CALDER

National
Center *for* Analysis *of* Longitudinal Data *in* Education Research

TRACKING EVERY STUDENT'S LEARNING EVERY YEAR

Urban Institute

A program of research by the Urban Institute with Duke University, Stanford University, University of Florida, University of Missouri-Columbia, University of Texas at Dallas, and University of Washington

Teacher Salary Bonuses in North Carolina

Jacob L. Vigdor

WORKING PAPER 15 • FEBRUARY 2008

Teacher Salary Bonuses in North Carolina

Jacob L. Vigdor*
Duke University and NBER

Prepared for NCPI conference
Vanderbilt University
February 2008

This draft: February 4, 2008

Abstract

Since the 1996/97 school year, the state of North Carolina has awarded bonuses of up to $1,500 to teachers in schools that exhibit test score gains above certain thresholds. This article reviews the details of the bonus program, describes patterns of differences between schools that qualify for bonuses of differing amounts, and presents basic data to address the question of whether the bonus program has improved student achievement, or has led to a narrowing of racial or socioeconomic achievement gaps. There is some evidence to suggest an improvement in overall test scores, particularly in math, but less evidence to suggest that achievement gaps have narrowed. The bonus program has been associated with higher rates of turnover in low-performing schools; differential pay programs may be one way to avoid this unintended consequence.

---

* Box 90312, Durham NC 27708. Email: jacob.vigdor@duke.edu. I am grateful to Mia Bonarski for exceptional research assistance, and to participants in the Education Policy Initiative lecture series at the University of Michigan, who commented on some early results.

1. Introduction

What would happen if teacher salary schedules rewarded performance, as measured by standardized test score outcomes, rather than the acquisition of credentials? Would student test scores improve? Would these improvements be distributed in an equitable way, or would the program encourage teachers to abandon difficult-to-educate students, either by changing jobs or changing the way they teach? Are there any companion policies that could offset potentially regressive impacts?

Starting in the 1996/97 school year, the state of North Carolina implemented a system of performance incentives for all teachers in all public schools. While the specific details of the bonus program have changed over time, the general structure has not. All teachers in a given public school are awarded cash bonuses of up to $1,500 each year, depending on how the students in that school perform on end-of-grade examinations in math and reading, or on end-of-course exams in high school. The performance standard has always been based on the amount of improvement shown by students from one year to the next, rather than proficiency levels. In theory, at least, the experience of North Carolina public schools over the past ten years could provide valuable information on the empirical questions raised above.

This paper provides an overview of teacher bonus programs in North Carolina. Section 2 begins by providing basic details of the performance-based bonus program. Section 3 assesses basic time-series evidence on the impact of the bonus program, by examining trends in test scores for the high-stakes exams that form the core of the program, as well as trends on National Assessment of Educational Progress (NAEP) exams, which although similar in content to North Carolina's exams, have few if any stakes associated with them. Section 4 goes beyond the time-

series evidence, which is limited by the relative brevity of the bonus program and the lack of a clear control group, to consider the impact of the program on a cross-sectional sample of schools, using a regression discontinuity design to infer the impact of failing to receive a bonus on subsequent school performance. Section 5 considers the distributional implications of the bonus program, examining basic trends in achievement gaps in North Carolina and reviewing existing evidence on the bonus program's impact on teacher turnover in low-performing schools. Section 6 reviews a second bonus program, implemented for a three-year period in North Carolina, which illustrates a potentially useful strategy for offsetting any negative impact of performance bonuses on teacher turnover in high-poverty schools. Section 7 concludes.

2. The North Carolina Accountability Program

Beginning in the 1996/97 school year, the North Carolina accountability program, known formally as the ABCs of Public Education, began awarding salary bonuses to teachers in schools meeting specific targets for test score growth in their student body. In the initial year of implementation, teachers in elementary or middle schools were awarded the amount of $1,000 if the mean year-over-year test score gains in the school exceeded a threshold for “exemplary” growth. The formula for computing this threshold is described in section 2.1 below. In the following year, the bonus program was extended to high schools and the bonus was altered to have a two-tiered structure, with teachers in schools meeting “exemplary” growth receiving $1,500 and teachers in schools meeting “expected” growth receiving the amount of $750. This basic structure has been in place ever since, though the formula for computing the bonus eligibility threshold has changed, and the label “exemplary” was in 2001/02 replaced with the

term “high.”

The practice of awarding bonuses to teachers on the basis of the entire school's performance has a theoretically ambiguous effect on the strength of incentives present to improve test scores. On the one hand, tying bonus payments to group performance dilutes the impact of an individual teacher's effort on the probability of receiving a reward. This introduces a potential “free rider” problem, whereby teachers reduce their effort because the ultimate outcome is largely beyond their personal control. The existence of a free rider produces the theoretical prediction that the bonus program should have had a stronger impact in smaller schools.

On the other hand, to the extent that improving test score performance requires cooperation among the teachers in a school, the use of group-level incentives could encourage good habits. Moreover, the use of school-level performance sidesteps concerns about how to effectively incentivize teachers in untested grades (K-2) or in untested subjects (anything other than math and reading in middle schools).

The following two subsections provide basic information on the computation of bonus eligibility thresholds under two regimes, in place before and after the 2005/06 school year. Section 2.3 then describes basic patterns of bonus receipt over time: the frequency with which schools received performance bonuses, and the school-level correlates of bonus receipt.

2.1 The pre-2005/06 model for computing bonus eligibility thresholds

Prior to the 2005/06 school year, North Carolina elementary and middle schools, serving grades 3 through 8, were evaluated on the basis of their ability to improve students' test scores

from one year to the next by more than a pre-determined mean amount. The state used a simple formula of the form

$$(1) \quad \Delta y_{igst} = \Delta \bar{y}_{gs94} + b_1 ITP_{igt} + b_2 IRM_{igst}$$

where Δy_{igst} represents the target threshold for the year-to-year change in test scores in subject *s* for students in grade *g* in year *t* at school *i*, $\Delta \bar{y}_{gs94}$ is average change in test score for a student in grade *g* anywhere in North Carolina at the end of the 1993/94 school year, relative to that same student's score in grade *g*-1 at the end of the 1992/93 school year.[1] These were the first two years in which North Carolina administered statewide end-of-grade tests in reading and math. North Carolina's standardized tests employ a developmental scale, which permits scores from consecutive grades to be directly comparable to one another. Ignoring the second and third terms on the right hand side of equation (1) for a moment, the basic model rewarded teachers when average test score gains in their school exceeded the statewide average between 1992/93 and 1993/94.

The second and third terms in equation (1) are "correction" factors. The term ITP_{igt} refers to the "Index of True Proficiency" for students in grade *g* at school *i* in year *t*. The index of true proficiency does not vary by subject. It is obtained by subtracting the 1994/95 state average scale scores from the average scores of students in grade *g*-1 at school *i* in year *t*-1. The coefficient b_1 varies by grade and subject, but is universally positive.[2] Thus schools with students who achieved higher level test scores in grade *g*-1 in year *t*-1 had to attain a greater

1 Test score gains for third graders are computed by comparing their end-of-grade reading and math test scores to scores on a pretest in the same subjects. The pretest is administered at the beginning of the school year. The "benchmark" average growth for third graders was initially based on the results of pretests and end-of-grade tests administered in the 1996/97 school year. This benchmark was later replaced with results from the 2000/2001 school year.

2 In the 2003/04 school year, for example, the coefficient b1 was 0.22 for reading in all grades except third, where it was 0.47. In math, the coefficient was 0.26 in all grades except third, where it was 0.20.

degree of growth to be eligible for bonus payments, other things equal. The rationale for including this correction factor was the premise that higher-achieving students should have greater test score growth over time. This premise is debatable – it may be more difficult to produce significant gains from high-achieving students. As we will see, however, the higher standard imposed on high-achieving students by this correction was in practice offset by the second.

The second correction factor, IRM_{igst}, was intended to account for statistical noise in standardized test scores. Students who score unusually well in subject *s* in grade *g* in year *t* are more likely to exhibit slower test score growth over the subsequent year, simply because their initial test score was more likely to have been high for idiosyncratic reasons. Some high-scoring students, for example, may simply have guessed a number of correct answers on multiple-choice tests. Similarly, some low-scoring students may have guessed poorly, or may have been negatively affected by poor health or other distractions on the day of the exam. When aggregated to the school level, concerns about mean reversion are lessened, as the idiosyncratic factors producing noise in test scores cancel out at least to some extent. The degree of "canceling out" rises in proportion to the size of the school.[3] For this reason, a rational correction for mean reversion would have treated schools of different sizes differently. Small schools with an unusually high previous year mean should have received a greater discount on their growth threshold than large schools with similarly high test scores.

The North Carolina formula did not use a rational correction for mean reversion. Instead,

3 Statistically speaking, if one student's test score equals his or her true achievement plus an error term with mean zero and variance σ^2, then the mean test score in a school with *n* students is equal to the mean true achievement plus an error term with mean zero and variance σ^2/n, so long as the students' error terms are independent of one another. If student errors are perfectly correlated with one another – an unlikely scenario – then the school-level error term is independent of size.

the index of mean reversion, calculated separately by subject, was nothing more than the difference between the average score of school i students in subject s in grade g-1 in year t-1 and the statewide average on the same test in 1994/95. The coefficient b_2 varies by grade and subject, but is always negative.[4] Moreover, the coefficient b_2 is universally greater in absolute value than the coefficient b_1. Thus comparing any two schools with equal reading test scores, the school with higher math test scores faced a lower threshold for test score growth over the subsequent year.

Does this imply that the pre-2005/06 system penalized schools serving students with low initial performance? This question turns out to be very difficult to answer. It may well be the case that it is easier to produce test score gains with lower-performing students. Evidence that more disadvantaged schools were less likely to cross the bonus threshold could be taken as evidence that the playing field was tilted against them, but might also reflect lower quality of instruction at those schools. What is less controversial is that this system incorporated a feedback mechanism. Schools that achieved high growth in year t not only received bonus payments, but were also rewarded by having a lower threshold set for the subsequent year.

Regardless of where thresholds were set, schools faced a straightforward incentive to increase the mean test score in reading and math. For a school with G tested grades, a set of $2G$ mean test scores were produced every year, to be compared with the set of $2G$ bonus thresholds. To reduce this information to a single indicator of bonus eligibility, the differences between actual test score gains and the target threshold were standardized (by dividing by the standard deviation of this difference across all schools in the state) and averaged, with the average

4 In the 2003/04 school year, the coefficient was -0.58 for math in all grades, -0.60 in reading for all grades except third, and -0.98 for third grade.

weighted by the number of students taking each test. If this weighted average, the "expected growth composite," exceeded zero, teachers in the school were eligible for bonus payments of $750.[5] A second composite measure was computed by multiplying each of the 2*G* growth thresholds by 1.1, then executing a similar procedure of subtracting the second threshold from the actual test score growth in each grade and subject, standardizing, and taking the weighted average. If this second weighted average exceeded zero, teachers in the school were eligible for bonus payments of $1,500.

