NBER WORKING PAPER SERIES

THE ACADEMIC ACHIEVEMENT GAP IN GRADES 3 TO 8

Charles T. Clotfelter
Helen F. Ladd
Jacob L. Vigdor

Working Paper 12207
http://www.nber.org/papers/w12207

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
May 2006

The authors wish to thank Robert Malme, Leona Christy, Patten Priestley, Aaron Hedland, and Marco Hernandez for excellent research assistance, David Aman for useful information on test administration in North Carolina, Frank Levy and an anonymous referee for helpful suggestions, workshop participants at the Russell Sage Foundation, the New School, CUNY, Amherst, Columbia, NYU, and the University of Chicago for valuable comments, and the Spencer Foundation for financial support.

The Academic Achievement Gap in Grades 3 to 8
Charles T. Clotfelter, Helen F. Ladd, and Jacob L. Vigdor
NBER Working Paper No. 12207
May 2006, Revised November 2007
JEL No. I21,J15

ABSTRACT

Using data for North Carolina public school students in grades 3 to 8, we examine achievement gaps between white students and students from other racial and ethnic groups. We focus on successive cohorts of students who stay in the state's public schools for all six years, and study both differences in means and in quantiles. Our results on achievement gaps between black and white students are consistent with those from other longitudinal studies: the gaps are sizable, are robust to controls for measures of socioeconomic status, and show no monotonic trend between 3rd and 8th grade. In contrast, both Hispanic and Asian students tend to gain on whites as they progress through these grades. Looking beyond simple mean differences, we find that the racial gaps in math between low-performing students have tended to shrink as students progress through school, while racial gaps between high-performing students have widened for black and American Indian students.

Charles T. Clotfelter
Sanford Institute of Public Policy
Box 90245 Duke University
Durham, NC 27708
and NBER
charles.clotfelter@duke.edu

Helen F. Ladd
Sanford Institute of Public Policy
Box 90245 Duke University
Durham, NC 27708
hladd@pps.duke.edu

Jacob L. Vigdor
Terry Sanford Institute of Public Policy
Box 90245 Duke University
Durham, NC 27708
and NBER
jacob.vigdor@duke.edu

The Academic Achievement Gap in Grades 3 to 8

Charles T. Clotfelter, Helen F. Ladd, and Jacob L. Vigdor
Duke University

I. Introduction

No topic goes to the heart of American concerns about equity in K-12 education more than the racial achievement gap. In 2004 Secretary of Education Rod Paige (2004) stated: "The academic achievement gap is the major driver of racial inequity in this country," and commentators across the political spectrum have expressed alarm over its size and consequences.[1] The racial gap itself has been a major impetus for federal education policy as embodied in the No Child Left Behind Act of 2001, and has entered into countless state and local debates regarding school finance equalization, academic tracking, and school testing and accountability programs.

Ever since the Coleman Report in 1966, researchers have known that average achievement test scores of black students lag well behind those of white students, but this issue has been taken up with renewed energy in the last decade. Efforts to describe and explain the achievement gap include studies of differences across cohorts (Lee 2002; Perie et al. 2005) and the progress of individual cohorts in the early school years (Fryer and Levitt 2004, 2005; Murnane et al. 2006) or in later years (Hanushek and Rivkin 2006). These studies show a large and persistent gap in achievement test scores between white and black students, but they differ

[1] For example, Jencks and Phillips (1998, p.3-4) state, "if racial equality is America's goal, reducing the black-white test score gap would probably do more to promote this goal than any other strategy that commands broad political support"; Thernstrom and Thernstrom (2003, p. 235) write, "Closing the skills gap is the key to real racial equality in American society." See also NAACP Legal Defense and Educational Fund (2005).

regarding the size of the gap at the beginning of school, how much of the gap is explained by socioeconomic status, and whether the gap widens as students advance through school grades.

This paper contributes to existing literature by studying racial/ethnic achievement gaps, based on tests that are linked to the state's standard course of study, exhibited by five consecutive cohorts of North Carolina public school students as they progressed from 3^{rd} to 8^{th} grade. We confirm some findings in existing literature: test-score gaps between black and white students are sizable, even after controlling for several important student covariates; Hispanic and American Indian students also lag behind whites, albeit less dramatically; and Asian students often score higher than whites, particularly in later grades. Unlike other studies of which we know, we find that the regression-adjusted mean test scores of minority groups other than blacks rise relative to whites as students age, at least over the range of grades that we observe. The regression-adjusted black-white gap in math test scores declines by less than 1% between 3^{rd} and 8^{th} grade but grows by about 11% in reading.

Mean differences in test scores can be misleading if the test score distribution is truncated or compressed at the high or low end. Mean differences are also potentially sensitive to nonlinear transformations of the underlying variable. Although we show that the first concern is not particularly relevant in North Carolina, we present additional evidence on measures of test score disparities that are invariant to monotonic transformations of the underlying variable. The proportion of black students scoring below the median of the white distribution, in both reading and math, fluctuates within a narrow range between 79 and 83% as students progress from third to eighth grade. The proportion of Hispanic students scoring below the white median declines from 71% to 66% in both reading and math over the same interval. Thus our main conclusions regarding central tendency are not sensitive to monotonic transformations in the test score scale.

Looking beyond measures of central tendency, we find that the trend towards improved relative minority test scores in math is concentrated at the low end of the distribution. Disparities in math scores between disadvantaged minority groups and whites shrink at the bottom of the test distribution as students progress through school, but in most cases actually increase at the top. The test score distributions for black and American Indian students become more compressed relative to the white test score distribution over these grades. We suggest a possible explanation for this phenomenon – the possibility that predominantly minority schools have redistributed resources toward lower-performing students in response to policy incentives – but leave hypothesis tests to subsequent work.

Previous research on the racial achievement gap is summarized in section II of the paper. Section III describes our data. Section IV shows the size of racial achievement gaps in North Carolina applying to different groups, before and after adjusting for socioeconomic differences among students. Section V looks beyond mean differences to examine other aspects of achievement distributions, and section VI concludes the analysis.

II. Previous Research on the Racial Achievement Gap

Research studies have examined racial and ethnic gaps in achievement using various groups, but by far the most attention has been paid to the black-white gap.

The Black-White Gap

Starting in the 1990s researchers used nationally representative samples to document the extent of and change in racial and ethnic gaps in achievement test scores. Several studies examined data from the National Assessment of Educational Progress (NAEP). As summarized by Phillips and Chin (2004, pp. 468-470), this research showed that the black-white gap at the 4th grade narrowed during the 1970s and into the 1980s, after which it stagnated or grew again slightly. As of 2000, this gap was 0.90 of a standard deviation in math and 0.83 in reading, as shown in Table 1. The corresponding gap for 8th graders was more than a full standard deviation (1.06) in math and 0.85 in reading.[2]

In a pair of studies using the Early Childhood Longitudinal Study (ECLS), Fryer and Levitt (2004, 2005) find a gap of 0.66 in math and 0.40 in reading at the beginning of kindergarten, as shown in Table 1. Over successive grades, however, these gaps grew. In contrast, Murnane et al. (2006) find larger initial gaps when they examine a very similar sample of students, but they do not find any growth in the gaps with grade level. As explanation for those conflicting findings, these authors point to differences in the type of tests used in the two studies: in comparison to the more general test used by Murnane et al., the one used by Fryer and Levitt was more closely aligned to items that are learned in school, thus giving students from all family backgrounds a more or less equal start in kindergarten.[3]

Three previous studies of achievement gaps rely on administrative data from school systems. Using data from Pasadena, Bali and Alvarez (2004) find black-white gaps roughly 0.55 s.d. and 0.35 in math and reading at 1st grade and about 0.50 and 0.45 at 4th grade.[4] Hanushek

---

[2] Precise gaps provided by Meredith Phillips. From this point forward, the units in which achievement gaps are measured will be standard deviations unless otherwise indicated.

[3] Whereas the test used in the ECLS sample examined by Fryer and Levitt emphasizes skills learned in school, the test used in the Murnane et al. study stresses a wider set of skills.

[4] These figures were derived by dividing the reported gaps, measured as test scores, by the standard deviation of the 4th grade reading test, the only standard deviation reported in the article. See Bali and Alvarez

and Rivkin (2006) and Stiefel et al. (2007) employ much larger data sets. The former study, covering the entire state of Texas and examining only math achievement, finds black-white gaps of 0.70 in 3rd grade and 0.76 in 8th. The latter, covering New York City, finds slightly smaller raw gaps in math, and quite a bit smaller adjusted gaps for reading in grade 5 than Murnane et al.

Since socioeconomic characteristics such as income and parental education tend to be correlated with race, it is likely that at least a portion of the observed gaps between racial and ethnic groups can be accounted for by non-racial factors. The research on achievement gaps has sought to determine just how large this portion is, in part because it is the portion of the currently observed gap that presumably will wither away over time as socioeconomic differences recede.[5] In their study covering grades K-3, Fryer and Levitt (2005) find that adding a small set of controls – including age, gender, birth weight, mother's age at first birth, and indicators of socioeconomic status – reduces estimated black-white gaps by more than half and actually eliminates the pure racial component at the beginning of kindergarten, after which it grows at the rate of about a tenth of a standard deviation through 3rd grade. Murnane et al. (2006) reach a different conclusion, based, as noted above, on a different sort of achievement test. When they control for socioeconomic status and other student covariates, they find a relatively constant racial gap, a bit smaller than one standard deviation, in both kindergarten and 5th grade.

Gaps between Whites and Other Groups

Hispanic students nationwide now comprise a larger minority group than African

---

(2004, Tables 2 and 3).

[5] For an analysis that discusses the relationship between changes in the achievement gap and long-term socioeconomic differences, see Krueger, Rothstein and Turner (2005).

Americans. Although the historical circumstances and policy issues may differ between these two groups, the issues related to measuring the test score gap with whites and adjusting the gap for socioeconomic differences are quite parallel. As a general matter, the size of the Hispanic-white gap tends to be smaller than the black-white one. In their analysis of the NAEP, for example, Phillips and Chin (2004) find gaps on the order of 0.70 standard deviations on both math and reading at the 4th grade level (compared to the 0.90 and 0.83 black-white gaps in math and reading, respectively). At 8th grade, the corresponding Hispanic-white differences were nearly 0.90 and 0.80, smaller than the black-white gap in math but about the same as the black-white gap in reading. In their 4th grade calculations, Bali and Alvarez (2004) find Hispanic-white gaps about half the size of the black-white one. Although smaller than the black-white gap, the Hispanic-white gap has been seen as stubbornly constant in recent decades.[6] Nor does research suggest any shrinkage in this gap as students progress through school. As a percentage of the corresponding black-white gap on the NAEP in 2004, the Hispanic-white gap was 74% as large at 4th grade and 88% at 8th grade (Perie et al. 2005, pp. 41-44). In their somewhat gloomy National Research Council study, Tienda and Mitchell (2006, pp. 82-85) report that this gap remains constant through elementary school, as Hispanic students suffer from disadvantageous home environments, teacher biases, and low motivation.[7]

In a similar study of students in publicly supported English schools, Wilson, Burgess and

[6] Lee (2002, Figures 2 and 3) shows that Hispanic-white gaps based on math and reading NAEP scores at three age levels have not declined since the late 1980s and the gap in SAT scores increased in the 1990s. Kao and Thompson (2003, p. 431) find that, after controlling for socioeconomic status, the achievement of white and Hispanic students was very close.

[7] They state (p. 85): "Weak relations with teachers diminish students' motivation to pursue academic work, and in turn lower teachers' expectations in a self-perpetuating cycle of academic disengagement and under-achievement."

Briggs (2005) examine the gaps in tests taken between ages 7 and 16, between white students and those from several ethnic minority groups. They find that raw achievement scores for whites exceeded those for most minority groups, there was little gap with respect to Indians, and whites were consistently behind Chinese students. When scores are adjusted for differences in socioeconomic status, both of these nonwhite groups consistently outperformed white students. As students progress through school, most nonwhite groups gained relative to white students over most years.[8]

III. Data

In this paper we analyze administrative data from North Carolina, one of the first and most prominent of the states to develop a mandatory statewide testing and accountability program. Having begun statewide testing in the 1970s and designed its own tests beginning in 1993, the state launched its accountability program, called the ABCs of Education, in the fall of 1996, which required the testing of all students in grades 3 to 8 and offered monetary rewards to teachers in high-performing schools.[9] The state is both large and ethnically diverse. In 2002 its public schools enrolled 1.3 million public school students, giving it the 11th largest public school enrollment among the 50 states (U.S. Department of Education 2005, Table 37). Its largest racial minority is African American, with these black students making up 31.4% of the state's public

[8] The adjusted gap measured in standard deviation units, for example, shows Pakistani students 0.22 behind whites at age 7 but 0.02 ahead at age 11; black Caribbean students remain behind by 0.42 and 0.19. By contrast, Indian and Chinese students are ahead of whites at both ages (Wilson, Burgess and Briggs 2005, Table 5).

[9] For further background on testing in North Carolina, see North Carolina State Board of Education(n.d). or Ward (n.d.).

school students in 2004/05. Another 7.5% were Hispanic, this share having grown rapidly in recent years. Asian students comprised 2.0% of the total. And, reflecting concentrations of Cherokee in the west and Lumbee in the east, another 1.5% of students were American Indian (North Carolina Public Schools 2005, p. 24).

