with a medium-intensity testing experienced a 2.7 percentage point decline in their earnings, while blacks in states with a high-intensity of testing experienced a 4.6 percentage point decline.

However, given that individuals with different taste preferences for work or individuals that experienced unemployment spells will both report lower total earnings for the past year, an analysis on hours worked was also conducted to ensure that differences in this variable was not the driving force behind the lower log earnings of whites and blacks. Here, hours worked is calculated by multiplying "weeks worked in the previous year" with "usual hours worked in one week". The results in Table 11 suggest that there may be some truth to the proposed hypothesis. The total number of hours worked by was unaffected by accountability programs in the first

three specifications[12]. Results for the final specification listed in column (iii) however, find that blacks treated in programs that required a high level of testing did indeed work fewer hours than others in the sample. After accounting for this hours differential, the effect of accountability programs on log hourly wages (calculated as log earnings divided by hours worked) for blacks treated in states with high-intensity testing is now actually positive and statistically significant. However, due to the fact that black's log hours wages increased by only 0.3 percentage points, the impact of accountability programs on log hours wages is economically insignificant.

7. Sensitivity Analysis

The overall results from the paper suggest that accountability programs were effective in encouraging black students to complete more schooling, thereby increasing the number of black students who complete their high school education and earn a diploma. However these positive effects do not encourage more blacks to attend college, nor do they ultimately have any large impacts on students' long-run outcomes in the labor market. Although these estimates tell a reasonable story that is consistent with much of the literature, it is important that the results presented here are not sensitive to any assumptions imposed on the analysis. In the final analysis of the paper, several of the key assumptions will be relaxed to see if the significance and magnitude of the estimates are sensitive to these changes.

The main assumption used throughout the paper hinges on the methodology used to assign individuals to treatment. Individuals were said to have been treated by accountability if the state in which they were born in had adopted accountability programs prior to their 17th birthday. If one believes that using the state of birth is not necessarily the most accurate method

[12] This is with the exception of results for blacks with 6 years of accountability treatment. However, it is significant only at the 10% level of significance and imprecisely estimated due to the small number of individuals in that treatment cell (there are only 131 black individuals with 6 years of accountability exposure).

of assigning treatment, the estimates presented in Tables 6 through 12 may be inaccurately attempting to capture the effects of accountability programs. To test if the results are sensitive to the way treatment is assigned, a second set of analyses were conducted using the individual's current state of residence. This second set of estimates (although not presented in the paper, but are available by requests to the author) is similar to the results discussed earlier in the paper. All coefficients that are significant in tables 6 through 12 are remarkably similar in significance and magnitude of the results where treatment is assigned according to the individuals' current state of residence. They too suggest that accountability programs were effective in increasing graduation rates for blacks, but did not have significant impacts on outcomes in the long-run.

8. Discussion and Conclusions

The existing accountability literature has thus far focused the majority of its attention on the treatment effects of state accountability programs on student test scores, and in some cases, how it has affected the widely documented black-white test score gap. However, it has been difficult to disentangle the effects of accountability programs, given the documented evidence of gaming and cheating. While the test score gap is of worthy concern, it is presumably the socioeconomic gap that policy makers are ultimately concerned with. This paper looks to study whether accountability programs may have had lasting effects on treated individuals, which would be reflected through their level of educational attainment and labor market outcomes. And in particular, special attention is given to determining whether accountability programs affected blacks and whites differently. The comparison of treatment effects across the two groups will provide some insight into whether accountability programs were able to narrow the black-white educational and socio-economic gap.

Results from the analysis conclude that accountability programs were successful in improving the short-run outcomes of black students, but did not have a large impact on individuals' long-run labor market outcomes. On average, treatment led blacks to increase their completed years of schooling by a modest one-tenth of a year, but unfortunately also led whites to decrease their schooling by approximately the same magnitude. Though the increase is small, the additional education received was enough to increase the rate at which black students completed their high school education and earned a high school diploma by 2.6 percentage points. The results suggest that the positive effect of accountability programs on high school completion rates increases as treatment length increases and also increases as the intensity of testing increases. Blacks treated in low- and medium-intensity testing states experienced increases in graduation rates by 2.2 to 2.4 percentage points, while individuals treated in high-intensity testing states experienced a 3.5 percentage point increase. Although the completed years of schooling decreased very slightly for white students, accountability programs did not have a large impact on the likelihood of completing high school. Only white students that were treated in high-intensity testing states experienced an increase in graduation by 1 percentage point. Unfortunately, the same positive effects of accountability programs did not encourage more students to attend college. Enrollment rates for blacks were unaffected by accountability treatment, and even had negative effects on white students, decreasing their college attendance rates by nearly 1 percentage point, on average.

Many of the estimates recovered from labor market outcomes are insignificant and suggest that accountability programs did not have much of an impact on long-run outcomes. With the exception of a 1.1 percentage point increase in being employed for blacks, analyses on earnings and hours worked find that accountability treatment did not lead workers to earn higher

wages. The initial analysis on log earnings found that blacks treated in states with rigorous accountability programs reported earnings that were 4.6 percentage points lower than individuals in the control group. However, the results on total hours worked imply that lower total number of hours worked by blacks in high-intensity testing states was the driving force behind this result. Estimating the treatment effect on log-hourly wages in fact suggests that blacks in these states earned 0.3 percentage points more than individuals not exposed to accountability programs. Although this estimate is statistically significant at the 5% level of significance, the magnitude of the effect is small and does not have much economical significance.