The procedure for evaluating high schools differed in the pre-2005/06 model. Students stop taking uniformly scaled end-of-grade standardized tests in 8th grade. In high school, students take end-of-course examinations in a limited number of subjects. The threshold for bonus eligibility is based on student performance on these end-of-course (EOC) tests, as well as information on dropout rates and student performance on 10th grade school-wide exams. Beginning in 2000/2001, thresholds for performance on EOC exams were computed in a manner analogous to the end-of-grade growth formulas. The threshold was set equal to the state average score on each test, plus a correction factor based on the 8th grade test performance of students enrolled in the course. The threshold for eligibility was generally set higher for schools serving students who scored better on the relevant 8th grade test. As in K-8 schools, teachers received a $750 bonus when the weighted average of differences between actual performance and subject-specific thresholds exceeded zero, and $1,500 bonuses when the weighted average difference between actual performance and slightly higher subject-specific thresholds exceeded zero.

5 Schools meeting this criterion were ruled ineligible for bonus payments if they claimed an "excessive" number of exemptions from testing, or tested fewer than 98% of all eligible students.

2.2 The post-2004/05 model for computing bonus eligibility thresholds

The accountability system was evaluated during the 2003/04 state legislative system, and several perceived flaws were noted. Among other things, the formula proved difficult to adapt to changes in the underlying standardized tests. Although not explicitly stated in official reports, the system could also be criticized for poorly addressing the concern of mean reversion, and using a formula that rewarded schools even when large test score gains were concentrated among a small minority of students. In response to these flaws, the department of public instruction modified the formulas for determining whether schools were eligible for bonus payments. The new formulas went into effect in the 2005/06 school year.

The primary change in the formula was to stop using the difference in developmental scale score as the main measure of a student's progress from grade *g*-1 in year *t*-1 to grade *g* in year *t*. Instead, the new formula effectively transforms each student's test score into a *Z*-score, using a mean and standard deviation derived from the first year in which a given standardized test was used in North Carolina.[6] An individual student's "Academic Change" is then calculated according to the following formula:

(2) $\text{Academic change}_{gt} = Z_{gt} - d(Z_{g\text{-}1t\text{-}1}/2 + Z_{g\text{-}2t\text{-}2}/2)$.

The formula takes the average of the student's two prior *Z*-scores, multiplies this average by a discount factor *d*, and subtracts them from the current-year *Z*-score.[7] The discount factor is used to address mean reversion: students with prior scores further away from the average are expected to move towards the average over time. The procedure using EOC test score results for high

6 A *Z*-score is the difference between any one observation of a variable, such as a test score, and the mean of that variable, expressed in units of standard deviation. For example, on a test with mean 100 and standard deviation 10, a score of 90 would translate into a *Z*-score of -1.

7 In cases where only one prior year test score is available for a student, that single score is used in place of the average of the previous two. The discount factor is 0.92 when two years' worth of previous test scores are available, and 0.82 in years where a single year's data is available.

school students is similar.

In elementary and middle schools, teachers are eligible for $750 bonuses if the average academic change, across all students in all subjects, is greater than zero. High school eligibility also factors in dropout rates, the results of 10th grade competency exams, and the percent of graduates in college preparatory tracks.

Eligibility for $1,500 bonuses is determined by a different method. Conditional on eligibility for a $750 bonus, schools where the proportion of students with academic change greater than zero exceeds 60% receive the full $1,500. Thus, schools that achieve strong test score growth by raising the performance of a limited number of students will generally not receive the full bonus.

The "new" method of computing bonus eligibility can still be criticized for employing a crude correction for mean reversion. Idiosyncratic factors, such as the quality of random guesses or a student's health the day of the test, can explain part of the variation in student test scores. When aggregated to the school level, however, many of these idiosyncratic factors cancel out, to an extent that varies systematically with the size of the school. Instead of employing a correction factor that makes use of this statistical regularity, the new formula continues to effectively set a higher bar for below-average schools and a lower bar for above-average schools. Schools that manage to hold mean achievement steady from one year to the next receive bonuses if the prior achievement was above the mean, but do not if their prior year achievement was below the mean. For the same reasons stated above in reference to the original bonus eligibility criterion, it is not possible to determine whether this formula on net penalizes or rewards low-performing schools. It is also clear that the new formula incorporates a feedback mechanism: schools that raise

performance both receive rewards and make it easier to requalify in subsequent years.

2.3 Performance bonuses in practice

Figure 1 shows basic information on the proportion of schools who met their expected or exemplary growth standard in each academic year between 1996/97 and 2006/07. Data for the first two years reflects only the performance of elementary and middle schools; beginning in 1998/99, the data include all schools. In this graph, schools meeting the standard for exemplary growth are also counted as meeting the standard for expected growth. Schools that met the expected growth standard only received $750 bonus payments for each teacher; schools that also met the exemplary growth standard received payments of $1,500 for each teacher.

In each year, the majority of schools in North Carolina qualified for some form of bonus. In three of the eleven years shown, the majority of schools received the full $1,500 bonus payment. Thus, the bonus payments were relatively common, but far from universal. There is no evidence of a steady trend in the rate of bonus receipt over time. Bonus receipt peaked in the second year of the program, then declined over the next three years. As noted above, the structure of the bonus program may have contributed to the persistence of trends over time, as schools performing poorly in a given year were assigned higher thresholds for bonus receipt in the following year. Bonus receipt rates bottomed out in 2000/01, peaked again in 2002/03, then bottomed out at an even lower level in 2005/06.

Is it reasonable to think that the quality of instruction in North Carolina public schools varies so dramatically from year to year, and follows such cyclical patterns? There is a considerable amount of stability in the public school system. The great majority of students and teachers persist in the same school from one year to the next. Instructional practices do not vary

much from year to year. It seems more likely that these fluctuations in bonus receipt rates are artifacts of the structure of the bonus program itself, statistical noise inherent in standardized testing, or possibly consequences of minor alterations to the bonus program over time.

There is also evidence, however, that the bonus program was more than a system of randomly assigning rewards to teachers. Table 1 shows the distribution of schools by number of times qualified for $750 and $1,500 bonuses, over the five-year span between 2002/03 and 2006/07. During this time period, an average of 70% of schools received bonus payments each year. If distributed randomly, about 17% of schools would have received a bonus in all five years, less than 1% would have never received a bonus, and about 3% would have received the bonus exactly once. As shown, the number of schools in the extreme categories exceeds these benchmark figures, indicating that there is at least some persistence in bonus receipt. The statistics suggest that the 2,600 public schools in North Carolina consist of roughly 350 persistently low-performing schools, 300 or so persistently high-performing schools, and nearly 2,000 schools for whom receipt of the bonus was essentially a random draw.

Evidence suggests that the distinction between receiving a $750 or $1,500 bonus is closer to a random draw. Among schools that received a bonus only one time in five years, almost half received the $1,500 bonus rather than the $750. Among schools receiving bonuses in multiple years, statistics consistently show that the higher bonus amount was awarded half or slightly less than half the time. There is some evidence of persistence: the number of schools receiving the full $1,500 five times is higher than would be expected if the larger amount were awarded randomly to 50% of schools receiving any bonus. Overall, though, the knowledge that a school received a $1,500 rather than a $750 bonus appears much less meaningful than the knowledge

that a school received any bonus in the first place.

As mentioned previously, knowledge that schools serving disadvantaged populations were less likely to receive the bonus could imply that instruction quality truly is lower in those schools, or that the bonus program itself imparted a bias against those schools. With this important caveat in mind, Table 2 shows summary statistics for school/year observations in the interval between 2002/03 and 2006/07, by whether the school received any bonus in the given year, and the amount of the bonus if so. As foreshadowed, schools receiving no bonus payment served a higher proportion of black and hispanic students, and a higher proportion of students participating in the Federal free and reduced price lunch program. Consistent with the notion that the distinction between schools receiving $750 and $1,500 bonuses is largely random, schools in these two categories are largely indistinguishable along these three dimensions.

Given the collective nature of the bonus program, one might expect a stronger response in smaller schools, where the free-rider program is easier to overcome. In fact, this is not the observed pattern. Schools receiving bonus payments tend to be larger than others. This may reflect the fact that smaller schools tend to be located in rural areas of the state, which are generally poorer than the state's urban areas. As virtually all cities in North Carolina are served by county-wide school districts, high-poverty inner-city schools are relatively uncommon. It is interesting to note that among schools receiving a bonus, those qualifying for the full $1,500 are on average 10% smaller. This pattern may be an artifact of the free-rider problem.

Middle schools were disproportionately unlikely to receive bonus payments in any given year. The difficulties faced by middle school students in North Carolina and elsewhere have been widely established and discussed (see, for example, Cook et al., 2008). Moreover, middle

schools tend to have high rates of teacher turnover, which would support the hypothesis that instruction quality tends to be lower in those schools (Clotfelter et al., forthcoming). Schools serving a wider range of students, for example grades K-8 or 6-12, are also disproportionately represented in the no bonus category, which is unsurprising since these configurations almost always contain middle grades as well. Among the schools receiving at least some bonus, elementary schools are more likely to receive the full $1,500. Since elementary schools are typically smaller than high schools, this can be construed as further evidence that the "free rider" problem is an important limitation to the impact of school-level performance incentives.

3. Time-series evidence on the bonus program's impact

Have North Carolina's bonus payments, offered to a majority of the state's teachers in many years, improved student performance on standardized tests? This question is inherently difficult to answer. The effects of incentives such as this are systemic in nature: they should have increased teachers' efforts regardless of the ultimate outcome.[8] Moreover, the incentive system was put into effect simultaneously across the state, leaving no reliable control group to aid in the identification of treatment effects. Finally, the bonus program was implemented along with a more comprehensive system of school ratings. Schools are assigned one of several ratings each year based on the overall proficiency level attained by students in that year. It is therefore impossible to ascertain whether any purported effects of the accountability system are

8 One potential strategy for identifying the impact of the bonus program would be to exploit the free-rider hypothesis, which predicts that the impact of schoolwide incentives would be smaller in larger schools. There is some preliminary evidence in Table 2 above to suggest that the free rider problem has been a factor in North Carolina. A test based on the free rider problem would be weak in one critical respect: a failure to find that performance in small schools improved relative to larger schools could be taken as evidence either that the incentives had no impact, or that the free-rider problem was unimportant. North Carolina's crude implementation of corrections for mean reversion also threaten such an identification strategy. Other things equal, larger low-performing schools faced a higher hurdle for bonus qualification, while larger high-performing schools faced a lower hurdle.

attributable to the bonus payments or to the broader system of school ratings.