The data used in the present study are derived from administrative records created by North Carolina's Department of Public Instruction and maintained by the North Carolina Education Research Data Center.[10] The state required all students to take standardized achievement tests in both math and reading at the end of every grade between 3 and 8. Using unique student identifying numbers that had been assigned randomly to individual students by the Data Center, we were able to match a student's records over time, making it possible to compare trends in student achievement using an intact sample of students. Not only does the data set provide information on test scores and many of the usual set of demographic variables, it also contains information on parents' education and the school district attended. To facilitate comparisons across years, we normalized the scaled scores for each test in every year over all students in the state who took the test so that each test would have a mean score of zero and a standard deviation of one. On this normalized scale, positive scores denote above-average performances relative to the state-wide average, and negative scores denote below-average performance. Because we have achievement test data for grades 3 to 8 spanning the school years 1994/95 to 2004/05, we were able to examine five cohorts of students, most of whom progressed

---

[10] For more information on the Center, see its Web page at: http://www.pubpol.duke.edu/centers/child/NC_Education_Research_Data_Center/nc_education_res_ctr.html, visited 12/21/05.

normally from grade 3 to 8.[11]

So that our results would not be influenced by the movement of students in or out of the state or to and from private schools, we examined a series of intact cohorts of public school students. For both the math and reading tests we included in each cohort all students who took the corresponding 3rd grade end-of-grade test in a given year and end-of-grade tests in each succeeding year.[12] Thus our results apply only to students who were enrolled in North Carolina public schools for six years, a fact that is particularly important with respect to Hispanic students, as we discuss in more detail in the next section. We have explicitly chosen to include in the sample students who repeated one or more grades, provided they appeared for six years. Leaving them out of the analysis would produce a seriously unrepresentative sample, in light of the high rates of grade retention among black and Hispanic students. In order to deal with students who fall behind in grade from those in their age cohort, we exploit the fact that standardized tests in North Carolina employ a developmental scale explicitly designed "to measure growth in skills and knowledge throughout the grades" (North Carolina State Board of Education 1996, p. 31). Based on this feature, our basic estimates express a retained student's test outcome in terms of

---

[11] Appendix Table 1 shows the structure of the data by year and grade. Note that the sample sizes for the reading cohorts (those students who had six successive years' data) are slightly smaller than those for the corresponding math cohorts. Limited English proficient (LEP) students could be excused their first year in the U.S. for the end-of-grade test in reading, but not in math.

[12] The last two lines in Appendix Table 1 show the size of each year's cohort for each of the two tests. These cohorts grew in size much more rapidly than the state's overall public school enrollment, which increased at an average rate of 1.8% a year (calculated from NCES, *Digest of Education Statistics 2004*, Table 37, http://nces.ed.gov/programs/digest/d04/tables/dt04_037.asp, 1/9/06), owing to a significant increase in the rate of matching students' records across grades made possible by the availability of names and social security numbers beginning with data for the 1995/1996 school year.

standard deviations from the mean of his or her former peers.[13] Because the scale scores may not have been successfully aligned as intended, however, we offer several alternative measures of the achievement gap, discussed below. By focusing on students who remained in the state's public schools for six successive years, we are necessarily examining an unrepresentative group. These students on average had higher scores in 3rd grade and their parents had higher educational attainment than those students who did not qualify.[14] But we believe the benefits from our focus on intact cohorts – most importantly, the exclusion of newly arrived immigrants from the sample – outweighs this drawback.

IV. Mean Differences in Achievement Across Racial and Ethnic Groups

We begin our analysis of racial achievement gaps by examining levels and differences in mean achievement level by grade, with no corrections made for location or socioeconomic differences. We then note results using several alternative measures of raw gaps. Finally, we present regression-adjusted estimates of mean differences.

Raw Achievement Gaps

Table 2 presents mean achievement gaps for five racial/ethnic groups in grades 3 to 8, based on the cohorts who took their 3rd grade tests from 1995 to 1999. Figure 1 shows these gaps

---

[13] For a student held back in 4th grade, for example, we calculated a normalized score by applying the mean and standard deviation of 5th grade scores in that year to the student's actual scale score on the 4th grade test. Appendix A discusses our treatment of students who were retained in grade, and Appendix Figure 1 illustrates differences in retention by race and ethnicity. Calculations based only on students who made normal progress through grades not surprisingly produce higher average scores, but the resulting racial gaps are similar in size to those based on entire cohorts, as illustrated for black and Hispanic students in Figures 3and 3b.

[14] See Appendix Table 2 for a comparison of students included and excluded from the 1999 cohort.

graphically for the four most numerous groups, highlighting patterns by grade level.[15] The gap is largest and most persistent for black students. In math, this gap began at 0.783 s.d. in grade 3 and ended at 0.814 in grade 8; in reading the gap rose from 0.710 to 0.776. For neither test was there a monotonic progression, and what trends exist may largely if not entirely reflect statistical noise. These black-white gaps are well within the range established by existing literature – very close to those reported by Stiefel et al. (2007) for grades 5 and 8, somewhat smaller than those obtained by Fryer and Levitt (2005) for grade 3 and Phillips and Chin (2004) for grades 4 and 8, slightly larger than those observed by Hanushek and Rivkin (2006) for math in grades 3, 5, and 8, and quite a bit larger than Bali and Alvarez's (2004) 4th grade gap based on a single school district. It is instructive to compare our calculated black-white gaps, based on intact cohorts, with gaps based on data for all black and white students for whom data are available each year. As shown in Figure 3a (and Appendix Table 3a), the black-white gap based on such repeated cross-sections closely tracks those based on intact cohorts. Students of both races who moved in and out of the public schools generally did worse than those with stable enrollment patterns, but the black-white gap was not materially affected by including them.

For Hispanic students the pattern of gaps is quite different. Compared to black students, Hispanic students started with a smaller gap in 3rd grade and then reduced that gap further as they progressed into middle school. In math, they began an average of 0.516 standard deviation behind whites, and by 8th grade they were just 0.384 behind, a reduction of 25%. In reading, the

---

[15] Students were observed in six successive years. As explained in the Appendix A, those few who indicated different racial/ethnic groups were assigned to the group most often cited. Except in cases involving the multiracial category, we dropped any student who cited two groups an equal number of years. When the multiracial group was one of the groups in a tie, we assigned the student to the other racial group if the other group was cited in at least three years. Because of this asymmetric treatment, we give less attention to the gap for multiracial students.

gap started out larger than in math; between 3rd and 8th grade this gap fell by almost 30%, from 0.562 to 0.397 standard deviation.[16] Crucial to this result is our restricted focus on students who were in the public schools for six consecutive years. This restriction not only leaves out Hispanic students whose families moved out of North Carolina after the 3rd grade year, but, more importantly, it also leaves out those who moved into the state after 3rd grade. Reflecting the steady flow of immigrants into North Carolina over the years covered by our data, this second group of Hispanic students was numerous, and they tended to have lower scores than the Hispanic students in our cohorts. Thus an achievement gap based on repeated cross-sections would be larger than those we calculate based on intact cohorts and would grow rather than shrink with each grade. As shown in Figure 3b, Hispanic-white achievement gaps based on repeated cross-sections were slightly larger at 3rd grade than those based on cohorts. In later grades, however, the gap based on repeated cross-sections increased markedly (19% in math and 31% in reading) while that based on our cohorts *decreased*, as noted above.[17] This stark contrast illustrates why the pattern of raw gaps for Hispanic students shown in Table 2 differs so profoundly from the impression given by the NAEP and other repeated cross-section calculations.

Thus, the marked improvement we observe among Hispanic students relative to whites is highly dependent upon our decision to focus on unchanging cohorts of students. Those Hispanic students who began in 3rd grade and stayed in the state's public schools for five more years tended to learn English if they were not fluent before and proceeded to improve relative to whites

---

[16] Fryer and Levitt (2005, p. 3) cite a similar convergence.

[17] See Appendix Table 3b for detailed measures of these gaps.

in both math and English.[18] Our results serve as a warning not to obscure the promising gains being achieved by established Hispanic students by combining them with the lower scores of newly arrived immigrants – precisely the result that emerges from exclusive reliance on periodic snapshots such as those provided by the NAEP-based national report cards.[19]

Asian students made steady progress from grade to grade relative to whites. By 8th grade, the average Asian student had a math score more than a third of a standard deviation above that of the average white student and was slightly ahead (by 9.5% of a standard deviation) in reading. This pattern is remarkably similar to that found by Burgess and Wilson (2005) for Chinese and Indian students in English schools. The American Indian students in North Carolina remained more than half a standard deviation below white students in both tests. Multi-racial students experienced gaps smaller than those for disadvantaged minority groups, and these gaps remained rather stable as students progressed in school.

As noted above, our approach in assigning normalized scores to students who fall behind in grade is based on the assumption that the underlying scale scores are directly comparable across grades. This assumption will be violated if the scale scores on the state's tests for successive grades are not vertically equated so as to constitute a single measuring rod from grade to grade. To assess how sensitive our results are to this assumption, we made calculations using two alternative approaches that do not rely on this assumption. The first alternative is simply to omit from the sample all students who failed or skipped a grade. Although this approach avoids

---

[18] To illustrate, the percentage of the 1999 cohort of Hispanic students classified as limited English proficient fell from 37.9% in 1999 to 17.3% in 2002.

[19] Strikingly, other unchanging cohorts of students who first appear in the 4th and 5th grades reveal larger initial Hispanic-white achievement gaps but the same pattern of shrinking gaps in succeeding years.

the necessity of assigning a score to students who are behind in grade, it presents a distorted picture of racial achievement gaps because it does not reflect the differential retention rates by race, as noted above. As illustrated in Figures 3a and 3b, gaps based only on non-repeaters are smaller than those for our cohorts, but the patterns of change over the six grades are very similar.[20]

The second alternative uses a different assumption: that grade repeaters would have scored below the median for their race had they made normal progress.[21] Under this assumption, we do not actually need to observe these counterfactual test scores so long as we focus on comparing medians in the test score distribution, rather than means. While the median is a conceptually distinct measure of central tendency, it is also interesting in its own right.[22] Gaps based on medians, where grade repeaters are assigned arbitrary scores below the median for their racial/ethnic group, are shown for the black-white gap in Figure 3a and for the Hispanic-white gap in Figure 3b.[23] Median gaps are uniformly larger in all grades for both races in both subjects, implying that the test score distributions of whites are differentially skewed as compared to those

---

[20] These differences are shown in detail in Appendix Tables 3a and 3b.

[21] This assumption would be suspect if the proportion of students repeating a grade approached 50%. Appendix Figure A1 shows that more than 80% of students of any given race make normal progress from grade 3 to 8 over six years.

[22] Since virtually all grade repeaters obtain scaled scores below the median for their race anyway, an analysis of median gaps using this assumption is basically identical to an analysis of median gaps under the original assumption that scale scores are comparable across grades.

[23] Appendix Tables 3a and 3b report calculations for all of these alternative measures for black and Hispanic students. A third alternative measure addresses the sensitivity of our calculations to the possibility that the standard deviations of actual scores do not reflect the true variation in ability. We thus calculated gaps assuming a constant standard deviation of 10 scale points. (This contrasts with calculated values that ranged over the grades and years of our sample from 8.3 to 11.9.) The gaps measured using this metric, shown in Appendix Tables 3a and 3b, likewise retained the same pattern as our basic calculations.

of nonwhite groups. In section V below we confirm that the white test score distribution generally has a longer lower tail than the black test score distribution, while the black test score distribution has a longer upper tail. This difference is exactly consistent with a narrower gap in means than in medians.

In the Hispanic-white case, median gaps track mean gaps quite well, diminishing as students progress from 3rd to 8th grade, but black-white median gaps follow different patterns from mean gaps in some respects. In math, the median black-white test score gap shows more year-to-year fluctuation than the mean gap, and has a more noticeable downward trend. In reading, the median gap lacks the downward trend shown in the mean gap. These differences in the variability and trend between mean and median test scores imply that the relative shapes of the black and white test score distributions are changing over the course of grade progression. If the mean of one distribution increased relative to the other, but all other moments of the distributions remained the same, we would not observe these differences. In section V below, we show that the black test score distribution, particularly in math, becomes more compressed relative to the white test score distribution as students progress from 3rd to 8th grade. Given the skewness in the initial distribution, this compression has a stronger effect on the mean than the median.

Regression-adjusted Achievement Gaps

Because students in various ethnic and racial groups typically differ in other ways that are systematically associated with achievement levels, such as socioeconomic status, researchers have sought to control statistically for such factors so as to isolate the component of achievement gaps purely related to racial or ethnic category.[24] The controls used in these studies are of two types: measures of students' personal and family characteristics and descriptive measures of schools and teachers. A principal reason for including variables of the latter type is to explore the extent to which observed gaps may be due to quality differences in schools attended by white and minority students. In this paper, we use measures only of the first type, leaving for separate analysis the examination of how schools and teachers affect achievement gaps. If such socioeconomic measures were uncorrelated with school quality, unbiased estimates of adjusted achievement gaps could be obtained simply by including such measures in regressions explaining student achievement. However, there exists ample evidence that children of better educated and more affluent parents tend to be taught in schools with teachers who have better credentials and more experience.[25] To mitigate the likelihood of omitted variable bias stemming from such correlations, we estimate regressions explaining normalized achievement test scores, using as regressors indicator variables for each racial group other than white, a set of other student covariates for which we have data, and school fixed effects.[26] This list of variables includes gender, age in the spring of the 3rd grade, parental education, eligibility for free or reduced price lunch, and indicator variables signifying type of district and region within the state.[27] To obtain

[24] Appendix Table 5 illustrates for the 1999 cohort how measures such as parental education and free lunch status differ across the racial and ethnic groups used in this study.