More than just yielding interesting results, the findings from this study provide conclusions regarding state accountability programs that can have huge policy implications. For instance, accountability was found to have a particularly strong effect on the outcomes of black students. One appealing aspect that can be learned from these results is the fact that outcomes for individuals, and particularly blacks, treated in states with higher levels of testing actually fared better (especially with regards to educational attainment). From the perspective of program designers, they suggest that perhaps one way to increase the effectiveness of accountability programs is through more frequent assessment of students. As documented in Fletcher and Raymond (2002), state programs varied on a variety of other dimensions. Further analysis at a later time that accounts for these other cross-state variations in program strength may provide other policy implications on how accountability programs can be improved. Considering the extensive costs associated with running such a program, part of the goal is to ensure that school accountability lead not only to better test scores for students, but continues to also improve other more tangible educational outcomes, such as high school completion. Unfortunately, to this date, the same short-term positive effects of accountability programs on education outcomes cannot be

generalized to the long-run outcomes examined in this paper. Thus, although accountability programs may have been successful in narrowing the black-white gap in schooling, it has not been able to successfully narrow the socio-economic gap between the two races.

References

Betts, J., 2001. "The Impact of School Resources on Women's Earnings and Educational Attainment: Findings from the National Longitudinal Survey of Young Women" *Journal of Labor Economics* 19(3): 635-657.

Bertrand, M., E. Duflo & S. Mullainathan, 2004. "How Much Should We Trust Differences-in-Differences Estimates?" *The Quarterly Journal of Economics* 119(1): 249-275.

Blume, H. "Test scores up, but achievement gaps persist" *The Los Angeles Times* 6 June 2007.

Brasington, D., 1999. "Central City School Administrative Policy: Systematically Passing Undeserving Students." *Economics of Education Review* 18: 201-212.

Burgess, S., C. Propper, H. Slater, & D. Wilson, 2005. "Who wins and who loses from school accountability? The distribution of educational gain in English secondary schools" The Centre for Market and Public Organization, Working Paper No. 05/128.

Cameron, S. & J. Heckman, 1993. "The Nonequivalence of High School Equivalents" *Journal of Labor Economics* 11(1) Part 1: 1-47.

Carnoy, M. & S. Loeb, 2002. "Does External Accountability Affect Student Outcomes? A Cross-State Analysis." *Educational Evaluation and Policy Analysis* 24(4): 305-331.

Card, D. & A. Krueger, 1992. "Does School Quality Matter? Returns to Education and the Characteristics of Public Schools in the United States." *Journal of Political Economy* 100(1): 1-40.

Cawley, J., J. Heckman & E. Vytlacil, 1999. "On Policies to Reward the Value Added by Educators" *Review of Economics and Statistics* 81(4): 720-727.

Chakrabarti, R., 2005. "Vouchers, Public School Response and the Role of Incentives: Evidence from Florida" *Doctoral Dissertation*, Kennedy School of Government, Harvard University.

Chay, K., P.J. McEqan & M. Urquiola, 2003. "The Central Rose of Noise in Evaluating Interventions that Use Test Scores to Rank Schools." NBER Working Paper W10118.

Clark, M. & D. Jaeger, 2002. "Natives, the Foreign-Born and High School Equivalents: New Evidence on the Returns to the GED" *Journal of Population Economics* 19(4): 769-793.

Cullen, J. & R. Reback, 2002. "Tinkering Toward Accolades: School Gaming Under a Performance Accountability System." NBER Working Paper No. W1226.

Donovan, C., D. Figlio & M. Rush, 2006. "Cramming: The Effects of School Accountability on College-Bound Students" NBER Working Paper No. W12628.

Fletcher, S. & M. Raymond, 2002. "The Future of California's Academic Performance Index." CREDO, Hoover Institution, Stanford University (April).

Figlio, D. & L. Getzler, 2002. "Accountability, Ability and Disability: Gaming the System." NBER Working Paper No. W9307.

Figlio D. & C. Rouse, 2006. "Do Accountability and Voucher Threats Improve Low-Performing Schools?" *Journal of Public Economics* 90: 239-255.

Goertz, M., M. Duffy & K. Le Floch, 2001. Assessment and Accountability Systems in the 50 States: 1999-2000. Consortium for Policy Research in Education (CPRE) Research Report Series RR-046.

Hanushek, E. & M. Raymond, 2004. "Does School Accountability Lead to Improved Student Performance?" NBER Working Paper No. W10591.

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

Jacob, B. & S. Levitt, 2002. "Rotten Apples: An Investigation of the Prevalence and Predictors of Teacher Cheating" NBER Working Paper No. W9413.

Kane, T. & D. Staiger, 2002. "The Promise and Pitfalls of Using Imprecise School Accountability Measures." *Journal of Economic Perspectives* 16(4): 91-114.