With these caveats in mind, this section presents some basic time-series evidence on student proficiency rates in North Carolina. If the bonus program had a positive impact on student test scores, we would expect proficiency rates to grow over time, as cohorts exposed to the program for at most a brief period of time are replaced by cohorts for whom the bonus program has always been in effect. These across-cohort comparisons are hampered by additional trends in North Carolina public schools, most notably the rapid growth of the Hispanic population. Hispanic students generally attain lower scores on standardized tests in North Carolina, although they also show some progress conditional on remaining in the public school system for a period of multiple years (Clotfelter, Ladd and Vigdor, forthcoming).

This is not the first attempt to estimate the impact of performance incentives on student outcomes. Previous studies have generally focused on much smaller programs, however. Eberts, Hollenbeck and Stone (2000) evaluate a program implemented by a single high school. Figlio and Kenny (2006) evaluate numerous programs implemented by public and private schools nationwide, lumping various programs into categories on the basis of the strength of the incentive. Programs similar to North Carolina's, implemented in Dallas (Ladd 1999), Israel (Lavy 2002), and Kenya (Glewwe, Ilias and Kremer, 2003), have been evaluated previously, with mixed evidence on effectiveness in raising test scores.

Figure 2 illustrates the time series trends in proficiency rates for 8th grade students in North Carolina on two different reading tests: the North Carolina end-of-grade tests used for purposes of determining bonus eligibility, and the lower-stakes National Assessment of Educational Progress (NAEP) test. Proficiency rates are shown beginning in 1998, and

continuing through 2007. For the 8th grade students of 1997/98, the bonus program began in their 7th grade year. For the 8th grade students of 2006/07, the bonus program predates their entry into the public school system. Thus, even though this chart offers no variation in the existence of a bonus program at the time of test implementation, the hypothesis that the bonus program's impact would cumulate over time suggest that we should observe some difference across cohorts.

According to North Carolina's own test results, there have been significant improvements in student reading over time. Across the cohorts shown here, proficiency rates increase from 80% to 88%. Most of this gain occurred in the first few years, when each successive cohort represented an additional years' exposure to the bonus program. Proficiency rates level off in 2004; the four cohorts exposed to the bonus program since their first grade year perform at nearly identical levels as 8th graders.[9] Taken in isolation, this pattern suggests that the implementation of the bonus program raised proficiency rates, with an extra years' worth of exposure to the program associated with a one percentage point increase in proficiency.

The second time series displayed on this graph suggests that North Carolina's own test results should not be considered in isolation. Reading proficiency on the NAEP test is generally much lower for North Carolina 8th graders, with proficiency rates hovering around 30 percent rather than above 80 percent. Moreover, while there is some evidence of an uptick in proficiency ratings among earlier cohorts, recent NAEP results have been comparatively poor, with the most recent results indicating a proficiency rate of 28%, relative to a rate of 30% for the 1998 cohort. Overall, then, the bonus program appears to have led schools to improve performance on the

9 It is interesting to note that these four cohorts do vary in their exposure to the No Child Left Behind (NCLB) program, which more directly targeted proficiency as the basis for school sanctions. This basic evidence thus suggests that the system of transfers, supplemental tutoring, and school restructuring imposed on poorly performing schools in the NCLB regime has had little impact on the proficiency of 8th grade students.

high-stakes test, with at best no impact on performance as measured by a more impartial test.[10] This pattern has been observed in other studies comparing student gains on high-stakes and low-stakes tests (Figlio and Rouse, 2006; Jacob 2007)

Figure 3 presents analogous evidence on trends in 8th grade math proficiency ratings using both North Carolina end-of-grade tests and NAEP. Here, the evidence is more consistent. According to the EOG results, proficiency rates increased from 76% to 88% between 1998 and 2007.[11] As with the reading results, most of the increase had occurred by 2004, although the 2007 cohort appears to have made significant progress relative to its predecessors. In this case, the NAEP scores follow a similar pattern of improvement, with proficiency increasing from 27% to 34% between 2000 and 2007. Overall, then, the time-series evidence is less ambiguous in the case of math relative to reading.

4. Cross-sectional evidence on the impact of failing to receive a bonus

As stated above, there are clear limitations to time-series analysis of the impact of North Carolina's bonus program. Cross-sectionally, there is no variation in exposure to the bonus program across public schools. There is, however, variation in the actual receipt of bonus payments. One might expect that the initiative to change teaching practices or personnel is particularly strong in schools that do not receive a bonus in a particular year. The question of whether failure to receive a bonus leads to some improvement in instruction quality, though

10 This contrast is more striking in light of the purported similarity between the stated criteria used both by the state of North Carolina and the NAEP to judge proficiency in reading for 8th grade students. Published criteria in both cases refer to making inferences and drawing conclusions from text, and identifying and evaluating literary devices.

11 Effective in 2005/06, North Carolina redefined the proficiency standard in mathematics, resulting in a drop in reported proficiency rates on EOG tests. The proficiency rates shown for the 2005/06 and 2006/07 cohorts are extrapolated from proficiency rates on 7th and 6th grade EOG math tests, respectively.

perhaps of less ultimate interest from a policy perspective, is easier to answer, particularly given the 2005/06 revisions to the structure of the bonus program. After 2005/06, the implementation of a strict standard for bonus eligibility, coupled with the reporting of the criterion variable in state reports, enables a regression-discontinuity (RD) analysis of the impact of failure to receive a bonus on subsequent student performance. The RD analysis takes advantage of the fact that schools with nearly criterion variables that are nearly identical, but just on either side of the eligibility threshold, are treated very differently. One group receives a bonus, the other does not. The analysis presented here will determine whether schools that just missed bonus eligibility in one year are better or worse the following year, relative to schools that barely qualified.

As described in section 2.2 above, the post-2005 bonus criterion variable is a modified version of the mean change in *Z*-score for students with at least one prior years' test score in a given subject. The modification deflates the change in *Z*-score for students initially below the mean, and inflates it for students initially above the mean, to account for mean reversion. Schools receive bonus if the mean modified change in *Z*-score, hereafter referred to as ΔZ, exceeds zero. Beginning in the 2005/06 school year, the state of North Carolina began reporting this ΔZ on school report cards, along with information on whether schools met "expected growth," the standard for receiving a $750 bonus.

Figure 4 shows the relationship between ΔZ and the likelihood of receiving at least a $750 bonus after the 2005/06 school year. Each point in the graph is an unweighted mean for schools in an interval of width 0.025. In the intervals just below and above zero on this graph, there are 181 and 198 schools, respectively.[12] As expected, there is a clear, sharp discontinuity

12 The sample is trimmed to exclude the lowest performing 25 schools and the highest performing 50 schools, which as outliers have the potential to unduly influence the regression discontinuity analysis. These schools represent just over 3% of all potential observations. The specific criterion for exclusion is a 2005/06 ΔZ value

at $\Delta Z = 0$. Schools just below this threshold never receive bonuses; schools just above this threshold do. Further away from the discontinuity, there is some evidence of "non-compliance;" this can be attributed to the fact that ΔZ is not the sole criterion for bonus eligibility in high schools.

Regression discontinuity is a viable identification strategy under the assumption that all potential covariates influencing the outcome of interest vary smoothly over the interval containing the discontinuity. Figure 5 presents a basic check on this assumption, showing the relationship between the 2005/06 ΔZ and three school-level covariates – percent black, percent hispanic, and percent receiving free or reduced price lunch. Once again, data points have been collapsed into bins based on the ΔZ into bands of width 0.025. Fitted to each data series is a cubic in ΔZ, augmented by a indicator variable for whether ΔZ is positive.[13]

Figure 5 shows a pattern consistent with the summary statistics in Table 2. Schools with a higher proportion of black students, or a higher proportion of students receiving free or reduced price lunch, tend to post lower values of ΔZ. There is no obvious relationship between ΔZ and the share of Hispanic students at a school. More importantly for this analysis, there is little evidence of significant differences between schools that barely make or barely miss the threshold for receiving a $750 bonus. In two of three cases, the regressions underlying the fitted relationships fail to reject the hypothesis of no difference between schools with positive ΔZ and schools with negative ΔZ. The *p*-value for this hypothesis test is 0.976 in the case of percent black, and 0.25 in the case of percent Hispanic. In the case of percent free or reduced price

above 0.25 in absolute value.

13 The underlying regression specifications are weighted by school enrollment, which explains the occasional tendency for the fitted curve to be consistently above or below a series of data points.

lunch, the *p*-value is smaller, at 0.082, at the margins of conventional statistical significance. The associated coefficient is in this case positive, indicating that schools just above the margin are if anything more disadvantaged than those just below. Results reported below will include specification checks that control for potentially complex nonlinear relationships between the proportion of students receiving free or reduced price lunch and ΔZ.

Figure 6 shows the basic results of the analysis. The horizontal axis continues to display ΔZ for the 2005/06 school year. The vertical axis now measures the same variable for the 2006/07 school year. The question addressed here: did schools that just barely missed qualifying for a bonus in 2005/06 perform better than schools that barely qualified in the subsequent academic year? Data points are once again collapsed into bins of width 0.025, and the fitted curve represents a cubic in 2005/06 ΔZ augmented with an indicator for whether that variable was positive.

Before discussing the result of interest, note that the data points are for the most part greater than zero. This may indicate that the quality of instruction improved in 2006/07 relative to 2005/06, or that teachers and administrators learned more about the new incentive system after its first year of implementation and restructured their efforts to improve the likelihood of receiving a bonus. A general increase in the probability of bonus receipt between the two years is also seen in Figure 1.

The evidence in this case points to a clear and statistically significant discontinuity at the bonus threshold for 2005/06. Among schools above the threshold in the initial year, there is a very prominent, nearly linear relationship between ΔZ for 2005/06 and the same variable for 2006/07. At the threshold itself, this relationship weakens considerably. The three data points

immediately to the left of the discontinuity are roughly equivalent to the three data points immediately to the right, whereas these latter three points are considerably lower than the three points to their right. The underlying regression specification indicates that failure to receive a bonus in 2005/06 is associated with a 0.028 increase in ΔZ for 2006/07. In other words, relative to the cubic trend, schools below the threshold improved enough to move themselves into the next highest bin. The estimated effect is significant with a *p*-value of 0.03.

To assess the robustness of this finding to a potential discontinuity in percent of students on free and reduced lunch, the regression equation was re-estimated with a set of 98 indicator variables separating schools into percent free and reduced-lunch bins of width 0.01. In this specification, the estimated magnitude of the discontinuity effect is reduced by about one-fourth, to 0.023, with a *p*-value of 0.095. With a more conventional linear control for percent free and reduced lunch, the estimated effect is 0.025, with a *p*-value of 0.047. Overall, then, the results suggest that the failure to receive a bonus spurs teachers and administrators to alter their practices in ways that produce an average gain of 2 to 3 percent of a standard deviation for each student in each course.