[25] See, for example, Clotfelter, Ladd, Vigdor, and Wheeler (2007).

[26] To be sure, school fixed effects may "overcorrect" for differences in school quality in that they will tend also to correct for differences in neighborhood or community attributes.

[27] Shown in Appendix Tables 4a and 4b, the 12 estimated regressions in sets of six for math and reading scores, one per grade, pool observations from the five cohorts and restrict the estimated effect of covariates to be equal across all racial groups. Regressions estimated with only two racial groups at a time, which effectively relax this restriction, produce similar results. Regressions estimated separately for each cohort show little evidence of

gaps that are adjusted for differences in personal and family covariates but not school quality, we hold constant the effects of covariates but allow mean school fixed effects to differ by racial/ethnic group, thus eliminating only the effects of the covariates from the raw gaps.[28] This approach also allows us to decompose the raw gaps into three parts. One portion represents the effect of differences in student covariates, and is simply the difference between the raw and adjusted gaps shown in Figures 1 and 2. The adjusted gap can be further divided into the portion attributable to school fixed effects and the portion that is otherwise unexplained by included variables other than racial/ethnic indicators.

Figure 2 shows estimated gaps adjusted for differences in student covariates. Not surprisingly, this statistical correction reduces the size of the black-white gap. For math, that adjusted gap averages about half a standard deviation, roughly three-fifths the size of the raw gap, which averaged about -0.80 s.d. over the six years covered (shown in Table 2). Similarly, the adjusted gap in reading averages somewhat less than half a standard deviation, compared to an average of about three quarters without adjustments. Although reduced in size, therefore, the adjusted gap between white and black students remains sizable. While there is no discernible trend over the grades in the reading gap, the math gap does decline slightly, by 0.08 s.d. The decomposition noted above is shown in Table 3. For example, it shows that the total black-white gap for math in 3rd grade (-0.783) can be divided into three parts: -0.231 due to differences in student covariates, -0.021 due to school fixed effects, and -0.530 otherwise unexplained by

meaningful differences in the level or trend of any racial achievement gap.

[28] Using estimated coefficients from regressions with school fixed effects to calculate gaps in mean predicted scores by group when all covariates are set to zero.

included variables. As is evident, the school fixed effect is relatively small for both tests over all grades, averaging just 7% of the total raw math black-white gap and 6% of the raw reading gap.[29] Our findings suggest that variation in school quality and other aspects of communities correlated with school attendance does not account for a large part of the total black-white gap in North Carolina, though it does grow in importance over the grades we study.

For Hispanic students, the effect of the statistical adjustment for covariates is striking: by 5th grade these students were on a par with whites in both math and reading. By 8th grade, adjusted scores for Hispanic students in the state surpassed those of observationally equivalent whites by roughly a tenth of a standard deviation. These gains relative to whites are parallel to those observed above in the raw gaps, only they begin with a smaller gap. Thus, once income and educational background differences are taken into account, Hispanic students in North Carolina were outperforming whites by the end of middle school.[30] Two features of the analysis explain this striking result, and serve to qualify it. First, the estimated equation underlying the adjustment includes both family income and parental education, two characteristics on which Hispanic students differ markedly from whites.[31] Second, to repeat the point emphasized above,

---

[29] The small contribution of school fixed effects in the present study is similar to the findings of Fryer and Levitt for 3rd grade (2005, Tables 1 and 6) but considerably smaller than those implied by Hanushek and Rivkin (2006, Table 3) for Texas and Stiefel et al. (2007, Table 3) for New York City. Since public schools in Texas are characterized by higher rates of segregation than those in North Carolina (Orfield and Lee 2004, Tables 11 and 14), the large difference in Texas could be attributable to large differences in community characteristics or school quality between schools attended by black and white students.

[30] Little if any of the relative improvement by Hispanic students over the six grades can be attributed to their improved economic standing. A year-by-year comparison of subsidized lunch rates for the 1999 cohort, the only one for which information was available in every year, shows the white-Hispanic gap falling slightly (from 51.4 to 49.1%) between 3rd and 5th grade and then remaining constant thereafter.

[31] See Appendix Table 4 for a comparison of mean values by racial/ethnic group.

our focus on intact cohorts of Hispanic students necessarily ignores newly arrived immigrants and their likely lower average achievement scores. Decomposing the Hispanic-white gap shows that, averaging over all six grades, school fixed effects explain only 6% of the total gap in math and 4% in reading, suggesting again that differences in school quality play a minor role in explaining achievement gaps in North Carolina.

For Asian students, the statistical adjustment has only a small effect on the size of the gap, because of their greater socioeconomic similarity with whites. The resulting estimates show Asian students surpassing whites on both tests and in all years except 3rd and 4th grade reading. For American Indians, the adjustment markedly reduces the gap with whites, cutting it by more than half in most grades, owing to the large differences in income and education levels between white and Indian families in the state. In this case school fixed effects play a larger role than with other disadvantaged minority groups, accounting for 12% of the raw gap in math and 18% in reading. This difference could reflect the comparatively high level of segregation for American Indian students in the state: more than half of those students attend schools in two relatively small counties (Robeson and Swain). The regression-adjusted estimates also show stronger trends towards convergence for this group. The adjusted gap between multiracial students and whites similarly shows convergence and is reduced in overall size.

Although the evidence of convergence for most disadvantaged minority groups conditional on observable characteristics might be construed as promising, it necessarily implies that the importance of certain other characteristics in determining test scores must be increasing over time. Indeed, the estimated equations in Appendix Table 4a and 4b show that the gap between poor and non-poor students (classified here according to their eligibility for subsidized lunches in 7th or 8th grade) increases marginally as students get older.[32] This increase might mean that a portion of the racial gap is morphing into an economic gap as students age. Why a similar change does not emerge from the parental education coefficients, however, is unclear.

A second factor showing increasing importance over time is student age. Students who are old relative to their cohort perform worse than their younger counterparts, and this gap grows over time. Disadvantaged minority students show a very slight tendency toward being older than whites (see Appendix Table 5), possibly because they are more likely to be retained prior to entering our analysis as 3rd graders. Thus, while the pattern of convergence in test score gaps is encouraging in some respects, researchers and policy-makers should continue making efforts to understand the widening gaps along other dimensions.

V. Racial Gaps at Other Points in the Achievement Distribution

Much of the existing literature on racial achievement gaps focuses on differences in raw or regression-adjusted means. While the mean is certainly a useful, intuitive statistic for measuring differences between racial and ethnic groups, there are several reasons to pay attention

---

[32] We considered the possibility that this result might be due to measurement error, but this appears unlikely. As noted in the text above, we used eligibility for free or reduced price lunch in grades 7 or 8 because those were the only two grades for which we had such information for all five cohorts. If families move in and out of poverty in years when students are in these grades, our measure could be a less precise indicator of poverty for students in early grades than when they were in grades 7 and 8. A consequence would be classic errors-in-variables that would bias the coefficient of the subsidized lunch indicator in the earlier grades toward zero, which could in turn raise the estimated coefficient on the race/ethnicity indicators to the extent they were correlated with income level.

To test for this possibility, we examined data for the 1999 cohort only (the only cohort for which we had subsidized lunch information for each year). We found that our grade 7-8 measure of subsidized lunch eligibility is indeed more highly correlated to contemporaneous eligibility (0.66 for 3rd grade math versus 0.86 for 8th grade math, for example). This variation appears to exert little bias, however. We estimated regressions of the form of those in Appendix Tables 4a and 4b, comparing the results when the indicator for actual contemporaneous eligibility for subsidized lunches was substituted for our indicator based on status in grades 7 and 8. We found no significant difference between any of the pairs of estimated coefficients for the subsidized lunch or any of the race/ethnicity indicators.

to other measures of racial disparities in test scores. Monotone transformations of a test score scale can have very large effects on differences in means or medians. Compression or truncation of the test score distribution at the high or low end can skew mean differences. Measures of central tendency, whether means or medians, can also obscure offsetting relative movements occurring at varying points in the test score distribution.

To address these concerns, Figure 4 and Tables 4a and 4b use a different methodology to analyze test score disparities between whites and other racial groups, a cross-referencing of percentiles. This method has the advantage of being entirely invariant to monotone transformations of the test score distribution. Figure 4 shows the proportion of black or Hispanic students with test scores at or below the median of the white distribution. Were there no racial disparity in test scores, we would expect 50% of each group to score at or below the white median. In both reading and math, the figure shows that four out of every five black students have test scores below the white median. These significant disparities remain roughly constant across grades, supporting the general conclusion that there is little net improvement or worsening of the black-white achievement gap as students age. As for Hispanic students, Figure 4 shows improvement relative to white students through the six grade levels, consistent with our findings above based on intact cohorts. The proportion of Hispanic students scoring below the white median in 3rd grade is about 71% in both math and reading. Six years later, the percentage if only 66%.

Tables 4a and 4b present similar statistics focusing at the upper and lower tails of the test score distribution. In both tables, each row tracks the experience of a single cohort of students between 3rd and 8th grade, in a single test score subject. To track trends at the lower tail of the achievement distribution, we report the proportion of students of a given race who score below the 10th percentile of the white distribution. Were there no racial disparity in test scores, we would expect 10% of students of any given race to fall below the white 10th percentile. Higher numbers indicate an over-concentration of minority students in the lower tail of the distribution. To track trends in the upper tail, we report the proportion of white students who fall above the 90th percentile of a given minority group. Again, we would expect a value of 10% in a world with no racial disparities. Higher numbers indicate that whites are concentrated at the high end of the test score distribution.

Table 4a reveals substantial and persistent disparities in math scores between blacks and whites, Hispanics and whites, and American Indians and whites, at each cohort and each point in time. The black-white disparities are largest. At any given point in time, more than a quarter of blacks test below the 10th percentile of the white distribution, and more than a third of whites score above the black 90th percentile.

Tracking the experience of individual cohorts over time, we find important differences at the low and high ends of the test score distribution. For each of the five cohorts tracked here, the proportion of black students testing below the white 10th percentile decreased over time, by between 2 and 4 percentage points. By contrast, the proportion of whites scoring above the black 90th percentile increased in four out of five cohorts. Thus, the black-white test score gap in math narrows at the low end of the distribution, but widens at the high end of the distribution. In other words, the black distribution becomes compressed relative to the white distribution. Similar, if not more dramatic, compression occurs in the American Indian test score distribution. Hispanic students show more consistent evidence of advancing relative to whites at both the high and low ends, but in four out of five cohorts progress is more rapid at the low end.

Figures 5a and 5b illustrate the compression of the distribution of black math scores relative to the white distribution, shown as kernal density plots of unnormalized scale scores based on the 1999 cohort. Figure 5a presents 3rd grade distributions, while Figure 5b shows the 8th grade distributions. In third grade, the two distributions have very similar peaks, at a density of 0.04. By the time these students reach 8th grade, the black test score distribution has a slightly higher peak, clearly above 0.04, while the white peak is clearly below 0.04. Over time, the white test score distribution transforms from one that is clearly skewed to one that is roughly symmetric. The black test score distribution, if anything, switches from having an elongated lower tail to having an elongated upper tail. This shift in distribution implies that the black mean increases more rapidly than the black median over time. This, in turn, explains why the black-white gap in median test scores expands relative to the black-white gap in mean test scores in Figure 3a. The black mean keeps pace with the white mean; the slower advance of the black median relative to the mean implies that it falls behind the white median.[33]

Asian students clearly do not fit the mold of other minority groups. In virtually every instance, Asian students start ahead of whites – with fewer than 10% of students starting below the white 10th percentile and fewer than 10% of whites scoring above the Asian 90th percentile – and advance still further between 3rd and 8th grade. Unlike other groups, there is no evidence that Asian advances relative to whites are concentrated at the low end of the test score distribution.

---

[33] The elongation of the upper tail referenced in the text explains the difference between the trend in Figure 3a and the steadiness observed in Figure 4. The black math median is not keeping up with the white math median, so the gap between medians increases. The elongation of the upper tail of the black distribution allows the black 80th percentile to keep up with the white median at the same time.

Table 4b presents the corresponding set of findings in the reading test score distribution. Once again, there are sizable and persistent gaps between blacks and whites, Hispanics and whites, and American Indians and whites. One-quarter or more of the black students in each cohort score below the white 10th percentile, and roughly one-third of white students in each cohort score above the black 90th percentile.

The only evidence of narrowing in the black-white reading test score gap is found at the high end of the distribution, where black advances relative to whites are observed in four out of five cohorts. At the low end of the distribution, disparities remain steady or increase, which is a direct contrast to the math test score distribution. Thus, in reading, the black test score distribution is widening, if anything, relative to the white test score distribution. Hispanic and American Indian students show more consistent evidence of advancing relative to whites between 3rd and 8th grade, at both the high and low ends of the distribution. Asian students, once again, are at or above white achievement levels in 3rd grade and improve relative to white students over the next five years, at both ends of the distribution.