Kane, T. & D. Staiger, 2003. "Unintended Consequences of Racial Subgroup Rules" in P.E. Peterson & M.R. West (Eds.), *No Child Left Behind?* (pp.152-176). Washington, DC: Brookings Institutional Press.

Ladd, H., 1999. "The Dallas School Accountability and Incentive Program: An Evaluation of its Impacts on Student Outcomes" *Economics of Education Review* 18:1-16.

Lemke, R., C. Hoerandner & R. McMahon, 2006. "Student Assessments, Non-test-takers, and School Accountability." *Education Economics* 14(2): 235-250.

Moulton, B., 1990. "An Illustration of a Pitfall in Estimating the Effects of Aggregate Variables on Micro Unit" *The Review of Economics and Statistics* 72(2): 334-338.

Park, J.H., 1999. "Estimation of Sheepskin Effects using the Old and the New Measures of Educational Attainment in the Current Population Survey" *Economic Letters* 62:237-240.

Schiller, K. & C. Muller, 2000. "External Examination and Accountability, Educational Expectations, and High School Graduation." *American Journal of Education* 108(2): 73-102.

Stern, D., I. Paik, J. Catterall & Y. Nakata, 1989. "Labor Market Experience of Teenagers With and Without High School Diplomas" *Economics of Education Review* 8(3): 233-246.

Tyler, J., 2004. "Does the GED Improve Earnings? Estimates from a Sample of Both Successful and Unsuccessful GED Candidates" *Industrial Labor Relations Reviw* 57(4): 578-598.

Table 1: Description of Accountability By State		
State	**Year of Program Adoption**	**Rating by # of Grades Levels Assessed**
Alabama	1997	Best
Alaska	2001	Better
Arizona	2000	Best
Arkansas	1999	Better
California	1999	Best
Colorado	2002	Better
Connecticut	1993	Minimum
Delaware	1998	Better
District of Columbia	1997	Better
Florida	1999	Better
Georgia	2000	Minimum
Hawaii	2001	Minimum
Idaho	1998	Best
Illinois	1997	Better
Indiana	1995	Minimum
Iowa	1996	Minimum
Kansas	1995	Better
Kentucky	1995	Better
Louisiana	1999	Better
Maine	1999	Minimum
Maryland	1999	Better
Massachusetts	1998	Better
Michigan	1998	Better
Minnesota	1996	Minimum
Mississippi	1994	Best
Missouri	1997	Better
Montana	1998	Minimum
Nebraska	2001	Minimum
Nevada	1996	Minimum
New Hampshire	2000	Minimum
New Jersey	1998	Minimum
New Mexico	2003	Better
New York	1998	Minimum
North Carolina	1993	Better
North Dakota	2003	Minimum
Ohio	1998	Minimum
Oklahoma	1996	Better
Oregon	2000	Minimum
Pennsylvania	1995	Better
Rhode Island	1997	Better
South Carolina	1999	Better
South Dakota	2003	Best
Tennessee	1996	Best
Texas	1994	Better
Utah	2003	Better
Vermont	1999	Better
Virginia	1998	Better
Washington	1997	Better
West Virginia	1998	Best
Wisconsin	1993	Minimum
Wyoming	1999	Minimum

A rating of minimum indicates that students were tested less than 5 times in K through 12; better corresponds to states that evaluated their students between 5 and 8 times, while states with a best rating tested their students 9 or more times throughout primary and secondary school.