5. Has the bonus program closed achievement gaps?

While one goal of the North Carolina school accountability program has been to increase test scores across the board, a second goal has been to reduce achievement gaps between students of different races, or between students of varying socioeconomic status. Inferring the program's impact on these outcomes is rendered difficult by the same factors that complicate the analysis of the policy's overall effect.[14] Figures 7 through 10 show basic time series evidence drawn from

14 In theory, the regression discontinuity design of section 4 could be applied to the study of achievement gaps.

NAEP 8th grade reading and math administrations, under the hypothesis that cohorts exposed to the bonus program for a longer period of time will demonstrate stronger impacts.

Figure 7 shows trends in 8th grade reading test scores for black and white students on NAEP administrations between1998 and 2007. The 8th grade cohort of 1998 had been exposed to the bonus program for two years at the time of administration, while the 2007 cohort entered kindergarten with the program in place. There is no evidence of a narrowing of the gap across these cohorts – if anything, the mean difference between blacks and whites has increased. The average score for white students has remained constant, while the average score for blacks has declined slightly.

Figure 8 shows trends over the same time period for students receiving free or reduced price lunch and all others. Once again, there is no evidence of a narrowing of the gap across cohorts. Non-participants in the subsidized lunch program witness no change in average test scores over this period, while average scores for recipients has declined slightly.

Figures 9 and 10 present gap data for NAEP 8th grade math administrations between 1996 and 2007. Here, there is evidence of broad improvements in test scores for both advantaged and disadvantaged students, defined either by race or by socioeconomic status. The mean difference in NAEP test scores does not noticably decline over time in either case. To be fair, it is unclear whether, for example, the improvement in mean black test score from the 240s to the 260s is more or less socially valuable than the increase in white mean from the 270s to the 290s. The data on math test score gaps are thus best described as inconclusive. Reading test score gaps, on the other hand, can be assessed more confidently, since scores remained constant for the advantaged groups and declined for disadvantaged groups. Even if the importance of two

Such a study must await the release of student-level microdata for the 2006/07 school year.

increases cannot be ranked, stasis is clearly preferable to decline.

The failure of the North Carolina bonus program to demonstrably close test score gaps may reflect a pattern of teacher responses to the program documented by Clotfelter, Ladd, Vigdor and Aliaga (2004). This study analyzes teacher turnover in North Carolina public schools before and after the implementation of the bonus program. The introduction of the bonus program is associated with a significant increase in the rate of teacher departure from lower-performing schools. Figure 11, reprinted from the original study, shows that the survival rate of teachers beginning spells of employment at low performing schools in 1996/97 is lower than the survival rate of teachers who began spells at comparable schools two years earlier.

Earlier studies have documented a broad tendency for teachers to leave jobs in lower-performing schools in order to take positions at more advantaged campuses, often at the same or lower salary (Loeb and Page 2000; Hanushek, Kain and Rivkin 2004). The North Carolina bonus program, which as demonstrated above tended to steer rewards away from lower-performing schools, created yet another reason for teachers to prefer jobs in higher-performing schools. The intention of the bonus program was to spur teachers and administrators to exert greater effort to increase student test scores. The program appears to have had the unintended consequence of spurring teachers to abandon schools that serve lower-performing students. Differences in expected salary brought about directly by the bonus program most likely explain some of this effect; teachers may also have sought to avoid positions with a strong emphasis on preparing students for standardized tests.

6. A potential strategy for offsetting the impact of performance bonuses on turnover

Teachers generally prefer to avoid jobs in disadvantaged schools; this preference appears to be strengthened when disadvantage translates into a lower likelihood of receiving merit-based bonus payments. The most obvious potential policy lever for counteracting this preference is salary. The question of whether teachers would be more willing to work in a disadvantaged school if they were offered a higher salary to do so is in practice difficult to answer. Suppose, for example, that teacher turnover rates are uncorrelated with salaries among schools with similar observed working conditions. One explanation for this pattern is that higher salaries do not reduce turnover rates. It is also possible, however, that the schools offering higher salaries for equivalent observed working conditions offer inferior unobserved working conditions. Student socioeconomic status, for example, is very easy to measure, but it is more difficult to quantify the degree of parent involvement in a school, or the competence of the district administration.

A recent study by Clotfelter, Glennie, Ladd, and Vigdor (forthcoming) uses a second North Carolina bonus program to address this important but difficult question. For a three year period beginning in 2001/02, the state offered annual bonuses of $1,800 to certified teachers of math, science, and special education who took and remained in jobs in middle or high schools that met one of two criteria: high rates of participation in the free or reduced price lunch program, or high rates of failure on end-of-course examinations in Algebra and Biology. The program thus created within-school variation in salaries, breaking the potential correlation between salary levels and unobserved working conditions (so long as working conditions within a school do not vary substantially across teacher subject area). The analysis by Clotfelter et al. uses compares turnover rates of teachers before and after the implementation of the bonus program, across eligible and ineligible subjects, and between eligible and just-barely ineligible

schools – a methodology often referred to as differences-in-differences-in-differences.

Table 3 shows the study's basic results, derived from a statistical model predicting the probability of a teacher's departure after the t^{th} year of a spell of continuous employment at a single school, conditional on surviving to year t.[15] Table entries are hazard ratios. When below one, the hazard ratio indicates a factor that reduces the likelihood of departure. The basic estimate here indicates that the bonus program reduced the likelihood of departure by 15 percent. More refined estimates, presented in Clotfelter et al. (forthcoming), tend to indicate an even larger effect on teacher turnover. Converted into an elasticity, this estimate suggests that a 1% increase in salary at low-performing schools would lead to a 3-4% reduction in turnover rates.

It should be emphasized that the math, science, and special education bonus program was not performance-based. There were no provisions to ensure that bonus payments were made to the most effective teachers, except in the fact that uncertified teachers were not eligible to receive them. Clotfelter et al. report that the program had the highest proportional impacts on experienced teachers, who have repeatedly been associated with greater effectiveness at improving student test scores relative to novices.

7. Conclusions

North Carolina's accountability bonus program is the nation's largest, and perhaps the longest-running, initiative to reward teachers for producing gains in student test scores. Over the past decade, the program has paid millions of dollars' worth of bonuses to tens of thousands of teachers throughout the state. The creators of the bonus program can be praised for certain aspects of its design, particularly the focus on test score improvements rather than the focus on

15 The statistical model is a Cox proportional hazard models. Clotfelter et al also present results derived from parametric hazard models.

straight proficiency espoused by the No Child Left Behind Act. Repeated tinkering with the incentive system also reflects a willingness to address concerns as they are raised. It is also clear, however, that certain aspects of the bonus program are statistically perplexing, threaten to place disadvantaged schools at a further disadvantage, or weaken the program's potential incentive effect.

The bonus program has always based rewards on the performance of a school, rather than an individual teacher. Economic theory suggests that this emphasis on collective outcomes will lead to a free-rider program, as any one teacher's effort has only a small impact on the school as a whole. Moving to a teacher-level incentive system, however, is not necessarily advisable, given the statistical noise in standardized tests, and the demonstrated failure of many public schools to allocate students in a fair or even way across classrooms (Clotfelter, Ladd and Vigdor 2006). There are clear tradeoffs between the strength of incentives faced by any individual teacher, and the relative importance of luck or political maneuvering relative to effort in determining rewards. Further research is necessary to quantify these tradeoffs, and indeed to determine whether the flaws are sufficient to warrant abandoning efforts to incentivize teachers.

Beyond this tradeoff, perhaps the most glaring flaw in the North Carolina program's design has been its treatment of mean reversion. Mean reversion is a real concern; however, unless noise in student test scores can be attributed entirely to school-level shocks, the proper correction for mean reversion should take account of school size. The state's crude efforts to address mean reversion likely have the unintended effect of penalizing low-performing schools, though the magnitude of the effect is impossible to identify.

There is at least some evidence that the bonus program has led to an improvement in test

scores, though the evidence in this article should be considered less than definitive. Math proficiency rates have increased both on the high-stakes test used to determine bonus eligibility and on the lower-stakes NAEP exam. Reading proficiency rates have improved only on the state's own examination. The regression discontinuity analysis of failure to receive a bonus suggests that schools do implement changes that lead to improvements following a negative outcome.

Hopes that the bonus program would help ameliorate racial or socioeconomic differences in achievement have not been realized, quite possibly because teachers have reacted to the uneven playing field by departing disadvantaged schools in increased numbers. According to NAEP results, achievement gaps in 2007 were just as wide or wider than they had been a decade earlier.

What lessons does the North Carolina experience offer to other states, districts, or individual schools seeking to incentivize teacher effort? Above all else, the results discussed here suggest that incentive programs, when adopted in an effort to raise the performance of disadvantaged students, can be a two-edged sword. If teacher perceive bonus programs as yet another factor making jobs in advantaged schools more attractive, increased turnover rates in low-performing schools are a predictable consequence. This unintended side effect could be avoided so long as teachers perceive the bonus program as a fair reward for their effort, rather than a reward for student background or other inputs over which they have no direct control.

This implication, in turn, suggests a fruitful avenue for further policy-relevant research. To craft a bonus program that presents a truly level playing field, policy-makers require evidence on the expected test score gains of individuals at varying points in the achievement distribution,

under constant instructional quality. Such evidence could be based on "within-class" empirical models, based solely on students enrolled in the same classroom, under the assumption that instructional quality does not vary within classrooms. If instructional quality varies within classrooms, however, more sophisticated methods will be required to derive these estimates.

Finally, given the political controversy surrounding the use of performance bonuses in public schools, it should be noted that the accountability bonus program enjoys broad support in North Carolina. The state does not have a teachers' union with collective bargaining power, which undoubtedly eased the path toward implementing the bonus program, but there is a professional association of teachers, the North Carolina Association of Educators, which engages in policy advocacy on a number of fronts. In its published agenda for the 2007/08 legislative session, there is no opposition to the bonus program. In fact, the NCAE explicitly advocates maintaining the bonus program, and expanding it to certain state-run schools that do not currently participate. The NCAE's stated policy stances may reflect political reality as much as their own preferences, but their expression of explicit support, rather than tacit acceptance, is noteworthy. While there is some evidence of effectiveness in spite of its flaws, it is the sheer popularity of the bonus program that provides the most heartening evidence to jurisdictions contemplating similar initiatives.

References

Clotfelter, C.T., E. Glennie, H.F. Ladd and J.L. Vigdor. "Would Higher Salaries Keep Teachers in High-Poverty Schools? Evidence From a Policy Intervention in North Carolina." Forthcoming, *Journal of Public Economics*.

Clotfelter, C.T., H.F. Ladd and J.L. Vigdor. "Teacher-Student Matching and the Assessment of Teacher Effectiveness." *Journal of Human Resources* v.41 n.4 (Fall 2006) pp.778-820.