Why do racial math test score gaps tend to close at the bottom end of the distribution and widen at the top end? Why don't reading test score gaps follow the same pattern? Although a full evaluation of these findings is beyond the scope of this paper, one reasonable hypothesis about the pattern for math is that it reflects efforts to meet standards imposed by school accountability programs, such as North Carolina's ABCs program and the federal No Child Left Behind Act, which assign disproportionate weight to low-performing students. These and other accountability programs include sanctions that punish schools where students fail to attain a minimum level of achievement. This emphasis on bringing all students up to a certain threshold may lead some schools to reallocate instructional resources away from high-performing children and towards low-performing children. If schools with high concentrations of low-achievers divert more resources away from high-performers than other schools, high-performers will be disadvantaged when they attend such schools. The observed erosion of the relative position of high-achieving black students might then reflect the tendency for these students to have more low-performing classmates than their white counterparts.

The absence of compression in the reading test score distribution at first seems to contradict this hypothesis, but it might reflect the marginally larger growth in the black-white gap over the grades in reading. The fact that the black-white math gap grows less than that in reading might reflect a more successful redistribution of instructional resources in math towards low-performing children.[34] We must, however, leave further analysis of this phenomenon to future work. In any case, the possibility that raising the test scores of low-performing students in math may come at the expense of scores of high-performing students merits additional research.

---

[34] A straightforward test of this hypothesis would be to compare cohorts educated before and after the implementation of North Carolina's accountability system, to see if the system's implementation is associated with a decline in test scores for high-performing students in schools with a high proportion of low-performing students. We are unable to test this hypothesis because our earliest cohort did not reach 8th grade before the implementation of accountability in North Carolina.

In considering the possible effects of accountability programs, it is instructive to ask how the mean achievement gaps changed over the five cohorts. To test for secular trends in these gaps, we estimated regressions with student covariates, interacting a linear time trend with indicators for each racial/ethnic group. The estimated coefficients for those interaction terms, shown in Appendix Table 6, indicate the average annual change in each gap at grades 3 and 8. Positive coefficients show improvement relative to white students, and negative ones show worsening gaps.

The table shows that the black-white gap gradually improved over the period in math at both grade levels and in reading at grade 8, but in no case changing by as much as 2% of a standard deviation per year. The gap for Hispanic students reveals a seemingly contradictory pattern, with a growing gap in math in grade 3 but a shrinking gap for reading only at 8th grade. For Asian students, the only statistically significant trend is in reading at 8th grade. The group showing the most consistent progress in relation to whites is American Indians, who reduced their gap in both subjects at 3rd grade and in math at 8th grade. The rate of improvement in 8th grade math for this group, 2.5% of a standard deviation a year, was about twice the rate experienced by black students and the rates of improvement in grade 3 were even higher. If continued, these rates imply that half the 3rd grade raw gaps with whites could be eliminated in a decade.

VI. Conclusion

Our analysis of North Carolina administrative data adds to the body of research on racial achievement gaps by focusing on several very large cohorts of students in an ethnically diverse state, by examining gaps with respect to students in four different minority groups, and by looking beyond the mean of the distribution to uncover a more complicated pattern. Because the cohorts contain students who remained in the state's public schools for six consecutive years, the results are not influenced by immigration, private school enrollment, or movements across state lines, any of which could change the composition of students in the public schools. To be sure, restricting ourselves to students who remain in the state and in the public schools necessarily limits the generalizability of our findings, but we believe this drawback is vastly outweighed by the advantages of observing the academic performance of intact groups of students over time. Only by comparing such intact groups can one identify how gaps change as students progress through school.

Like previous studies, we find large gaps in mean achievement between white and black students. In contrast to some of those studies, however, we find no appreciable growth in the gap as students progress through school. Other disadvantaged minority groups, namely Hispanics and American Indians, also display lower test scores than whites, but these gaps are uniformly smaller than the black-white gap and show evidence of dissipating as students age. Asian students surpass whites between 3rd and 8th grade. Our findings are also consistent with previous research in that socioeconomic factors explain a sizable portion of these racial test score gaps, for example, about a third of the black-white raw gaps. Indeed, these factors grow more important as students advance in school. Our work departs from previous research, however, in showing that the gap between cohorts of white and Hispanic students narrows markedly between 3rd and 8th grades.

In the case of the black-white gap in math, the relative stability of the gap in means masks two divergent trends in the tails of the achievement distribution. At the low end of the achievement distribution, the gaps between white and black students shrink. At the high end, however, test score gaps tend to increase by similar amounts as students age. This tendency may reflect any number of different factors, but one important question for further research is whether these divergent trends reflect tradeoffs that are being made in response to accountability programs, such as NCLB and North Carolina's ABCs program, designed to raise the achievement of those at the bottom of the distribution.

References

Bali, Valentia A. And R. Michael Alvarez, "The Race Gap in Student Achievement Scores: Longitudinal Evidence from a Racially Diverse School District," *The Policy Studies Journal* 32 (No. 3, 2004), 393-415.

Clotfelter, Charles T., Helen F. Ladd, Jacob L. Vigdor and Roger Aliaga Diaz, "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 (2003). pp.251-271.

Clotfelter, Charles T., Helen F. Ladd, and Jacob L. Vigdor, "Federal Oversight, Local Control, and the Specter of 'Resegregation' in Southern Schools," *American Law and Economics Review* 8 (Summer 2006), 1-43.

Clotfelter, Charles T., Helen F. Ladd, and Jacob L. Vigdor, "Teacher-Student Matching and the Assessment of Teacher Effectiveness," *Journal of Human Resources* 41 (Fall 2006), 778-820.

Clotfelter, Charles T., Helen F. Ladd, Jacob L. Vigdor, and Justin Wheeler, "High Poverty Schools and the Distribution of Teachers and Principals," *North Carolina Law Review* 85 (June 2007), 1345-1379.

Fryer, Roland G. and Steven D. Levitt, "Understanding the Black-White Test Score Gap in the First Two Years of School, *Review of Economics and Statistics* 86 (May 2004), 447-464.

Fryer, Roland G. and Steven D. Levitt, "The Black-White Test Score Gap through Third Grade," *American Law and Economics Review* 8 (Summer 2006), 249-281.

Hanushek, Eric A. and Steven G. Rivkin, "School Quality and the Black-White Achievement Gap," NBER Working Paper 12651, October 2006.

Jencks, Christopher and Meredith Phillips (eds.), *The Black-White Test Score Gap* (Washington: The Brookings Institution, 1998).

Kao, Grace and Jennifer S. Thompson, "Racial and Ethnic Stratification in Educational Achievement and Attainment," *Annual Review of Sociology* 29 (2003), 417-442.

Krueger, Alan, Jesse Rothstein, and Sarah Turner, "Race, Income and College in 25 Years: The Continuing Legacy of Segregation and Discrimination," NBER Working Paper 11445, June 2005.

Lee, Jaekyung, "Racial and Ethnic Achievement Gap Trends: Reversing the Progress

Toward Equity?" *Educational Researcher* 31 (January/February 2002), 3-12.

Murnane, Richard J., John B. Willett, Kristen L. Bub, and Kathleen McCartney, "Explaining Puzzling Patterns in Black-White Achievement Gaps," Harvard Graduate School of Education, January 2006.

NAACP Legal Defense and Educational Fund, *Closing the Gap* (Washington: NAACPLDEF, 2005). http://www.naacpldf.org/content/pdf/gap/Closing_the_Gap_-_Moving_from_Rhetoric_to_Reality.pdf

North Carolina Public Schools, *Statistical Profile 2005*. http://www.ncpublicschools.org/fbs/stats/statprofile05.pdf, 1/16/06.

North Carolina State Board of Education, *End-of-Grade Tests*, Technical Report #1, August 1996.

North Carolina State Board of Education, "History of the North Carolina State Board of Education," NC State Board of Education website, http://www.ncpublicschools.org/state_board/SBE_history/chapter3.html, 3/13/06.

Orfield, Gary and Chungmei Lee, *Brown at 50: King's Dream or Plessy's Nightmare?* Harvard Civil Rights Project, January 2004.

Paige, Rod, address to National Urban League, July 22, 2004. http://64.233.161.104/search?q=cache:ldWoiotQy64J:www.ed.gov/news/pressreleases/2004/07/07222004.html+Paige+calls+achievement+gap&hl=en, 12/22/05.

Perie, Marianne, Rebecca Moran, Anthony D. Lutkus, and William Tirre, *NAEP 2004 Trends in Academic Progress* (Washington: Department of Education, National Center for Education Statistics, July 2005). http://nces.ed.gov/nationsreportcard/pdf/main2005/2005464.pdf, 1/18/06

Phillips, Meredith and Tiffani Chin, "School Inequality: What Do We Know?" in Kathryn Neckerman (ed.), *Social Inequality* (New York: Russell Sage Foundation, 2004).

Public Schools of North Carolina, Website, *The ABCs Accountability Model*, http://abcs.ncpublicschools.org/abcs/, 12/23/05.

Public Schools of North Carolina, *Assessment Brief*, vol. 5, no. 3, March 1, 2004 , "Understanding North Carolina End-of-Grade Testing" http://www.dpi.state.nc.us/docs/accountability/testing/eog/AssessmentBriefEOG04.pdf, 12/23/05

Stiefel, Leanna, Amy Ellen Schwartz, and Ingrid Gould Ellen, "Disentangling the Racial Test Score Gap: Probing the Evidence in a Large urban School District," *Journal of Policy Analysis and Management* 26 (Winter 2007), 7-30.

Thernstrom, Stephan and Abigail Thernstrom, *No Excuses: Closing the Racial Gap in Learning* (New York: Simon and Schuster, 2003).

Tienda, Marta and Faith Mitchell (eds.), *Multiple Origins, Uncertain Destinies* (Washington: National Research Council, 2006).

Wilson, D., S. Burgess, and A. Briggs, "The Dynamics of School Attainment of England's Ethnic Minorities," unpublished paper, U. of Bristol, October 2005.

Williams, Valerie S.L., Mary Pommerich, and David Thissen, "A Comparison of Developmental Scales Based on Thurstone Methods and Item Response Theory," *Journal of Educational Measurement* 35 (Summer 1998), 93-107.

Williamson, Gary L., Charles L. Thompson, and Robert F. Baker, "North Carolina's Growth in Reading and Mathematics," paper presented at the annual meeting of the North Carolina Association for Research in Education, March 31, 2006.

U.S. Department of Education. 2005. *Digest of Education Statistics, 2004*. Web page: http://nces.ed.gov/programs/digest/d04/tables/dt04_037.asp, 12/21/05.

Ward, Martha S., "Policies and Standards, Their Role and Revision: The Case of Ethics in Testing in North Carolina," unpublished paper, n.d., http://www.natd.org/Ward96.htm,

c:\wp\Ach gap\achgapdraft80707

Appendix A
Detailed Data Description

The construction of our panel data sets involved forming cohorts of students who took six successive end-of-grade achievement tests. This appendix discusses the treatment of students who failed at least one grade, the characteristics of students not included in the cohorts, and the consistency of data for individual students.

Grade Retention

Any student who failed a grade necessarily did not take all tests in the same years as most of his or her peers. Simply eliminating these students would have yielded a seriously distorted sample, because rates of grade retention differ systematically by race. Figure A1 depicts the proportions of five racial/ethnic groups who progressed as expected from grade to grade. It shows that school careers with normal grade progression were markedly less common among black students than among white students, with Hispanic students falling between these groups. Five years after the 1998/99 school year, only 84% of the black students in the cohort had reached the 8th grade, compared to 89% of Hispanic and 94% of white students.[35]

Our approach to dealing with these repeaters is to base a student's achievement on the scale score received in a year regardless of the student's grade. More precisely, we computed for these students a score based on the student's actual test performance but standardized according to test performance in the student's "normal" rather than actual grade. For example, we applied the score a student obtained on the 3rd grade test after being retained in grade to the state distribution of the 4th grade scores to infer where that student would have fallen if he or she had taken the 4th grade test. We take this approach because it was the explicit aim of the state for its "developmental scale scores" to be used in just this way – "like a ruler that measures growth in reading and mathematics from year. Just like height in inches, the student's scores in reading and mathematics are expected to increase each year" (Public Schools of North Carolina 2004, pp. 1-2). In the early 1990s the state devised the tests by converting raw scores based on the number of correct answers on multiple choice tests at each grade level into a scale for each subject calibrated in part by comparing performance by students in adjacent grades on sets of identical questions, a process known as "vertical equating." The resulting scale "uses the same metric for student performance in different age groups or school grades, so that test performance may be compared across grade levels and growth may be assessed in terms of changes in average performance and variability from grade to grade" (Williams, Pommerich, and Thissen 1998, pp. 95-96, 93). As constructed in 1993, the mean scale score on the math end-of-grade tests increased from 139.9 in 3rd grade to 168.3 in 8th. On the reading tests the mean rose from 142.7 in 3rd to 158.7 in 8th (North Carolina State Board of Education 1996, Table 16, p. 37). Depending on the type of test and the grade level, the state's published tables assume that a

---

[35] As can be inferred from comparing the gender composition of groups A and B in the 1999 cohort shown in Table A1, boys fail grades with more regularity than girls. The proportion of boys reaching 8th grade in the normal five years among white students was 92% compared to 95% for girls. For black students the comparable rates were 80 versus 88%, and for Hispanic students they were 87 versus 92%.

student showing "consistent mastery" should see growth in scaled scores ranging from a fraction of 1% to about 3% per year. Although both the math and reading tests experienced at least one re-calibration over the period of our sample, they retained the same format, with scaled scores at a given mastery level showing gradual increases from one year to the next.