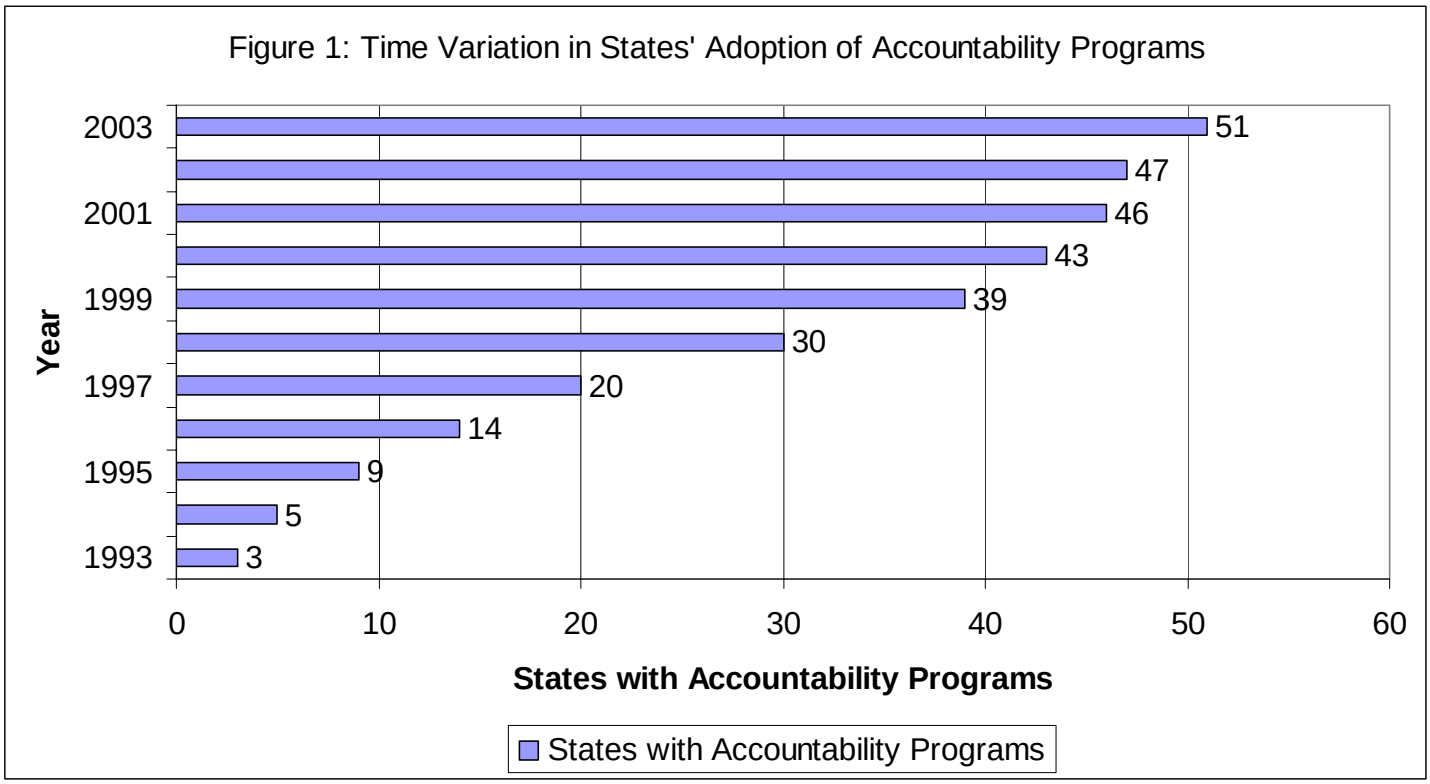

Table 2: Presence of Treated Individuals by Survey Years and Age

	Age when Surveyed												
Survey Year	18	19	20	21	22	23	24	25	26	27	28	29	30
1990	1989	1988	1987	1986	1985	1984	1983	1982	1981	1980	1979	1978	1977
1991	1990	1989	1988	1987	1986	1985	1984	1983	1982	1981	1980	1979	1978
1992	1991	1990	1989	1988	1987	1986	1985	1984	1983	1982	1981	1980	1979
1993	1992	1991	1990	1989	1988	1987	1986	1985	1984	1983	1982	1981	1980
1994	1993	1992	1991	1990	1989	1988	1987	1986	1985	1984	1983	1982	1981
1995	1994	1993	1992	1991	1990	1989	1988	1987	1986	1985	1984	1983	1982
1996	1995	1994	1993	1992	1991	1990	1989	1988	1987	1986	1985	1984	1983
1997	1996	1995	1994	1993	1992	1991	1990	1989	1988	1987	1986	1985	1984
1998	1997	1996	1995	1994	1993	1992	1991	1990	1989	1988	1987	1986	1985
1999	1998	1997	1996	1995	1994	1993	1992	1991	1990	1989	1988	1987	1986
2000	1999	1998	1997	1996	1995	1994	1993	1992	1991	1990	1989	1988	1987
2001	2000	1999	1998	1997	1996	1995	1994	1993	1992	1991	1990	1989	1988
2002	2001	2000	1999	1998	1997	1996	1995	1994	1993	1992	1991	1990	1989
2003	2002	2001	2000	1999	1998	1997	1996	1995	1994	1993	1992	1991	1990
2004	2003	2002	2001	2000	1999	1998	1997	1996	1995	1994	1993	1992	1991
2005	2004	2003	2002	2001	2000	1999	1998	1997	1996	1995	1994	1993	1992
2006	2005	2004	2003	2002	2001	2000	1999	1998	1997	1996	1995	1994	1993

Table 3: Sample Restrictions and Sample Size			
	All	**White**	**Black**
Total Sample	15822933	11665160	1531666
< 18 yrs old	-3940947	-2594143	-463689
> 30 yrs old	-9365418	-7370162	-798315
Attending HS	-147702	-91603	-21557
Not White or Black	-511509		
Non-Natives	-78994	-59360	-19634
Group Quarters	-73610	-57262	-16348
Restricted Sample	1,704,753	1,492,630	212,123

Table 4: Summary Statistics			
Variable	**All**	**White**	**Black**
Accountability Treatment (Dummy)	0.33	0.33	0.34
Years of Exposure to Account (for Treated)	4.06	4.05	4.12
Per-Pupil Spending	5,590	5600.05	5517.07
Ethnic Diversity	0.28	0.27	0.35
Age	24.4	24.5	24.2
Female	0.52	0.51	0.56
Married	0.34	0.37	0.18
Family Size	2.9	2.79	3.30
Number of Children in HH	0.5	0.5	0.6
Completed Years of Education	13.4	13.5	12.8
High School Diploma	0.89	0.90	0.81
College Attendance	0.58	0.60	0.45
Sample (Age 18-30)	1,704,753	1,492,630	212,123
Labor Market Sample			
Accountability Treatment (Dummy)	0.09	0.09	0.09
Employed	0.86	0.86	0.81
Hours Worked (wks worked x hrs/wk)	1710	1745	1442
Log Earnings (if employed)	$9.90	$9.94	$9.58
Hourly Log Wages (if employed)	$0.01	$0.01	$0.01
Sample (Age 25-30)	750,135	662,494	87,641