Clotfelter, C.T., H.F. Ladd, J.L. Vigdor and R.A. Aliaga. "Do School Accountability Systems

Make It More Difficult for Low Performing Schools to Attract and Retain High Quality Teachers?" *Journal of Policy Analysis and Management* v.23 n.2 (Spring 2004) pp.251-271.

Cook, P., R. MacCoun, C. Muschkin and J.L. Vigdor. "The Negative Impacts of Starting Middle School in Sixth Grade." *Journal of Policy Analysis and Management* v.27 n.1 (Winter 2008) pp. 104-121.

Eberts, R., K. Hollenbeck, and J. Stone "Teacher Performance Incentives and Student Outcomes." W.E. Upjohn Institute working paper #00-65 (2000).

Figlio, D.N. and L. Kenny "Teacher Incentives and Student Performance." National Bureau of Economic Research Working Paper #12627 (October 2006).

Figlio, D.N. And C.E. Rouse "Do Accountability and Voucher Threats Improve Low-Performing Schools?" *Journal of Public Economics* v.90 n.1-2 (January 2006) pp.239-255.

Glewwe, P., N. Ilias, and M. Kremer "Teacher Incentives." National Bureau of Economic Research Working Paper #9671 (May 2003).

Hanushek, E.A., J.F. Kain and S.G. Rivkin. "Why Public Schools Lose Teachers." *Journal of Human Resources* v.39 n.2 (Spring 2004) pp.326-354.

Jacob, B.A. "Test-Based Accountability and Student Achievement: An Investigation of Differential Performance on NAEP and State Assessments." National Bureau of Economic Research Working Paper #12817 (January 2007).

Ladd, H.F. "The Dallas School Accountability and Incentive Program: An Evaluation of its Impacts on Student Outcomes." Economics of Education Review v.18 n.1 (February 1999) pp. 1-16.

Lavy, V. "Evaluating the Effect of Teachers' Group Performance Incentives on Pupil Achievement." *Journal of Political Economy* v.110 (December 2002) pp.1286-1317.

Loeb, S. and M. Page "Examining the Link Between Teacher Wages and Student Outcomes: The Importance of Alternative Labor Market Opportunities and Non-Pecuniary Variation" *Review of Economics and Statistics* v.82 n.3 (August 2000) pp.393-408.

Table 1: Distribution of schools by frequency of bonus receipt over a five year period, 2002/03-2006/07

Made exemplary growth ($1500 bonus)	Made expected growth ($750 bonus)					
	0 times	1 time	2 times	3 times	4 times	5 times
0 times	169	152	98	59	70	50
1 time		145	190	248	198	130
2 times			75	133	166	206
3 times				29	92	175
4 times					25	112
5 times						66
Total all rows	169 (7%)	297 (11%)	363 (14%)	469 (18%)	551 (21%)	739 (29%)

Note: sample consists of all schools with five observations on expected or exemplary growth over the span between 2002/03 and 2006/07.

Table 2: Summary statistics for school/year observations by bonus eligibility status

Variable	No bonus	$750 bonus	$1,500 bonus
Percent free/reduced lunch	47.8%	36.9%	34.4%
Percent black	41.3%	30.7%	30.3%
Percent hispanic	7.7%	6.9%	6.7%
Enrollment	518 (280)	631 (403)	570 (334)
Percent elementary school	37.4%	42.8%	53.8%
Percent middle school	23.1%	15.0%	12.5%
Percent high school	10.0%	20.0%	14.6%

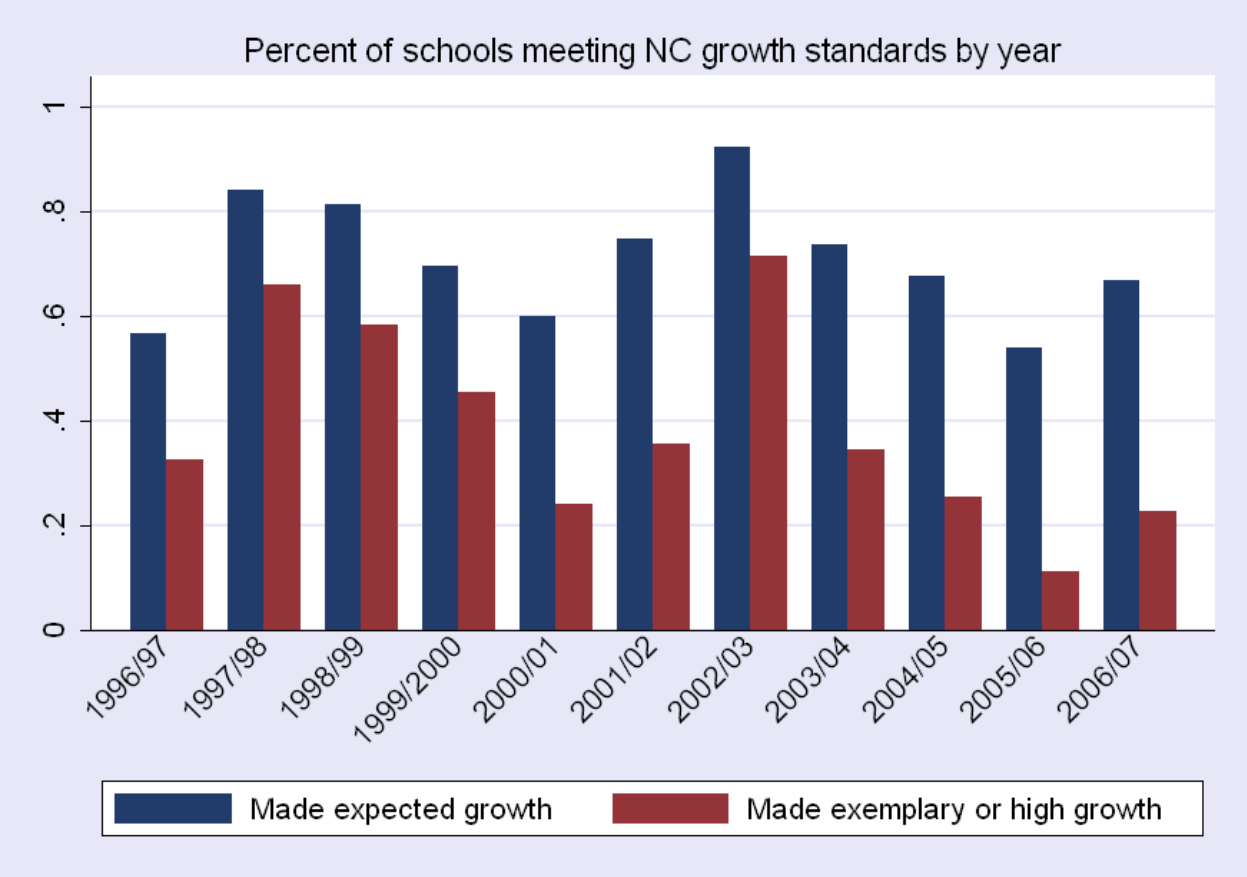
Note: Unit of observation is the school/year. Standard deviations in parentheses, where appropriate. Sample consists of school years between 2002/03 and 2005/06, inclusive. Means and proportions are unweighted. The sample size ranges from 8019 (free/reduced lunch) to 8747 (enrollment). Reductions in sample size can be attributed to missing data in the Common Core. The omitted category of school serves a non-traditional assortment of grades.

Table 3: Basic Estimate of the Math/Science/Special Education Bonus Program's Impact

Teacher Receives a Bonus Payment	0.848** (0.057)
Teacher is Certified in Math, Science or Special Education and is Employed by an Ever-Eligible School	1.005 (0.062)
Teacher is Employed by a Currently Eligible School	0.802** (0.034)
Teacher is Certified in Math, Science or Special Education in a Post-Program Year	1.114* (0.070)
Teacher is Certified in Math, Science or Special Education	0.996 (0.054)
Teacher is Employed by an Ever-Eligible School	1.286** (0.041)
Year is 2000	1.141** (0.050)
Year is 2001	1.560** (0.070)
Year is 2002	1.907** (0.076)
N	29,562
Log likelihood	-59,123.93

Note: Table entries are hazard ratios, with standard errors in parentheses. The hazard refers to the probability of exiting a school after period *t*, conditional on remaining in that school until period *t*. Unit of observation is the teacher/school/year. ** denotes a hazard ratio significantly different from 1 at the 5% level; * the 10% level.

Source: Clotfelter, Glennie, Ladd and Vigdor (forthcoming).


Percent of schools meeting NC growth standards by year
0
.2
.4
.6
.8
1
1996/97
1997/98
1998/99
1999/2000
2000/01
2001/02
2002/03
2003/04
2004/05
2005/06
2006/07
Made expected growth
Made exemplary or high growth
Figure 1

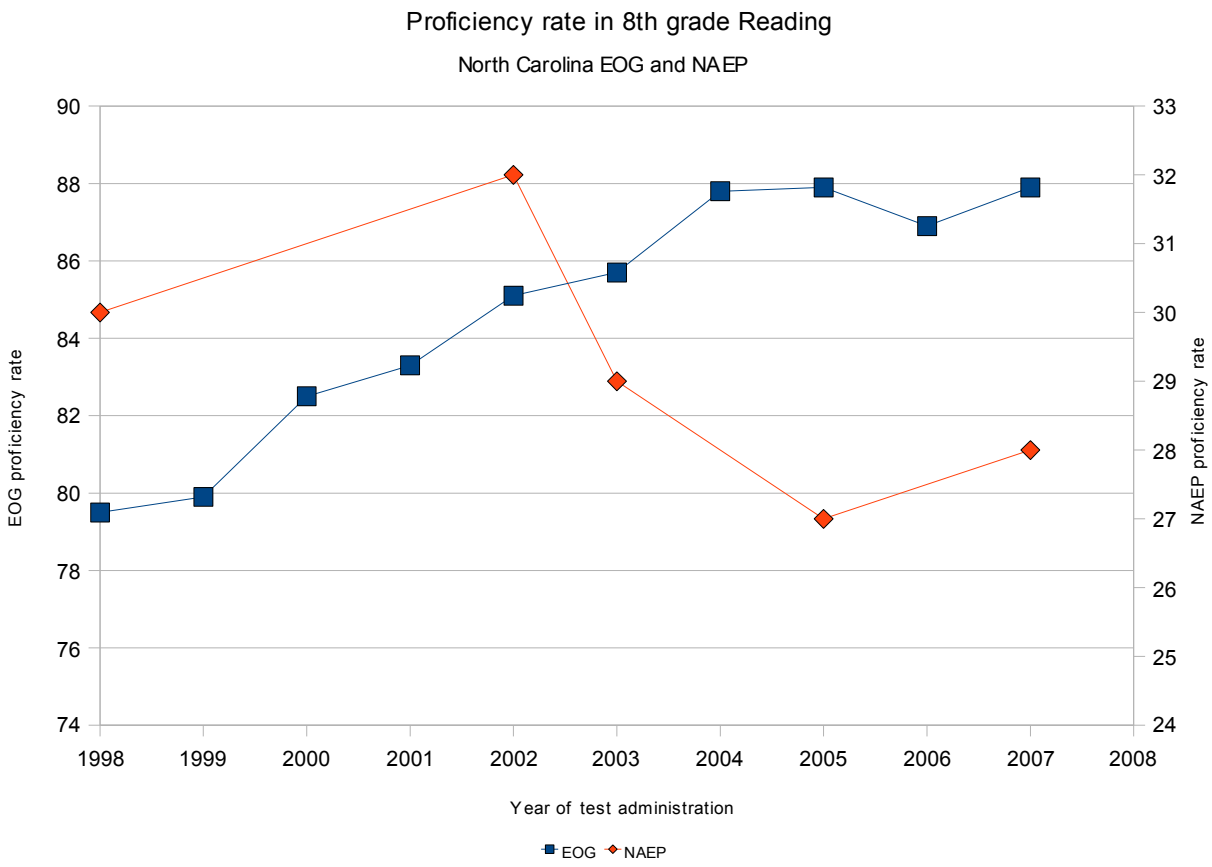

Proficiency rate in 8th grade Reading
North Carolina EOG and NAEP
EOG proficiency rate
NAEP proficiency rate
90
88
86
84
82
80
78
76
74
33
32
31
30
29
28
27
26
25
24
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
Year of test administration
EOG
NAEP