Our basic approach is strictly correct only to the extent that the intended aim of vertical equating was actually achieved. Despite the state's efforts, however, it is possible that the attempt to vertically equate scores was not entirely successful – that is, that every student would actually have achieved the same score by taking at the same time tests written for two different grade levels – simply because successive tests are not identical. To allow for the possibility that the scale scores were not, in fact, vertically equated, we present results based on several alternative means of assessing achievement gaps, as discussed in the text and illustrated in Appendix Tables 3a and 3b.

As noted in the text, simply omitting students who repeated a grade would affect the representativeness of the sample. This may be illustrated by examining the cohort of students who were 3rd graders in 1999. Appendix Table 2 gives sample sizes and mean values of several variables for four groups of students. The first three columns include three groups who took the 3rd grade test in 1999: those who made normal progress through all six grades, those in the cohort who repeated at least one grade, and those students who were excluded from the cohort because they were not present in the data set all six years. Of these 98,857 students who took the 3rd grade test in 1999, this last group numbered almost 20,000, or about a fifth of the total. Of the remaining 79,406 students – those qualifying for the cohort – almost a tenth had to repeat at least one grade. The table's last column covers the parallel group of students who were in a public school and took the math test in the last year, 2004, but who were not present all six years.

The contrast between the cohort's two groups, A and B, shows striking differences. Compared to those with normal grade progression, the grade repeaters (group B) averaged achievement scores in 3rd grade a full standard deviation below the normal-progress group. These repeaters were also disproportionately male (64%, compared to 49% among those with normal progress) and even more disproportionately black (52 vs. 29%). They were also more likely to be Hispanic or American Indian and less likely to be Asian. Their parents were markedly less likely to be college graduates and much more likely to have dropped out of school before receiving a high school diploma. The repeaters were also more likely to have family incomes low enough to qualify for subsidized lunches, and they were more likely to have been classified as exceptional in some way other than being deemed gifted. The sharp differences between groups A and B illustrate the importance of keeping the latter in the sample when examining the racial achievement gap.

Characteristics of Non-cohort Students

By comparison, the mean values for groups C and D shown in Appendix Table 2 suggest what kinds of students are omitted from the analysis by restricting our sample to those who are present in all of the six years. Looking at the first group of mobile students, the table shows that those in group C had slightly lower 3rd grade achievement than those with normal progress in the cohort (roughly a fifth of a standard deviation), were much less likely to have college-educated parents, but also less likely to have parents who were not high school graduates. The mobile students in group D, comprising mostly students who arrived in the state after 1999, differed most clearly in their larger share of Hispanic students (10.4%), with a correspondingly higher share of students who were ever classified as having limited English. One other feature marking both of these mobile groups is their high probability of being in one of the state's five largest school districts, a finding consistent with the high relative rates of growth and population turnover in the largest metropolitan areas.

Data consistency

We performed several checks to see how consistent the administrative data were over time. We were also interested in how students' free and reduced price lunch status changed. We report here results for the 1998 cohort, but other cohorts looked similar in these regards.

Three attributes that should not change over time are a student's gender, age, and race/ethnicity; yet occasional inconsistences do arise in all of these categories. Given the vagaries of data collection in a multi-layered state administrative structure, some inconsistencies are inevitable. When students change schools in the weeks before a test, for example, pre-printed forms with student information cannot be made available, so students themselves must provide information on gender, race, and date of birth, taking care to follow all instructions on how to fill in forms, all of which leads to occasional mistakes. Fortunately, such mistakes are relatively rare. In the case of gender, they are quite rare, with a mere 0.4% of the 1998 cohort having at least one mistake over the six years. In fact, only 0.03% of the sample had no modal gender. In these cases, for all but the 1999 cohort, staff of the North Carolina Education Research Data Center examined the names to determine probable gender. Where that examination proved inconclusive, and for all the cases in the 1999 cohort with no mode, gender was assigned randomly. Inconsistencies were more common in the reporting of birthdays, with 5.4% of the 1998 cohort having at least one inconsistency. In these cases the modal birthday was chosen. For the 0.14% of this cohort that had no mode, the midpoint was chosen.

The consistency of the race/ethnicity category is of special importance because of its centrality to the current analysis. The issue of racial categories has also assumed more general significance in light of the change in the census race categories in 2000 and the accompanying scholarly debate over racial classification itself.[36] North Carolina school records allowed for these six categories in all years covered by our data: non-Hispanic white and black, Hispanic, Asian, American Indian, and multi-racial. For purposes of reporting to the Department of Education, however, the state used only the first five of these. For our purposes, we classified students according to the group they were placed in four or more of the six years. For the vast majority of students, there was no ambiguity: of the students in the 1998 cohort, 99% had the same race designation in all six years. If students were classified as multi-racial for three years and another race in three years, we assigned the student to the latter group. Students for whom there was no majority designation were dropped from the analysis. Interestingly, consistency in racial designation was highest for white and black students. The percentage of students in the 1998 cohort who showed only one race category over the six-year period was: white, 99.86%;

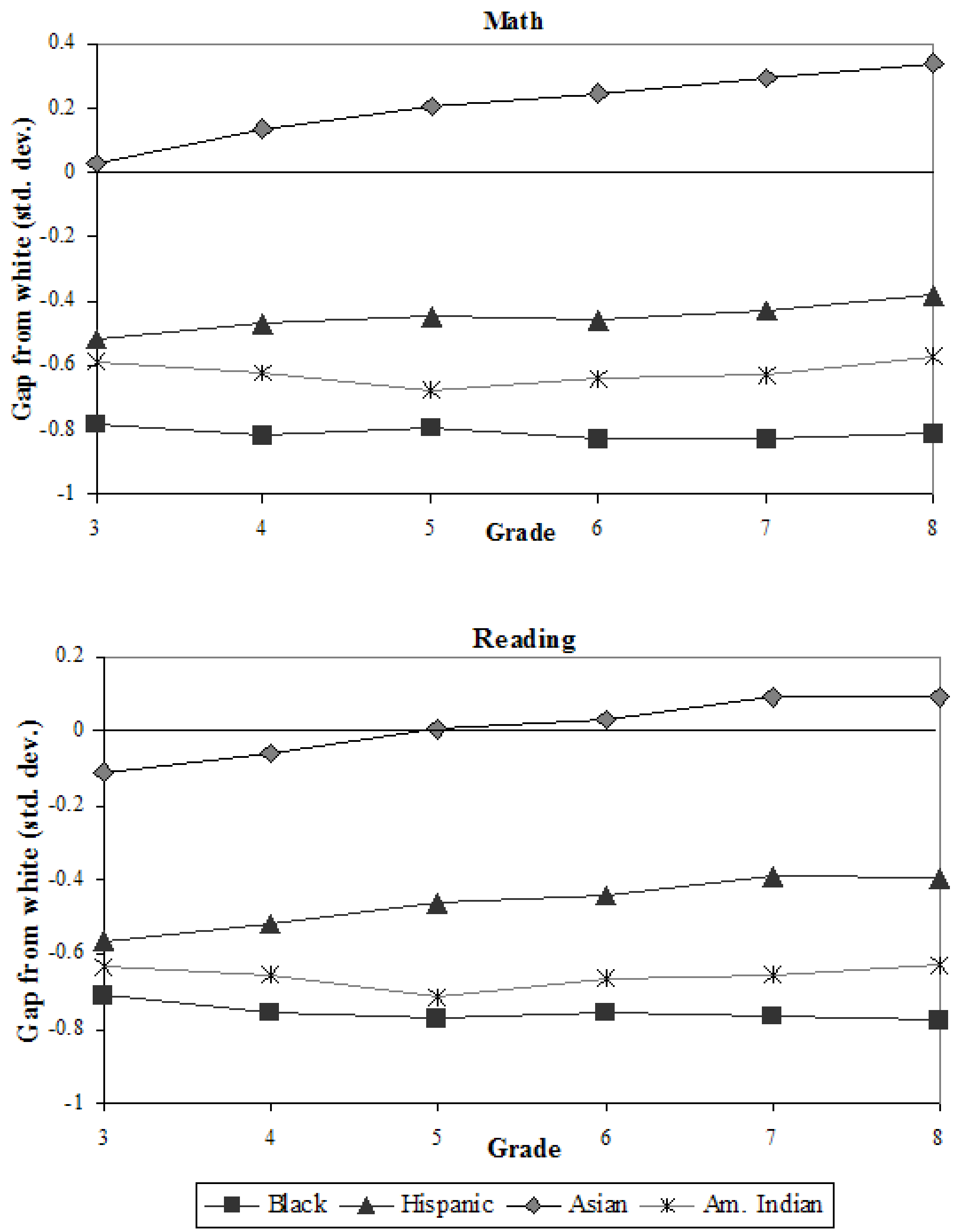
[36] See, for example, Perlmann and Waters 2002 and Barr (2005).

black, 99.67%; Hispanic, 94.6%; Asian, 94.6%; American Indian, 91.3%; and multi-racial, 57.0%.

Because information on which students were eligible for the free and reduced price lunch program was available only beginning in the 1998/99 year, we decided to define our measure of low family income based only on eligibility in a student's last two years, which for most would be their 7th and 8th grade years. This information was available equally for all five of our cohorts. Accordingly, we found that about 44% of students in the 1998 cohort were classified as low-income, by virtue of being eligible for free or reduced price lunch in either 2002 or 2003. To see how well this variable distinguished students, we compared these designations to information on eligibility in each year from 1999 to 2003. We found that three quarters of those designated low income by our definition in fact were eligible in at least four of the five years for which we could determine their status. On the other side, we found that only 16% of those not designated low income had ever qualified as eligible.

Our measure of parental education is based on information supplied by a student's teacher or counselor, who was asked to indicate the education level corresponding to "the highest level completed by either of this student's parents/guardians." To check for any general tendency to overstate or understate attainment by race/ethnicity, we compared the implied rates by group with those obtained in the 2000 census for the parents of students aged 8 to 11 in North Carolina. In general, the two distributions are very close to one another. Teachers and counselors were inclined to overstate the college completion rate of white and Asian parents (35.5% for whites compared to 30.9% from the census; 40.8% vs. 38.2% for Asians). But they tended to understate the proportion of black parents who had not completed high school (7.3% vs. 14.0% in the census). The tendency to overstate attainment for white and Asian parents could lead to an overestimate of the importance of parental education since their children have higher average achievement, but the tendency to overstate attainment for black parents would have the opposite effect.

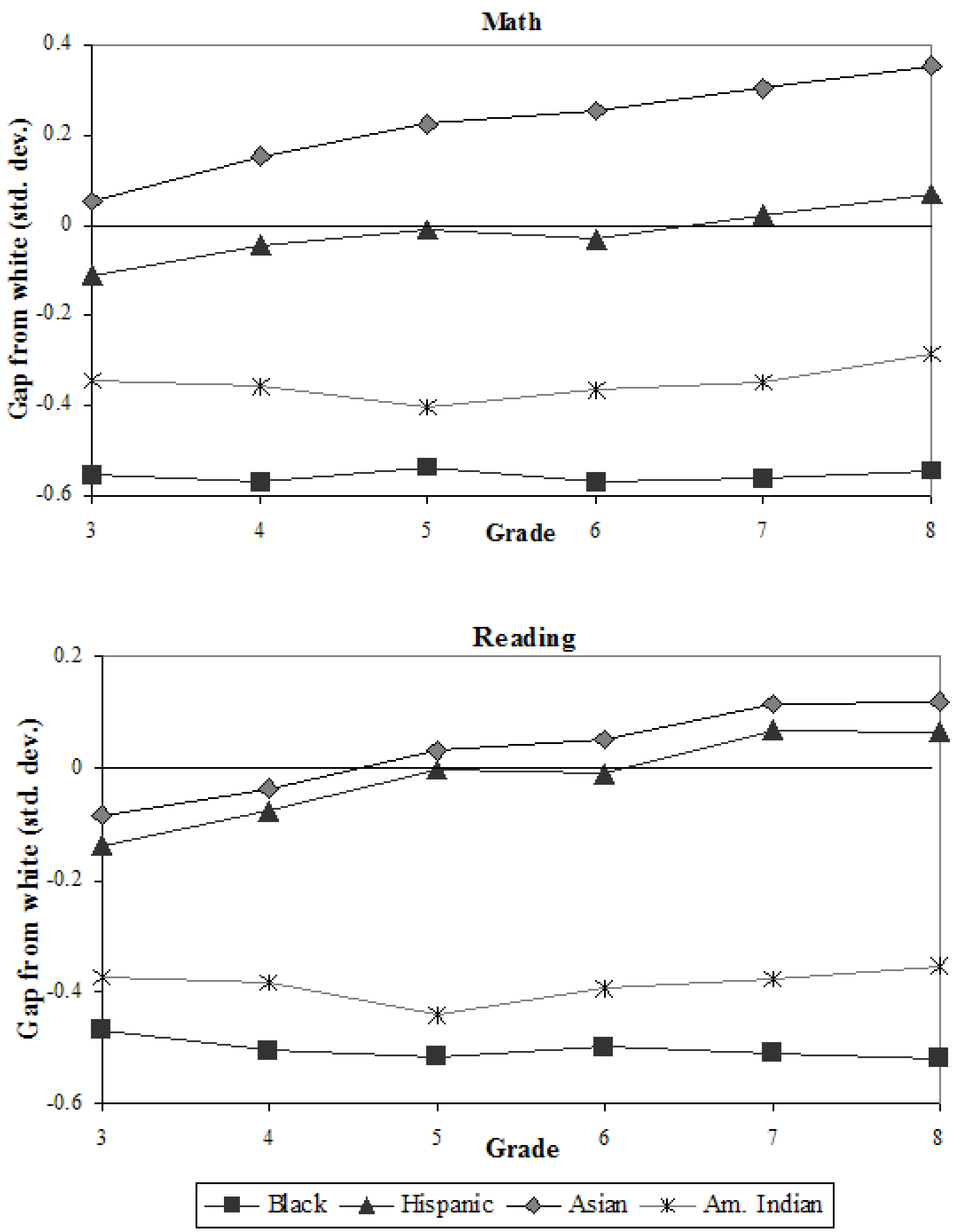
Figure 1. Raw Achievement Gaps, 1995-1999 Combined Cohorts, Math and Reading



Note: Achievement gaps are differences in mean normalized achievement scores, measured in standard deviation units, estimated in regressions combining five cohorts and containing indicators for racial/ethnic groups other than whites and for cohorts other than 1995. See Table 2.