Table 5: Summary Statistics on Length of Accountability Treatment			
Full Sample (Age 18 - 30)	**All**	**White**	**Black**
1 year of accountabiltiy treatment	99,334	86,840	12,494
2 years of accountability treatment	91,310	79,833	11,477
3 years of accountability treatment	82,355	71,891	10,464
4 years of accountability treatment	73,236	63,737	9,499
5 years of accountability treatment	64,279	56,061	8,218
6 years of accountability treatment	52,457	45,585	8,672
7 years of accountability treatment	38,537	33,514	5,023
8 years of accountability treatment	24,800	21,642	3,158
9 years of accountability treatment	16,331	14,235	2,096
10 years of accountability treatment	10,720	9,259	1,461
11 years of accountability treatment	6,066	5,138	928
12 years of accountability treatment	3,315	2,829	486
Average Length of Accountability Treatment	4.06	4.05	4.12
Sample (Age 18-30)	562,740	490,564	72,176
Labor Market Sample (Age 25 - 30)			
1 year of accountabiltiy treatment	28,065	24,937	3,128
2 years of accountability treatment	18,082	16,119	1,963
3 years of accountability treatment	12,361	10,973	1,388
4 years of accountability treatment	7,615	6,579	1,036
5 years of accountability treatment	3,768	3,245	523
6 years of accountability treatment	1,161	1,030	131
Average Length of Accountability Treatment	2.22	2.21	2.30
Sample (Age 25-30)	71,052	62,883	8,169

Table 6: Completed Years of Education			
	Columns		
Variables	(i)	(ii)	(iii)
Overall Treatment Effect			
Treatement Effect for Whites	-0.118** (0.030)	-0.070** (0.028)	
Treatment Effect for Blacks	0.049 (0.034)	0.072** (0.029)	
	[0.167** (0.039)]	[0.142** (0.034)]	
Low-Intensity Testing			
Treatement Effect for Whites			-0.108* (0.063)
Treatment Effect for Blacks			0.056 (0.067)
			[0.164** (0.074)]
Medium-Intensity Testing			
Treatement Effect for Whites			-0.064** (0.031)
Treatment Effect for Blacks			0.055 (0.044)
			[0.119** (0.040)]
High-Intensity Testing			
Treatement Effect for Whites			-0.013 (0.036)
Treatment Effect for Blacks			0.151** (0.041)
			[0.164** (0.043)]
School Characteristics	Y	Y	Y
State-Year Fixed Effects	N	Y	Y
Level of Testing Intensity	N	N	Y
Sample Size		1,704,505	

Variables	**White**	**Black**	
1 yr of accountability treatment	-0.055** (0.021)	0.020 (0.030)	[0.075** (0.036)]
2 yrs of accountability treatment	-0.070** (0.025)	0.092** (0.031)	[0.161** (0.030)]
3 yrs of accountability treatment	-0.058* (0.034)	0.052 (0.040)	[0.109** (0.035)]
4 yrs of accountability treatment	-0.078* (0.042)	0.105** (0.039)	[0.182** (0.052)]
5 yrs of accountability treatment	-0.098** (0.042)	0.0701 (0.045)	[0.169** (0.055)]
6 yrs of accountability treatment	-0.113** (0.046)	0.091 (0.065)	[0.204** (0.055)]
7 yrs of accountability treatment	-0.100** (0.050)	0.153** (0.073)	[0.253** (0.071)]
8 yrs of accountability treatment	-0.079 (0.049)	0.151* (0.086)	[0.230** (0.084)]
9 yrs of accountability treatment	-0.076 (0.057)	0.079 (0.114)	[0.155 (0.099)]
10 yrs of accountability treatment	-0.045 (0.069)	0.319** (0.097)	[0.364** (0.082)]
11 yrs of accountability treatment	-0.062 (0.083)	0.117 (0.143)	[0.178 (0.108)]
12 yrs of accountability treatment	-0.087 (0.090)	0.164 (0.123)	[0.250** (0.082)]
School Characteristics		Y	
State-Year Fixed Effects		Y	
Level of Testing Intensity		N	
Sample Size		1,704,505	

Notes: ** signifies a 5% level of significant, * signifies a 10% level of significance. Robust standard errors are clustered on the birth state level and are in parenthesis. The total treatment effect of accountability programs are reported separately for whites and blacks. Estimates reported in between square brackets are the marginal treatment effect for blacks (α_3). The total treatment effect of blacks is calculated by the linear combination of (α_1+α_3).