Figure 2

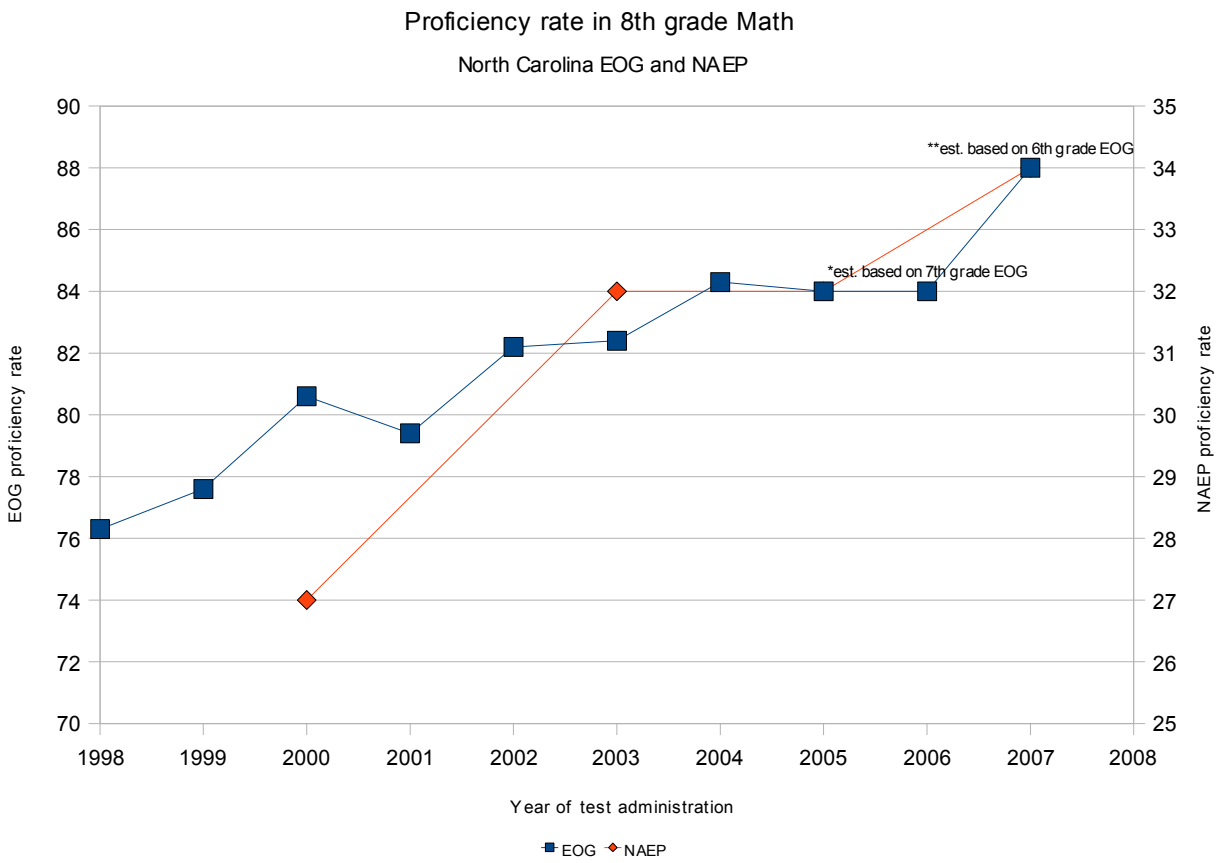
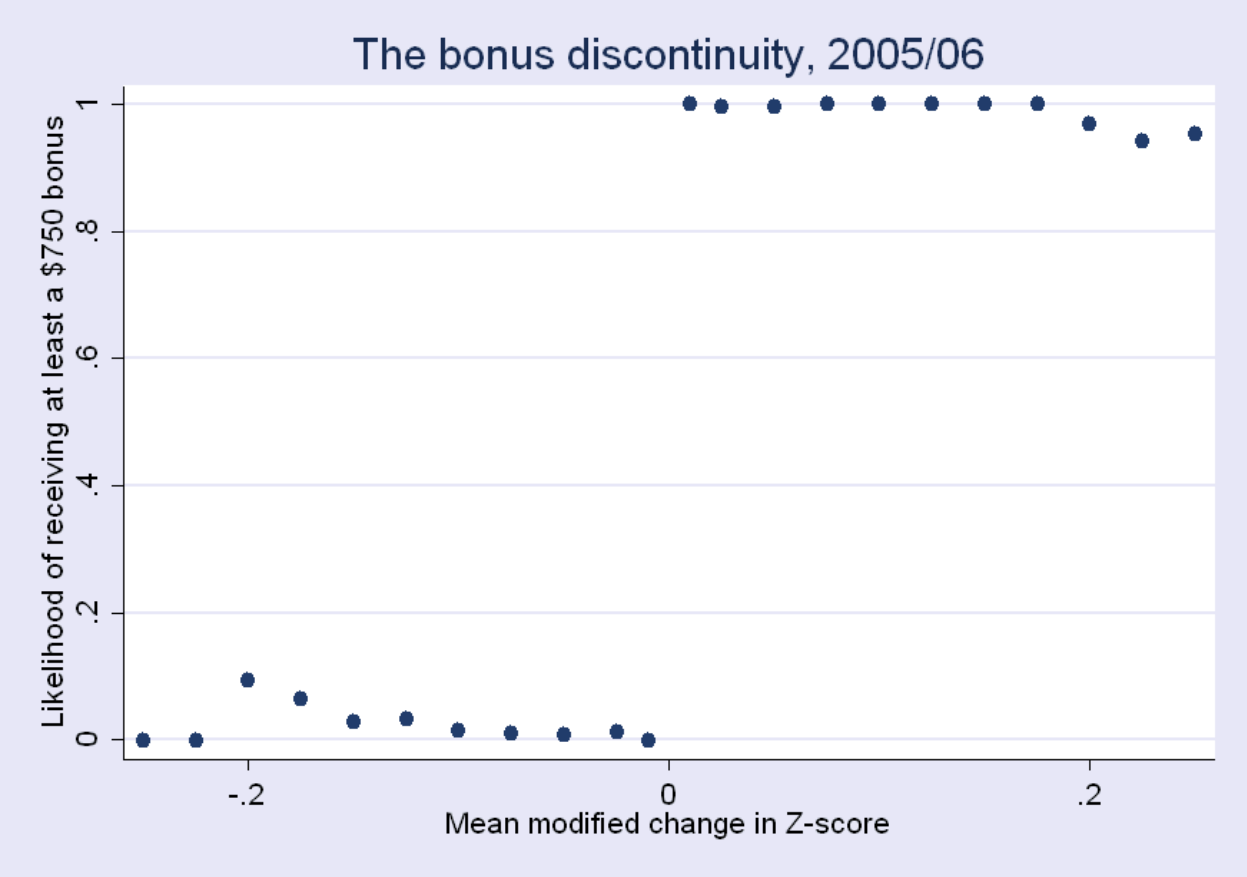

Proficiency rate in 8th grade Math
North Carolina EOG and NAEP
EOG proficiency rate
NAEP proficiency rate
Year of test administration
**est. based on 6th grade EOG
*est. based on 7th grade EOG
EOG
NAEP
90
88
86
84
82
80
78
76
74
72
70
35
34
33
32
31
30
29
28
27
26
25
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008

Figure 3


The bonus discontinuity, 2005/06
Likelihood of receiving at least a $750 bonus
0
.2
.4
.6
.8
1
-.2
0
.2
Mean modified change in Z-score
Figure 4