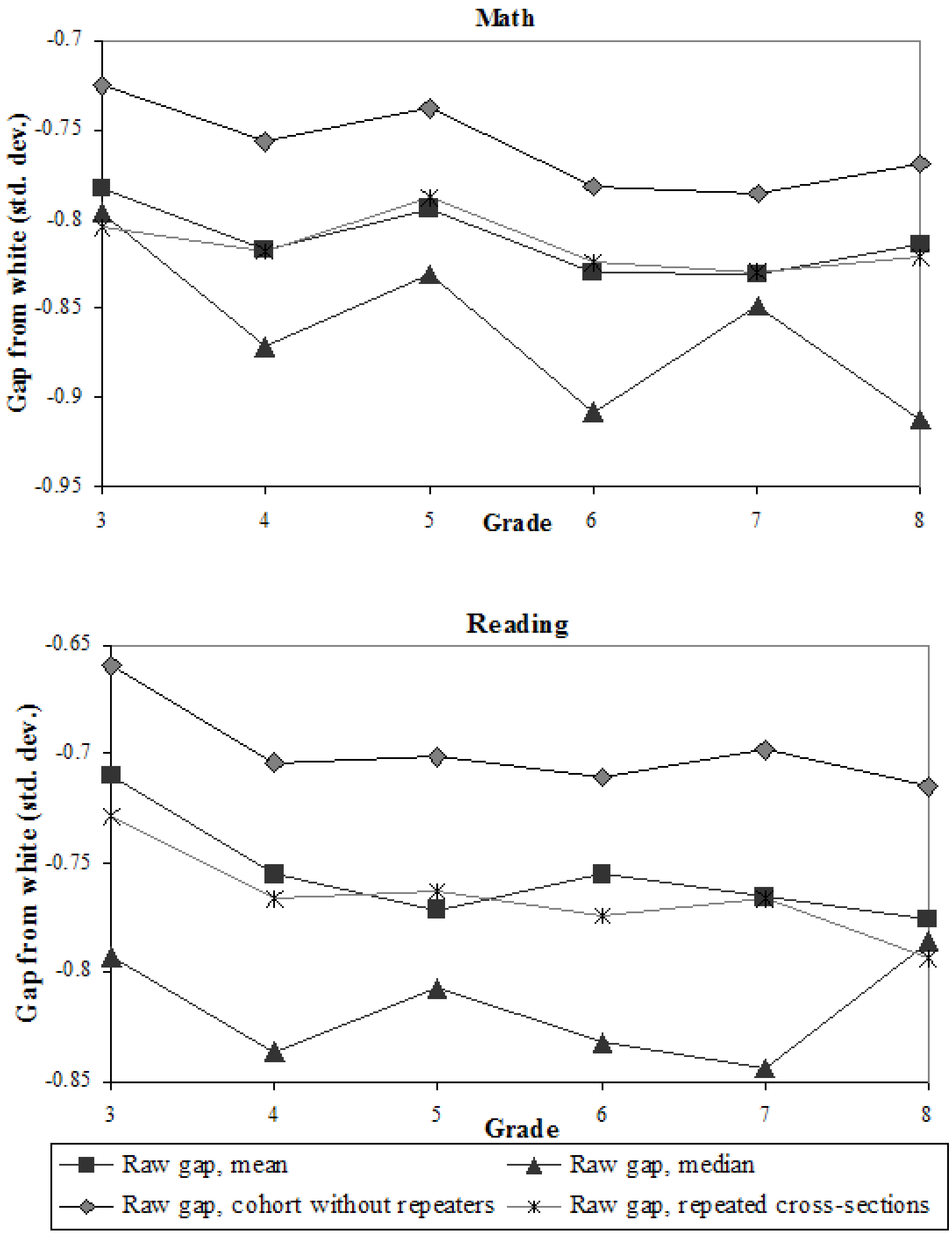
7/18/07

Figure 2. Covariate-adjusted Achievement Gaps, 1995-1999 Combined Cohorts, Math and Reading



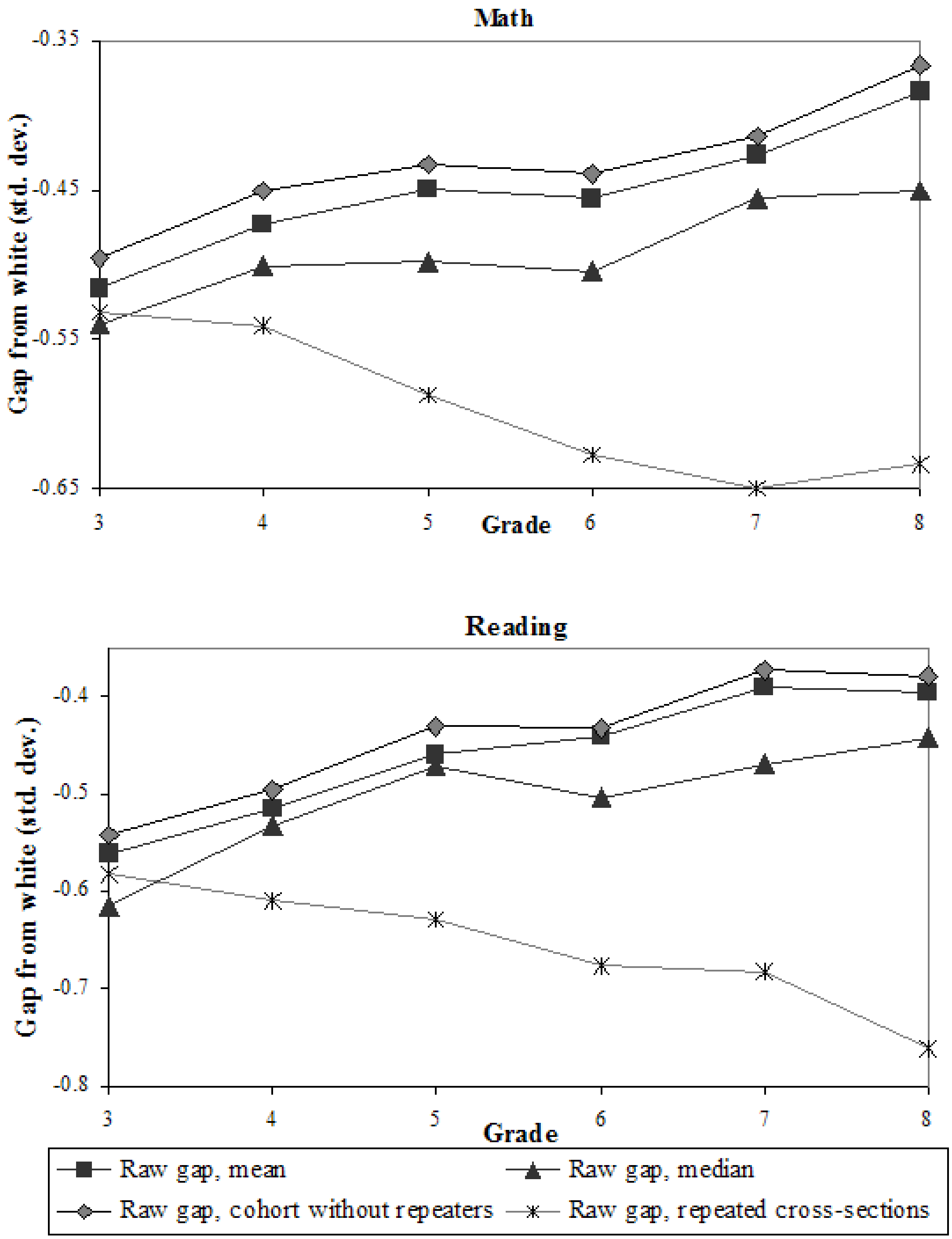
Note: Achievement gaps are measured in standard deviation units, estimated in regressions combining five cohorts and containing racial/ethnic indicators, other covariates, and school fixed effects. See text for method. See Appendix Tables 4a and 4b for estimated gaps.
7/18/07

Figure 3a. Black-White Raw Achievement Gaps, 1995-1999 Combined Cohorts, Math and Reading, Alternative Measures



Source: Appendix Table 3a.

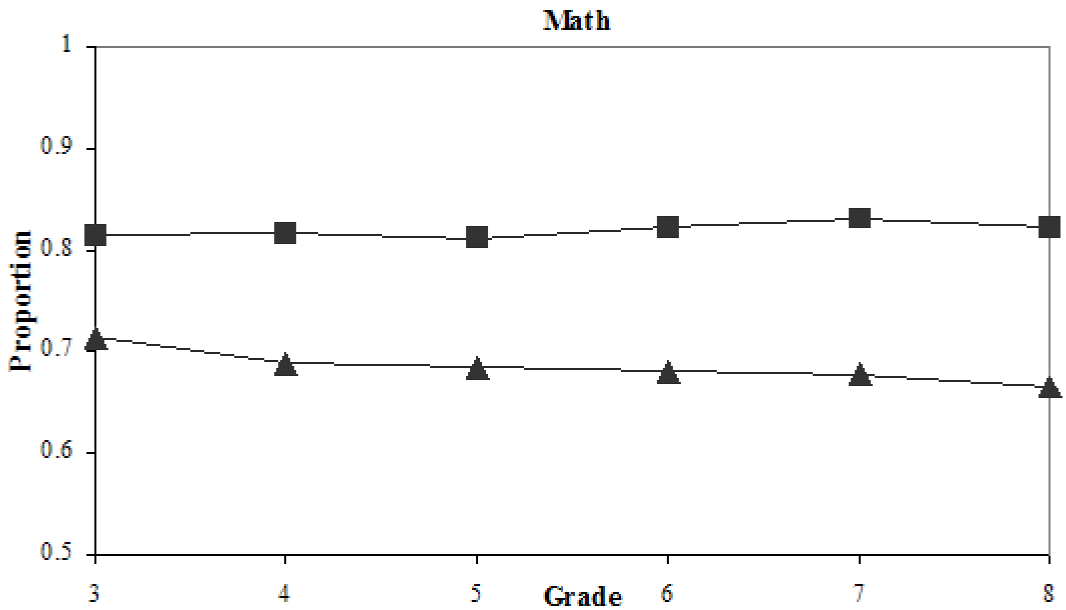
Figure 3b. Hispanic-White Raw Achievement Gaps, 1995-1999 Combined Cohorts, Math and Reading, Alternative Measures

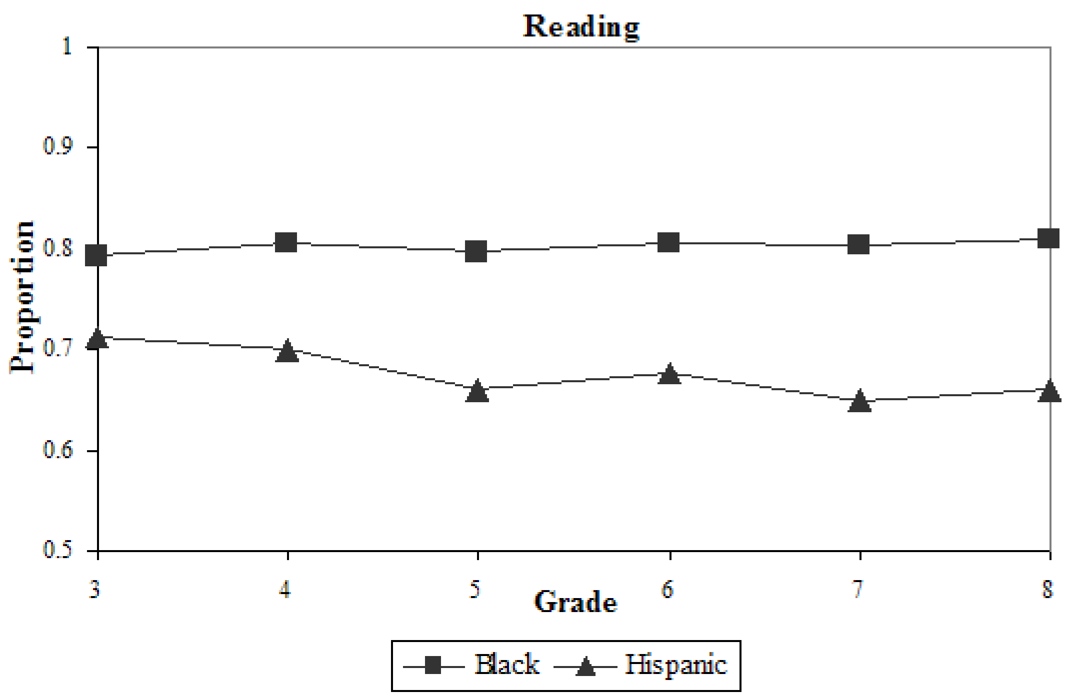


Source: Appendix Table 3b.