Table 7: High School Graduation

	Columns		
Variables	(i)	(ii)	(iii)
Overall Treatment Effect			
Treatement Effect for Whites	0.000 (0.004)	0.000 (0.002)	
Treatment Effect for Blacks	0.030** (0.007)	0.026** (0.004)	
	[0.030** (0.006)]	[0.026** (0.005)]	
Low-Intensity Testing			
Treatement Effect for Whites			-0.001 (0.004)
Treatment Effect for Blacks			0.022** (0.010)
			[0.023** (0.011)]
Medium-Intensity Testing			
Treatement Effect for Whites			-0.000 (0.002)
Treatment Effect for Blacks			0.024** (0.005)
			[0.025** (0.005)]
High-Intensity Testing			
Treatement Effect for Whites			0.008** (0.003)
Treatment Effect for Blacks			0.035** (0.009)
			[0.027** (0.009)]
School Characteristics	Y	Y	Y
State-Year Fixed Effects	N	Y	Y
Level of Testing Intensity	N	N	Y
Sample Size		1,704,552	

Variables	**White**	**Black**	
1 yr of accountability treatment	-0.000 (0.002)	0.017** (0.005)	[0.017** (0.006)]
2 yrs of accountability treatment	0.000 (0.003)	0.023** (0.006)	[0.023** (0.006)]
3 yrs of accountability treatment	0.003 (0.003)	0.024** (0.005)	[0.020** (0.006)]
4 yrs of accountability treatment	0.004 (0.003)	0.041** (0.007)	[0.036** (0.008)]
5 yrs of accountability treatment	0.002 (0.004)	0.035** (0.007)	[0.032** (0.008)]
6 yrs of accountability treatment	0.003 (0.004)	0.041** (0.009)	[0.037** (0.008)]
7 yrs of accountability treatment	0.010** (0.005)	0.051** (0.012)	[0.040** (0.012)]
8 yrs of accountability treatment	0.016** (0.004)	0.051** (0.020)	[0.034* (0.019)]
9 yrs of accountability treatment	0.019** (0.005)	0.040** (0.018)	[0.021 (0.018)]
10 yrs of accountability treatment	0.016** (0.005)	0.105** (0.017)	[0.089** (0.017)]
11 yrs of accountability treatment	0.030** (0.008)	0.079** (0.023)	[0.049** (0.019)]
12 yrs of accountability treatment	0.052** (0.014)	0.066** (0.024)	[0.013 (0.018)]
School Characteristics		Y	
State-Year Fixed Effects		Y	
Level of Testing Intensity		N	
Sample Size		1,704,552	

Notes: ** signifies a 5% level of significant, * signifies a 10% level of significance. Robust standard errors are clustered on the birth state level and are in parenthesis. The total treatment effect of accountability programs are reported separately for whites and blacks. Estimates reported in between square brackets are the marginal treatment effect for blacks (α_3). The total treatment effect of blacks is calculated by the linear combination of (α_1+α_3).

Table 8: College Enrollment			
	Columns		
Variables	(i)	(ii)	(iii)
Overall Treatment Effect			
Treatement Effect for Whites	-0.018** (0.005)	-0.008** (0.003)	
Treatment Effect for Blacks	-0.004 (0.010)	0.000 (0.008)	
	[0.013 (0.009)]	[0.008 (0.008)]	
Low-Intensity Testing			
Treatement Effect for Whites			-0.012* (0.007)
Treatment Effect for Blacks			-0.006 (0.010)
			[0.005 (0.013)]
Medium-Intensity Testing			
Treatement Effect for Whites			-0.005 (0.005)
Treatment Effect for Blacks			-0.001 (0.010)
			[0.004 (0.010)]
High-Intensity Testing			
Treatement Effect for Whites			-0.009 (0.008)
Treatment Effect for Blacks			0.009 (0.012)
			[0.018** (0.007)]
School Characteristics	Y	Y	Y
State-Year Fixed Effects	N	Y	Y
Level of Testing Intensity	N	N	Y
Sample Size		1,521,159	

Variables	**White**	**Black**	
1 yr of accountability treatment	-0.007** (0.002)	-0.011 (0.008)	[-0.004 (0.008)]
2 yrs of accountability treatment	-0.007** (0.003)	0.008 (0.008)	[0.016* (0.008)]
3 yrs of accountability treatment	-0.006 (0.004)	-0.000 (0.010)	[0.006 (0.008)]
4 yrs of accountability treatment	-0.009 (0.006)	0.004 (0.012)	[0.013 (0.014)]
5 yrs of accountability treatment	0.010 (0.007)	-0.000 (0.011)	[0.009 (0.012)]
6 yrs of accountability treatment	-0.011 (0.008)	-0.006 (0.018)	[0.005 (0.016)]
7 yrs of accountability treatment	-0.010 (0.009)	0.026 (0.020)	[0.036* (0.020)]
8 yrs of accountability treatment	-0.010 (0.011)	0.006 (0.019)	[0.016 (0.018)]
9 yrs of accountability treatment	-0.013 (0.010)	-0.011 (0.027)	[0.001 (0.028)]
10 yrs of accountability treatment	0.008 (0.013)	0.009 (0.023)	[0.001 (0.024)]
11 yrs of accountability treatment	-0.014 (0.013)	0.014 (0.026)	[0.028 (0.021)]
12 yrs of accountability treatment	-0.030 (0.023)	0.012 (0.024)	[0.043** (0.021)]
School Characteristics		Y	
State-Year Fixed Effects		Y	
Level of Testing Intensity		N	
Sample Size		1,521,159	

Notes: ** signifies a 5% level of significant, * signifies a 10% level of significance. Robust standard errors are clustered on the birth state level and are in parenthesis. The total treatment effect of accountability programs are reported separately for whites and blacks. Estimates reported in between square brackets are the marginal treatment effect for blacks (α_3). The total treatment effect of blacks is calculated by the linear combination of (α_1+α_3). Individuals included in the analysis are restricted to only those who earned a high school diploma.