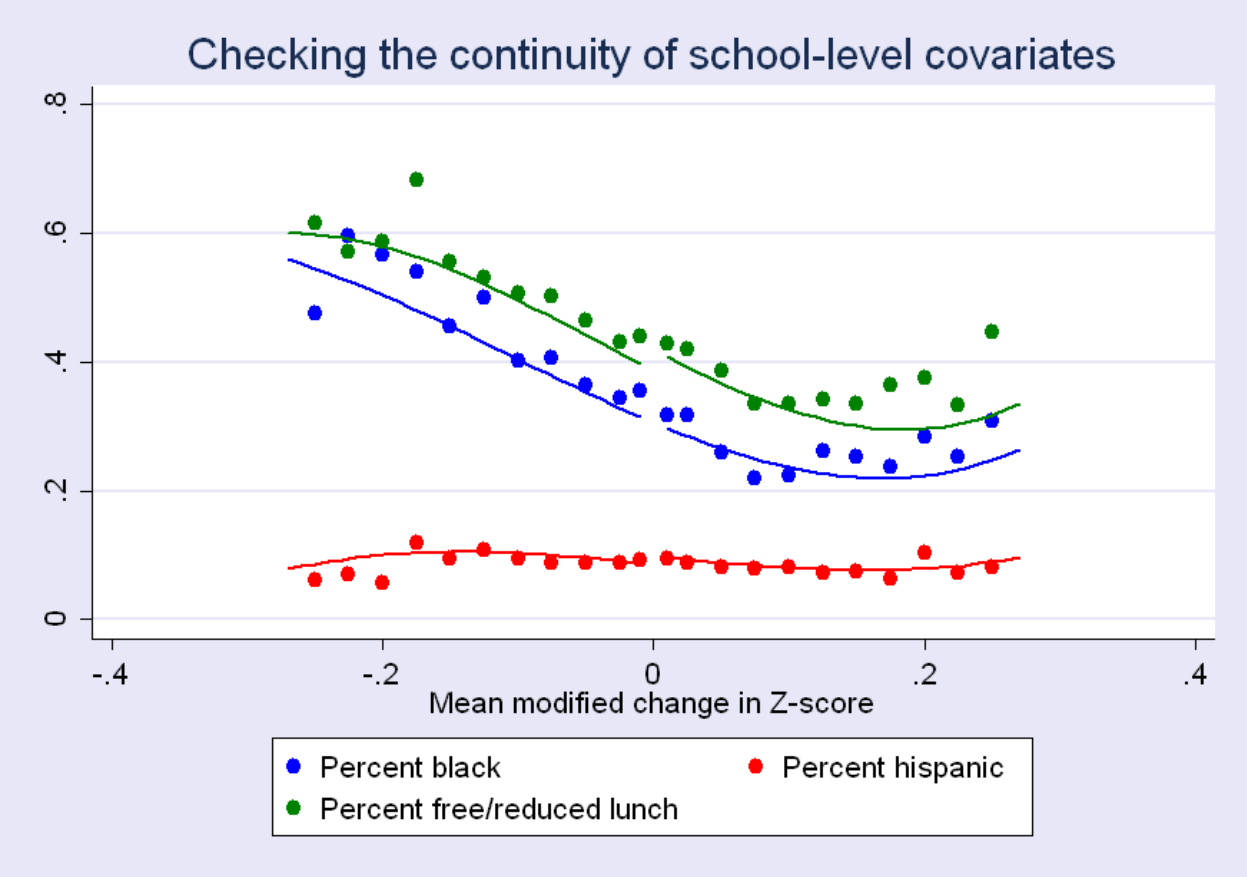
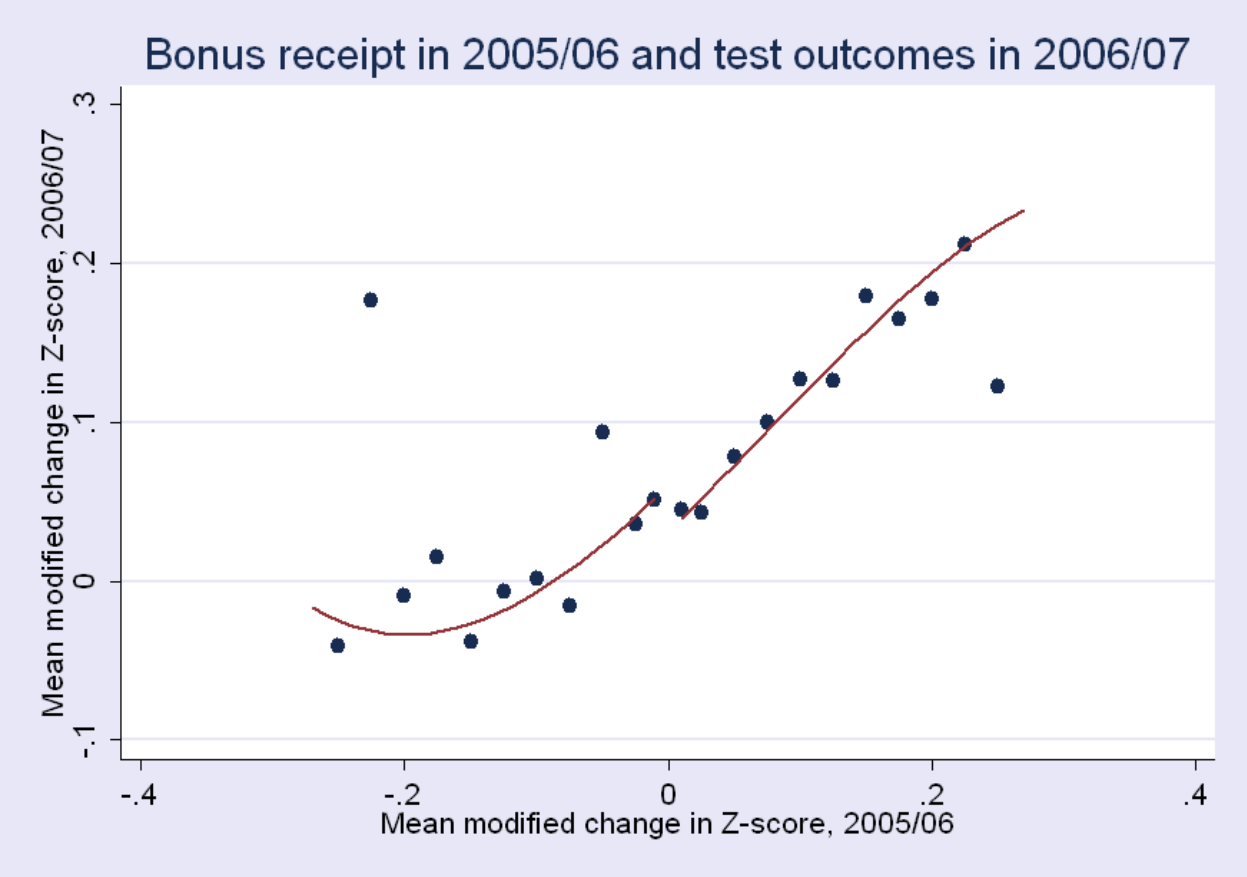

Checking the continuity of school-level covariates
0
.2
.4
.6
.8
-.4
-.2
0
.2
.4
Mean modified change in Z-score
Percent black
Percent hispanic
Percent free/reduced lunch

Figure 5


Bonus receipt in 2005/06 and test outcomes in 2006/07
Mean modified change in Z-score, 2006/07
-.1
0
.1
.2
.3
-.4
-.2
0
.2
.4
Mean modified change in Z-score, 2005/06
Figure 6

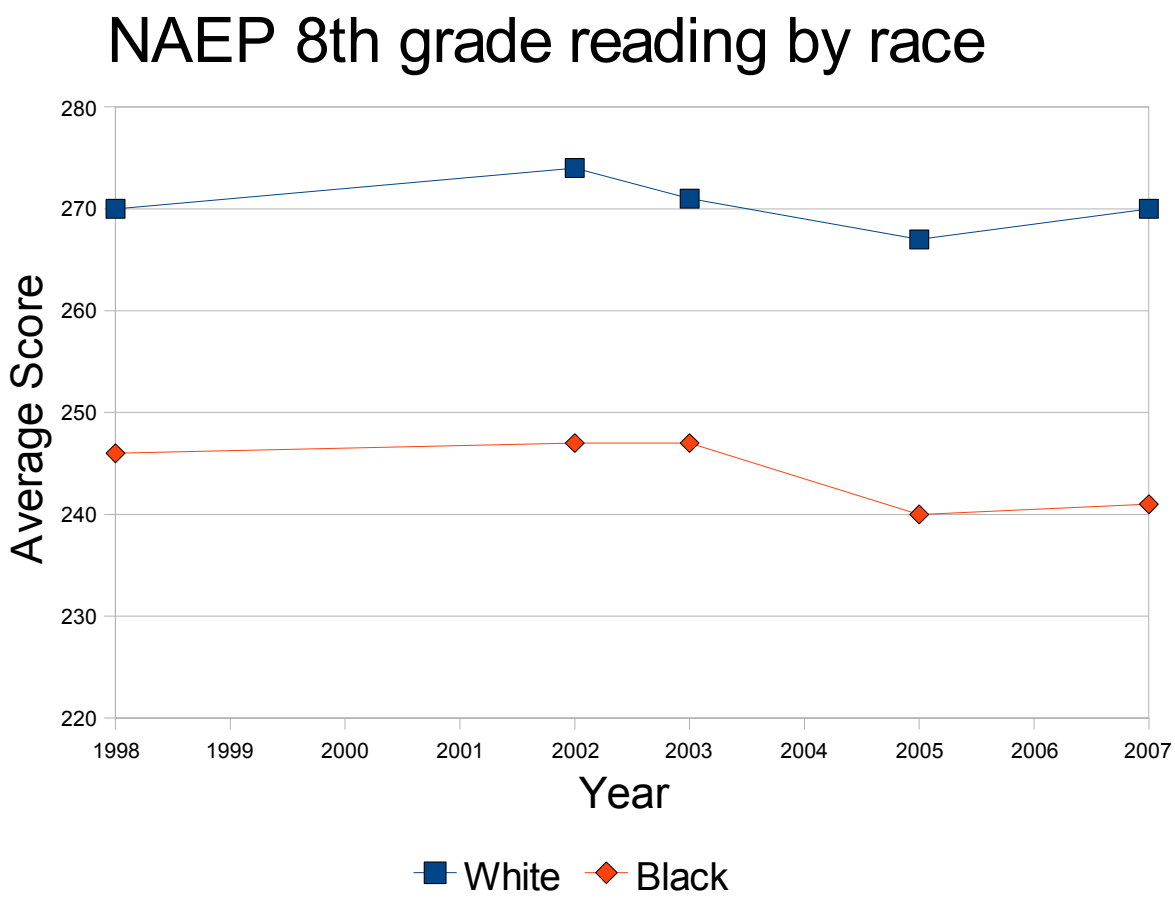

NAEP 8th grade reading by race
280
270
260
250
240
230
220
Average Score
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
Year
White
Black