8/6/07

Figure 4. Proportion of Group Scoring Below White Median, Black and Hispanic, 1995-1999 Combined Cohorts





Source: Appendix Tables 3a and 3b.

7/19/07

Figure 5a. Black and White Achievement Distributions, Math, 1999 Cohort, 3rd Grade

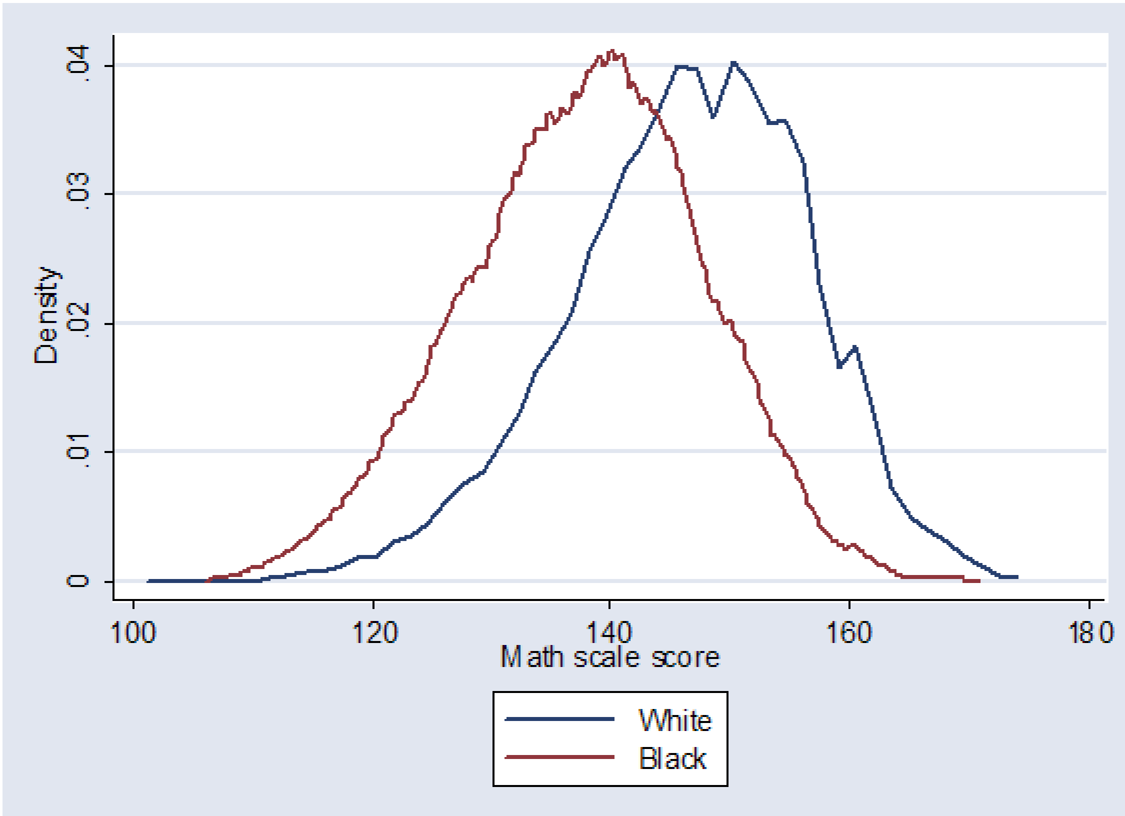


Figure 5b. Black and White Achievement Distributions, Math, 1999 Cohort, 8th Grade

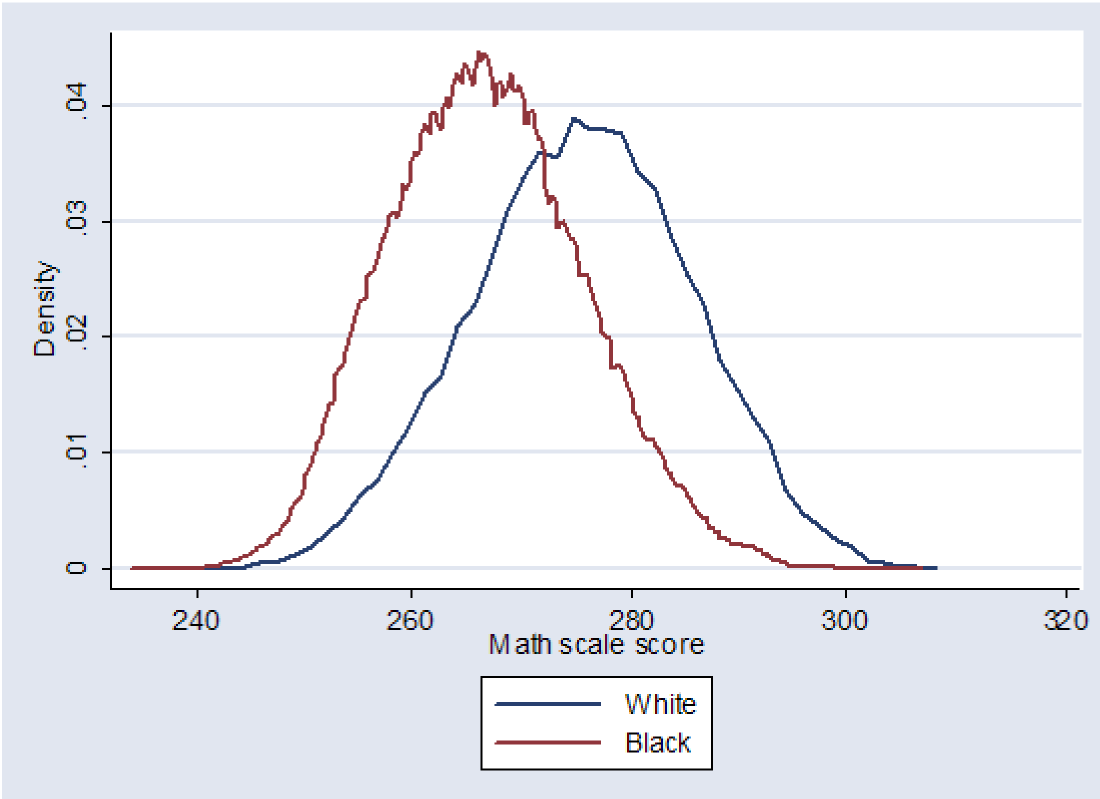


Table 1. Calculated Black-White Achievement Gaps, Selected Studies and Grades, in Standard Deviation Units

Grade	Study	Raw differences in means		Adjusted differences in means	
		Math	Reading	Math	Reading
K (fall)	Fryer-Levitt[1]	-0.66	-0.40	-0.10	0.13
K	Fryer-Levitt[1]	-0.73	-0.45	-0.21	0.00
K	Murnane et al.[2]	-1.00	-1.18	-0.74	-0.92
1	Fryer-Levitt[1]	-0.76	-0.52	-0.28	-0.08
1	Bali-Alvarez[3]	-0.55	-0.35	-0.28	-0.21
3	Fryer-Levitt[1]	-0.88	-0.77	-0.38	-0.28
3	Hanushek-Rivkin[4]	-0.70	----[a]		
4	Phillips-Chin[5]	-0.90	-0.83		
4	Bali-Alvarez[3]	-0.50	-0.45		
5	Hanushek-Rivkin[4]	-0.73	----[a]		
5	Murnane et al.[2]	-1.01	-1.09	-0.81	-0.87
5	Stiefel et al.[6]	-0.81	-0.73	----[a]	-0.48
8	Phillips-Chin[5]	-1.06	-0.85		
8	Hanushek-Rivkin[4]	-0.76	----[a]		
8	Stiefel et al.[6]	-0.84	-0.78	----[a]	-0.55

Note: Raw gaps are differences in mean achievement scores, and adjusted gaps are estimated coefficient of black indicator in regressions with student covariates.
Blanks indicate that studies did not calculate comparable adjusted gaps.
a. Gaps calculated just for math scores. Comparable gaps not presented for this subject.
Sources:
[1] Fryer and Levitt (2005, Tables 1, 2 and 3).
[2] Murnane et al. (2006, Tables 2, 8).
[3] Bali and Alvarez (2004, p. 409). Estimated by dividing reported gaps, measured in units of test scores, by the standard deviation of 4th grade reading test.
[4] Hanushek and Rivkin (2006, Table 3).
[5] Phillips and Chin (2004) and spreadsheet made available by the authors.
[6] Stiefel et al. (2007, Table 2; Table 3, without fixed effects).

Table 2. Raw Achievement Gaps, Grades 3 to 8, 1995-1999 Combined Cohorts

	Average normalized math score, by grade						Average normalized reading score, by grade					
	3	4	5	6	7	8	3	4	5	6	7	8
Black	-0.783	-0.817	-0.794	-0.830	-0.831	-0.814	-0.710	-0.755	-0.771	-0.755	-0.765	-0.776
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Hispanic	-0.516	-0.473	-0.449	-0.456	-0.427	-0.384	-0.562	-0.515	-0.460	-0.440	-0.391	-0.397
	(0.011)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
Asian	0.028	0.133	0.207	0.247	0.292	0.342	-0.108	-0.060	0.005	0.033	0.091	0.095
	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
Am. Ind	-0.589	-0.621	-0.674	-0.642	-0.632	-0.569	-0.634	-0.653	-0.714	-0.666	-0.652	-0.629
	(0.012)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.012)	(0.013)	(0.013)
Multiracial	-0.279	-0.282	-0.263	-0.301	-0.289	-0.275	-0.227	-0.221	-0.201	-0.205	-0.207	-0.203
	(0.018)	(0.018)	(0.019)	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)	(0.019)	(0.018)	(0.018)	(0.018)

Note: For math, the cohorts contain 56,591 students in 1995, 61,593 in 1996, 71,753 in 1997, 76,398 in 1998 and 79,147 in 1999. The cohort consists of NC public school students who took the grade 3 math end-of-grade test in the spring of 1998 and who took end-of-year math tests in each of the following five years. For reading, the cohort is defined analogously and contains 56,473 students in 1995, 61,332 in 1996, 71,334 in 1997, 75,853 for 1998 and 78,431 in 1999.

Source: North Carolina Education Data Center; author's calculations

7/18/07

Table 3. Decomposing the Raw Achievement Gap, Grades 3 to 8, 1995-99 Combined Cohorts

Group	Portion of gap	Math score, by grade						Reading score, by grade					
		3	4	5	6	7	8	3	4	5	6	7	8
Black	Total	-0.783	-0.817	-0.794	-0.830	-0.831	-0.814	-0.710	-0.755	-0.771	-0.755	-0.765	-0.776
	Unexplained	-0.530	-0.528	-0.503	-0.478	-0.486	-0.454	-0.446	-0.470	-0.478	-0.440	-0.460	-0.465
	Covariates	-0.231	-0.246	-0.255	-0.260	-0.270	-0.267	-0.242	-0.251	-0.254	-0.256	-0.255	-0.256
	School F.E.	-0.021	-0.043	-0.036	-0.092	-0.075	-0.093	-0.023	-0.034	-0.039	-0.059	-0.050	-0.055
Hispanic	Total	-0.516	-0.473	-0.449	-0.456	-0.427	-0.384	-0.562	-0.515	-0.460	-0.440	-0.391	-0.397
	Unexplained	-0.102	-0.025	0.009	0.020	0.055	0.101	-0.127	-0.064	0.011	0.023	0.084	0.084
	Covariates	-0.404	-0.431	-0.441	-0.427	-0.451	-0.454	-0.425	-0.437	-0.457	-0.432	-0.458	-0.461
	School F.E.	-0.010	-0.017	-0.018	-0.049	-0.031	-0.031	-0.010	-0.014	-0.015	-0.031	-0.017	-0.020
Asian	Total	0.028	0.133	0.207	0.247	0.292	0.342	-0.108	-0.060	0.005	0.033	0.091	0.095
	Unexplained	0.071	0.186	0.239	0.273	0.318	0.384	-0.079	-0.022	0.040	0.050	0.110	0.120
	Covariates	-0.023	-0.020	-0.016	-0.008	-0.010	-0.010	-0.024	-0.023	-0.025	-0.019	-0.026	-0.025
	School F.E.	-0.020	-0.033	-0.017	-0.019	-0.016	-0.032	-0.005	-0.014	-0.010	0.003	0.006	-0.000
Am. Indian	Total	-0.589	-0.621	-0.674	-0.642	-0.632	-0.569	-0.634	-0.653	-0.714	-0.666	-0.652	-0.629
	Unexplained	-0.299	-0.296	-0.285	-0.262	-0.262	-0.237	-0.291	-0.285	-0.281	-0.261	-0.251	-0.238
	Covariates	-0.246	-0.263	-0.272	-0.275	-0.285	-0.284	-0.260	-0.269	-0.273	-0.272	-0.275	-0.276
	School F.E.	-0.043	-0.062	-0.117	-0.105	-0.085	-0.048	-0.083	-0.099	-0.160	-0.133	-0.126	-0.115
Multiracial	Total	-0.279	-0.282	-0.263	-0.301	-0.289	-0.275	-0.227	-0.221	-0.201	-0.205	-0.207	-0.203
	Unexplained	-0.145	-0.139	-0.119	-0.133	-0.116	-0.091	-0.095	-0.090	-0.066	-0.060	-0.061	-0.054
	Covariates	-0.122	-0.129	-0.134	-0.133	-0.140	-0.138	-0.125	-0.128	-0.130	-0.128	-0.130	-0.131
	School F.E.	-0.012	-0.014	-0.010	-0.035	-0.033	-0.047	-0.007	-0.003	-0.005	-0.017	-0.016	-0.019

Note: Total gaps are taken from Table 2; the unexplained portions are the coefficients of race/ethnicity indicators in regressions explaining normalized end-of-grade test scores and including student covariates for gender, age, subsidized lunch, parental education, year of cohort, type of district, region, and school fixed effects. The School F.E. portion is the difference from whites in the average coefficient of the school effect, and the covariates portion is the difference from white based on covariates.