Table 9: Employment Status			
	Columns		
Variables	(i)	(ii)	(iii)
Overall Treatment Effect			
Treatement Effect for Whites	-0.003 (0.004)	-0.000 (0.003)	
Treatment Effect for Blacks	0.004 (0.006)	0.011** (0.004)	
	[0.007 (0.006)]	[0.011** (0.004)]	
Low-Intensity Testing			
Treatement Effect for Whites			0.004 (0.005)
Treatment Effect for Blacks			0.016 (0.011)
			[0.012 (0.011)]
Medium-Intensity Testing			
Treatement Effect for Whites			-0.001 (0.002)
Treatment Effect for Blacks			0.010** (0.005)
			[0.012** (0.006)]
High-Intensity Testing			
Treatement Effect for Whites			-0.008 (0.006)
Treatment Effect for Blacks			0.005 (0.006)
			[0.014* (0.008)]
School Characteristics	Y	Y	Y
State-Year Fixed Effects	N	Y	Y
Level of Testing Intensity	N	N	Y
Sample Size		750,051	

Variables	**White**	**Black**	
1 yr of accountability treatment	-0.002 (0.003)	0.000 (0.010)	[0.002 (0.011)]
2 yrs of accountability treatment	0.002 (0.004)	0.025** (0.011)	[0.022** (0.009)]
3 yrs of accountability treatment	0.002 (0.004)	0.014* (0.008)	[0.012 (0.007)]
4 yrs of accountability treatment	-0.011** (0.005)	0.022* (0.012)	[0.033** (0.012)]
5 yrs of accountability treatment	0.008 (0.005)	0.002 (0.020)	[-0.005 (0.023)]
6 yrs of accountability treatment	0.031** (0.007)	-0.003 (0.008)	[-0.034** (0.009)]
School Characteristics		Y	
State-Year Fixed Effects		Y	
Level of Testing Intensity		N	
Sample Size		750,051	

Notes: ** signifies a 5% level of significant, * signifies a 10% level of significance. Robust standard errors are clustered on the birth state level and are in parenthesis. The total treatment effect of accountability programs are reported separately for whites and blacks. Estimates reported in between square brackets are the marginal treatment effect for blacks (α_3). The total treatment effect of blacks is calculated by the linear combination of (α_1+α_3). The sample is restricted to individuals between the ages of 25 and 30 (referred to as the labor market sample) and individuals who do not report still being enrolled in school. Individuals are counted as employed if they reported earning positive log earnings and unemployed otherwise.

Table 10: Log Earnings			
	Columns		
Variables	(i)	(ii)	(iii)
Overall Treatment Effect			
Treatement Effect for Whites	-0.001 (0.020)	-0.009 (0.011)	
Treatment Effect for Blacks	-0.018 (0.027)	-0.038 (0.024)	
	[-0.017 (0.025)]	[-0.029 (0.022)]	
Low-Intensity Testing			
Treatement Effect for Whites			0.011 (0.018)
Treatment Effect for Blacks			-0.024 (0.042)
			[-0.034 (0.041)]
Medium-Intensity Testing			
Treatement Effect for Whites			-0.027** (0.011)
Treatment Effect for Blacks			-0.040 (0.033)
			[-0.013 (0.027)]
High-Intensity Testing			
Treatement Effect for Whites			0.032 (0.028)
Treatment Effect for Blacks			-0.046* (0.024)
			[-0.077 (0.047)]
School Characteristics	Y	Y	Y
State-Year Fixed Effects	N	Y	Y
Level of Testing Intensity	N	N	Y
Sample Size		641,445	

Variables	**White**	**Black**	
1 yr of accountability treatment	0.000 (0.008)	-0.034 (0.044)	[-0.034 (0.045)]
2 yrs of accountability treatment	-0.010 (0.012)	-0.044 (0.044)	[-0.033 (0.042)]
3 yrs of accountability treatment	-0.020 (0.023)	-0.024 (0.024)	[-0.003 (0.024)]
4 yrs of accountability treatment	-0.032 (0.023)	-0.063 (0.061)	[-0.031 (0.070)]
5 yrs of accountability treatment	-0.038** (0.017)	-0.079 (0.063)	[-0.041 (0.052)]
6 yrs of accountability treatment	-0.053 (0.045)	-0.044 (0.080)	[-0.009 (0.046)]
School Characteristics		Y	
State-Year Fixed Effects		Y	
Level of Testing Intensity		N	
Sample Size		641,445	

Notes: ** signifies a 5% level of significant, * signifies a 10% level of significance. Robust standard errors are clustered on the birth state level and are in parenthesis. The total treatment effect of accountability programs are reported separately for whites and blacks. Estimates reported in between square brackets are the marginal treatment effect for blacks (α_3). The total treatment effect of blacks is calculated by the linear combination of ($\alpha_1+\alpha_3$).The sample is restricted to individuals between the ages of 25 and 30 (referred to as the labor market sample), individuals who do not report still being enrolled in school and individuals who are employed.