Figure 7

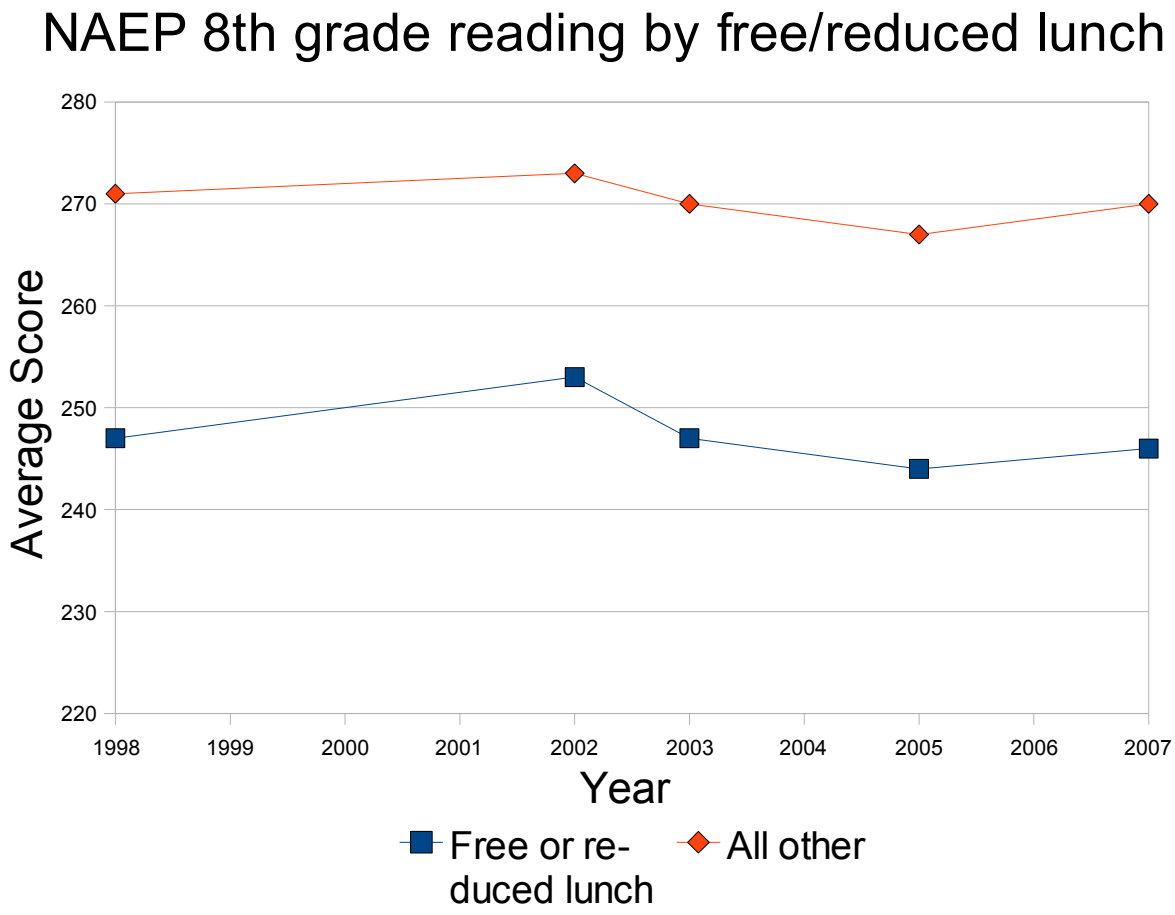

NAEP 8th grade reading by free/reduced lunch
Average Score
280
270
260
250
240
230
220
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
Year
Free or reduced lunch
All other

Figure 8

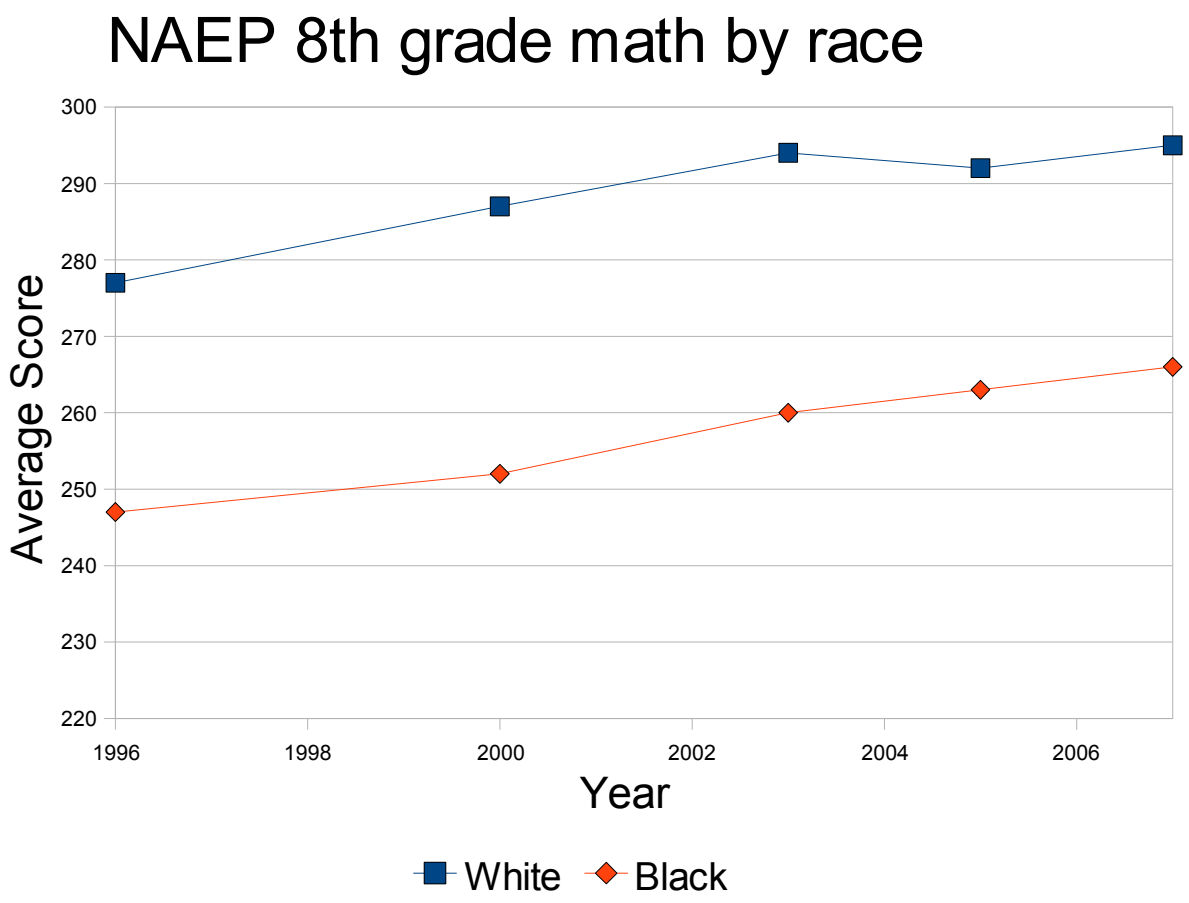

NAEP 8th grade math by race
300
290
280
270
260
250
240
230
220
Average Score
1996
1998
2000
2002
2004
2006
Year
White
Black

Figure 9

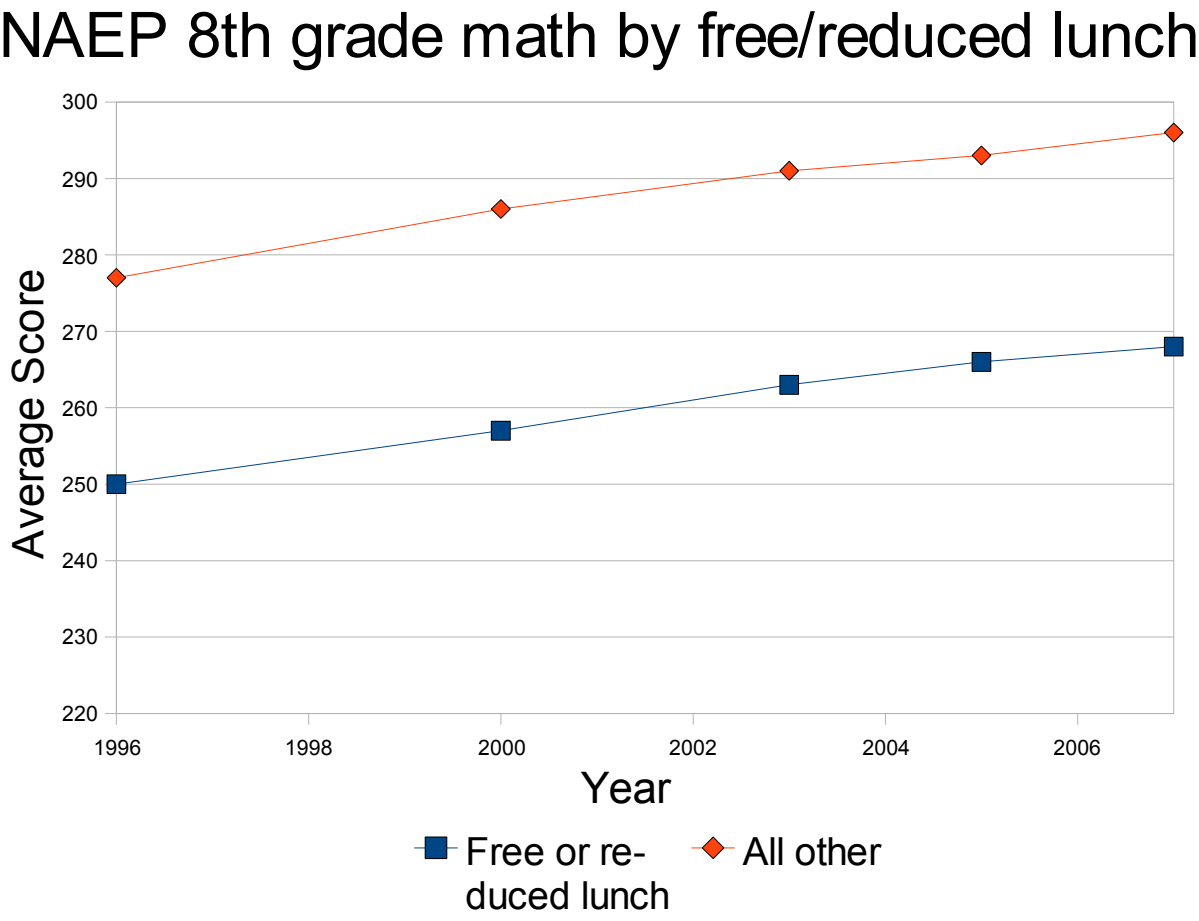

NAEP 8th grade math by free/reduced lunch
Average Score
300
290
280
270
260
250
240
230
220
1996
1998
2000
2002
2004
2006
Year
Free or re-duced lunch
All other

Figure 10

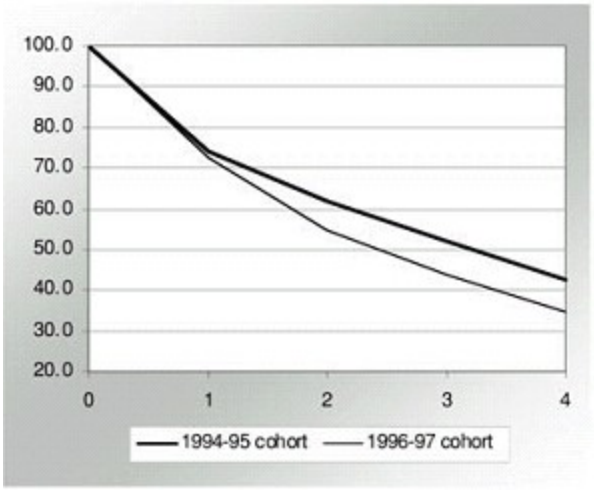


Figure 11: Comparison of teacher retention rates in low-performing schools, 1995 and 1997. Low performing schools are defined as schools with more than half of the students below grade level in the initial year. The horizontal axis refers to the number of years since the initial year for each cohort. Reprinted from Clotfelter, Ladd, Vigdor and Aliaga (2004).


CALDER