Table 11: Hours Employed Last Year			
		Columns	
Variables	(i)	(ii)	(iii)
Overall Treatment Effect			
Treatement Effect for Whites	-10.16 (12.20)	1.238 (9.694)	
Treatment Effect for Blacks	-31.94 (27.79)	-9.742 (25.57)	
	[-21.78 (27.80)]	[-10.98 (24.19)]	
Low-Intensity Testing			
Treatement Effect for Whites			18.27 (13.88)
Treatment Effect for Blacks			23.65 (32.73)
			[5.377 (38.99)]
Medium-Intensity Testing			
Treatement Effect for Whites			-6.17 (12.50)
Treatment Effect for Blacks			-10.98 (35.41)
			[-4.803 (31.07)]
High-Intensity Testing			
Treatement Effect for Whites			-10.72 (15.28)
Treatment Effect for Blacks			-37.47* (20.87)
			[-26.75 (30.92)]
School Characteristics	Y	Y	Y
State-Year Fixed Effects	N	Y	Y
Level of Testing Intensity	N	N	Y
Sample Size		750,051	

Variables	**White**	**Black**	
1 yr of accountability treatment	4.251 (8.736)	-32.36 (33.26)	[-36.61 (34.34)]
2 yrs of accountability treatment	6.127 (11.26)	9.161 (43.78)	[3.033 (39.71)]
3 yrs of accountability treatment	-8.109 (15.96)	12.19 (29.31)	[20.30 (29.80)]
4 yrs of accountability treatment	-20.65 (17.23)	5.326 (35.84)	[25.98 (31.67)]
5 yrs of accountability treatment	11.39 (27.07)	-48.48 (41.82)	[-59.87 (60.44)]
6 yrs of accountability treatment	28.57 (34.21)	71.75* (40.88)	[43.18** (18.62)]
School Characteristics		Y	
State-Year Fixed Effects		Y	
Level of Testing Intensity		N	
Sample Size		750,051	

Notes: ** signifies a 5% level of significant, * signifies a 10% level of significance. Robust standard errors are clustered on the birth state level and are in parenthesis. The total treatment effect of accountability programs are reported separately for whites and blacks. Estimates reported in between square brackets are the marginal treatment effect for blacks (α_3). The total treatment effect of blacks is calculated by the linear combination of (α_1+α_3). The sample is restricted to individuals between the ages of 25 and 30 (referred to as the labor market sample) and individuals who do not report still being enrolled in school. Here, hours employed in the previous year are calculated as the product of "weeks worked in the previous year" and "usual hours worked per week".

Table 12: Log Hourly Wages			
		Columns	
Variables	(i)	(ii)	(iii)
Overall Treatment Effect			
Treatement Effect for Whites	0.000 (0.000)	-0.000 (0.000)	
Treatment Effect for Blacks	0.003* (0.001)	0.002 (0.001)	
	[0.002 (0.001)]	[0.001 (0.001)]	
Low-Intensity Testing			
Treatement Effect for Whites			-0.000 (0.000)
Treatment Effect for Blacks			-0.000 (0.002)
			[0.000 (0.001)]
Medium-Intensity Testing			
Treatement Effect for Whites			-0.000 (0.000)
Treatment Effect for Blacks			0.002 (0.002)
			[0.002 (0.001)]
High-Intensity Testing			
Treatement Effect for Whites			-0.000 (0.000)
Treatment Effect for Blacks			0.003** (0.001)
			[0.003** (0.001)]
School Characteristics	Y	Y	Y
State-Year Fixed Effects	N	Y	Y
Level of Testing Intensity	N	N	Y
Sample Size		641,445	

Variables	**White**	**Black**	
1 yr of accountability treatment	-0.000 (0.000)	0.002 (0.002)	[0.003 (0.002)]
2 yrs of accountability treatment	-0.000 (0.000)	0.000 (0.001)	[0.001 (0.001)]
3 yrs of accountability treatment	0.000 (0.000)	0.001 (0.003)	[0.001 (0.003)]
4 yrs of accountability treatment	0.000 (0.000)	-0.000 (0.000)	[-0.000 (0.002)]
5 yrs of accountability treatment	-0.001 (0.000)	0.005 (0.006)	[0.006 (0.005)]
6 yrs of accountability treatment	-0.000 (0.000)	-0.003** (0.000)	[-0.002** (0.000)]
School Characteristics		Y	
State-Year Fixed Effects		Y	
Level of Testing Intensity		N	
Sample Size		641,445	

Notes: ** signifies a 5% level of significant, * signifies a 10% level of significance. Robust standard errors are clustered on the birth state level and are in parenthesis. The total treatment effect of accountability programs are reported separately for whites and blacks. Estimates reported in between square brackets are the marginal treatment effect for blacks (α_3). The total treatment effect of blacks is calculated by the linear combination of (α_1+α_3). The sample is restricted to individuals between the ages of 25 and 30 (referred to as the labor market sample), individuals who do not report still being enrolled in school and who are employed. Log hourly wages are calculated as log earnings divided by hours worked in the previous year.