7/27/07

Table 4a. Achievement Gaps in the Tails of the Math Distribution

		Proportion of students below the 10th percentile of whites			Proportion of white students above the 90th percentile of each racial group		
Group	Cohort	3rd grade	8th grade	Change	3rd grade	8th grade	Change
Black	1995	0.31	0.29	-0.02	0.35	0.39	0.04
	1996	0.30	0.27	-0.03	0.34	0.37	0.02
	1997	0.30	0.26	-0.04	0.34	0.35	0.01
	1998	0.31	0.29	-0.02	0.35	0.38	0.03
	1999	0.29	0.27	-0.02	0.37	0.35	-0.03
Hispanic	1995	0.18	0.14	-0.05	0.21	0.20	0.00
	1996	0.18	0.14	-0.04	0.22	0.20	-0.03
	1997	0.24	0.17	-0.07	0.21	0.18	-0.03
	1998	0.23	0.16	-0.07	0.24	0.18	-0.06
	1999	0.20	0.17	-0.03	0.22	0.19	-0.03
Asian	1995	0.07	0.05	-0.01	0.06	0.03	-0.03
	1996	0.09	0.04	-0.05	0.06	0.03	-0.03
	1997	0.09	0.06	-0.04	0.09	0.05	-0.04
	1998	0.11	0.05	-0.06	0.07	0.04	-0.03
	1999	0.09	0.05	-0.03	0.07	0.03	-0.03
Am. Ind.	1995	0.29	0.21	-0.08	0.27	0.26	-0.01
	1996	0.23	0.16	-0.08	0.22	0.25	0.03
	1997	0.29	0.19	-0.10	0.25	0.29	0.03
	1998	0.24	0.19	-0.04	0.27	0.27	0.00
	1999	0.20	0.18	-0.02	0.26	0.28	0.02

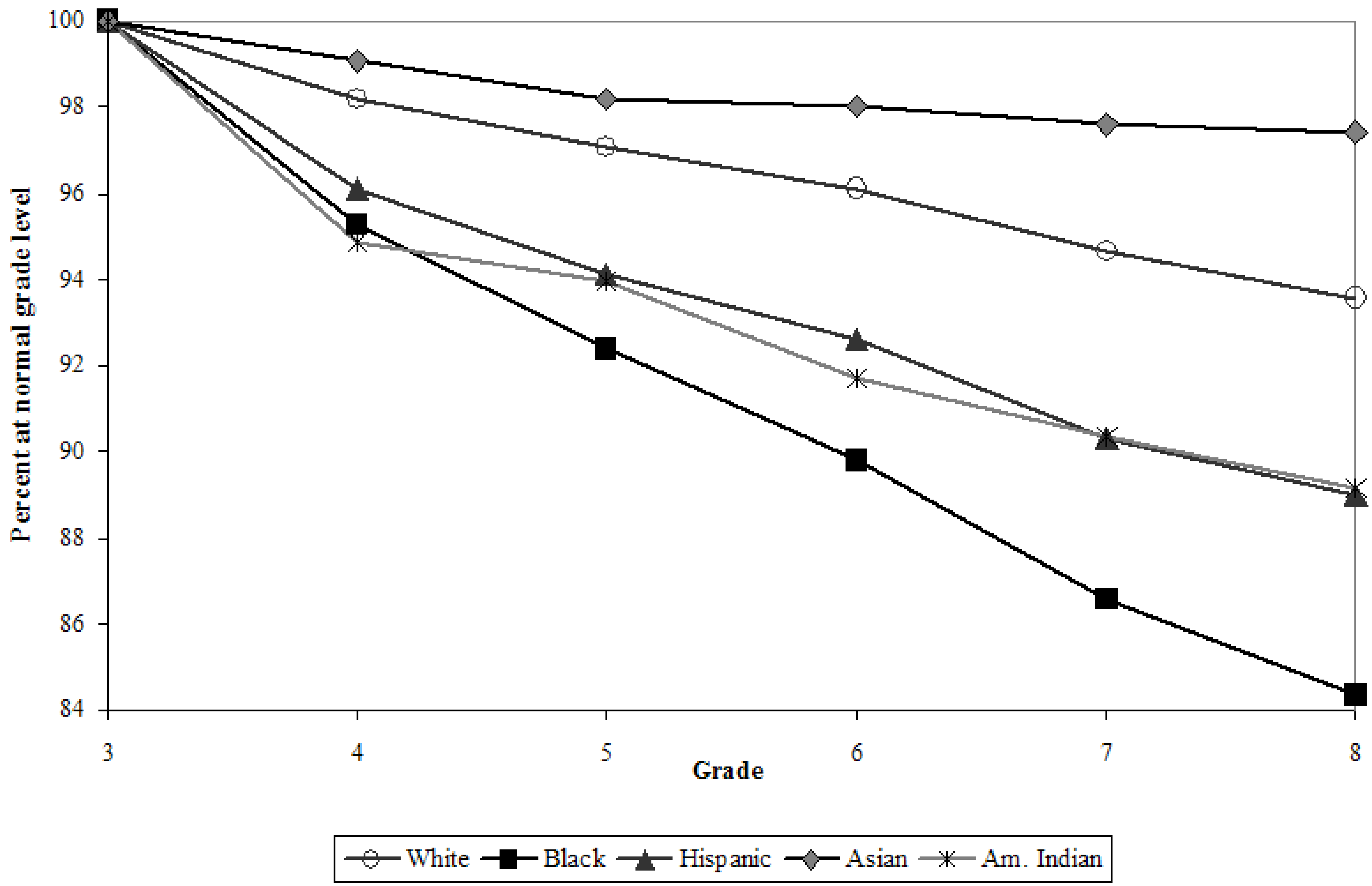
7/19/07

Table 4b. Achievement Gaps in the Tails of the Reading Distribution

		Proportion of students below the 10th percentile of whites			Proportion of white students above the 90th percentile of each racial group		
Group	Cohort	3rd grade	8th grade	Change	3rd grade	8th grade	Change
Black	1995	0.25	0.27	0.02	0.35	0.32	-0.03
	1996	0.25	0.25	0.00	0.32	0.33	0.01
	1997	0.24	0.28	0.04	0.35	0.34	0.00
	1998	0.27	0.27	0.00	0.36	0.33	-0.03
	1999	0.25	0.26	0.01	0.35	0.33	-0.02
Hispanic	1995	0.15	0.14	-0.02	0.19	0.14	-0.04
	1996	0.17	0.12	-0.05	0.24	0.16	-0.08
	1997	0.22	0.19	-0.03	0.26	0.25	-0.01
	1998	0.24	0.17	-0.07	0.27	0.18	-0.09
	1999	0.23	0.18	-0.05	0.26	0.23	-0.03
Asian	1995	0.07	0.04	-0.03	0.09	0.05	-0.04
	1996	0.09	0.04	-0.04	0.09	0.06	-0.03
	1997	0.11	0.06	-0.04	0.10	0.09	-0.01
	1998	0.12	0.05	-0.07	0.13	0.07	-0.07
	1999	0.12	0.06	-0.05	0.07	0.05	-0.02
Am. Ind.	1995	0.28	0.23	-0.04	0.27	0.28	0.01
	1996	0.24	0.19	-0.05	0.32	0.29	-0.04
	1997	0.26	0.23	-0.03	0.30	0.30	-0.01
	1998	0.26	0.20	-0.06	0.27	0.23	-0.04
	1999	0.19	0.18	-0.01	0.31	0.28	-0.03

7/19/07

Appendix Figure 1. Grade Progression by Racial/Ethnic Group, 1999 Cohort



Appendix Table 1. Normal Grade, by Year and Cohort

School year	Test date	Cohort, by Year in 3rd Grade				
		1995	1996	1997	1998	1999
1994/1995	1995	3				
1995/1996	1996	4	3			
1996/1997	1997	5	4	3		
1997/1998	1998	6	5	4	3	
1998/1999	1999	7	6	5	4	3
1999/2000	2000	8	7	6	5	4
2000/2001	2001		8	7	6	5
2001/2002	2002			8	7	6
2002/2003	2003				8	7
2003/2004	2004					8
Sample size						
Math		56,591	61,593	71,753	76,398	79,147
Reading		56,473	61,332	71,334	75,853	78,431

7/18/07

Appendix Table 2. Sample Sizes and Mean Values for Selected Variables, Groups of 3rd Grade Students, 1999; One Group of 8th Grade Students, 2004

	1999 cohort		Not present all six years	
	A Normal progress	B Repeated a grade	C Present in 1999	D Present in 2004
N	71,800	7,606	19,451	29,480
Normalized grade 3 score (a)				
Math	0.169	-1.023	-0.053	--
Reading	0.171	-1.008	-0.035	--
Male	0.488	0.637	0.537	0.549
Race/ Ethnicity				
White	0.646	0.423	0.633	0.521
Black	0.287	0.515	0.263	0.315
Hispanic	0.025	0.030	0.043	0.104
Asian	0.016	0.004	0.022	0.026
Am. Indian	0.015	0.017	0.011	0.009
Multiracial	0.012	0.011	0.019	0.020
Age May of grade 3 (b)	9.216	9.260	9.289	9.574
Parental education				
College	0.296	0.040	0.094	0.121
HS, no college degree	0.647	0.763	0.871	0.812
No HS degree	0.057	0.196	0.035	0.068
Subsidized lunch (c)	0.420	0.656	--	0.538
Ever exceptional	0.149	0.335	0.189	0.183
Ever limited English	0.018	0.023	0.026	0.059
District group				
Largest five	0.276	0.232	0.334	0.335
Urban Coastal	0.095	0.130	0.125	0.116
Urban Piedmont	0.101	0.100	0.092	0.108
Urban Mountain	0.077	0.071	0.075	0.070
Rural Coastal	0.061	0.076	0.058	0.059
Rural Piedmont	0.240	0.293	0.200	0.205
Rural Mountain	0.149	0.098	0.116	0.107

Note: The 1999 cohort includes all students who took the 3rd grade end-of-grade math or reading test in the spring of 1999, and the end-of-grade math and reading tests in each year through 2004. Group A includes all those who progressed one grade each year. Group B is comprised of students who were present in all six years but who repeated at least one grade.

The samples shown in columns C and D are comprised of students who did not have an end-of-grade math or reading score or were otherwise not in the data set in each of the six years.
a. Each end-of-grade test was normalized using the statewide mean and standard deviation of the scaled score, producing a normalized score with zero mean and unitary standard deviation. For students who were retained in grade, scaled scores were applied to the statewide mean and standard deviation applying to the grade in which they would have been had they made normal progress. See text. Very few in group D were in a state's public schools in their third grade year, so no mean is reported.
b. Age in years is exact only to the month since birth date was available only to the closest month.
c. Percentage of students who were eligible for free or reduced price lunch in either of the last two years covered by the sample, normally grades 7 or 8. Since very few in group C were in the state's public schools in the last two years of the grade span, the mean for this group is omitted.

Source: North Carolina Education Research Data Center; authors' calculations.

7/19/07

Appendix Table 3a. Alternative Gap Measures: Black-White, Grades 3 to 8, 1995-1999 Combined Cohorts

	Math, by Grade						Reading, by Grade					
	3	4	5	6	7	8	3	4	5	6	7	8
Fixed cohorts												
Raw gap, mean	-0.783	-0.817	-0.794	-0.830	-0.831	-0.814	-0.710	-0.755	-0.771	-0.755	-0.765	-0.776
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Raw gap, median	-0.796	-0.872	-0.831	-0.908	-0.849	-0.912	-0.793	-0.836	-0.807	-0.832	-0.844	-0.786
	(0.001)	(0.003)	(0.001)	(0.000)	(0.006)	(0.002)	(0.001)	(0.001)	(0.002)	(0.003)	(0.007)	(0.005)
Proportion scoring below white median	0.814	0.816	0.812	0.823	0.831	0.823	0.793	0.805	0.798	0.806	0.803	0.809
Raw gap, demeaned scale score / 10	-0.883	-0.858	-0.796	-0.871	-0.907	-0.903	-0.696	-0.716	-0.665	-0.713	-0.653	-0.653
	(0.004)	(0.004)	(0.003)	(0.004)	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Gap adjusted for covariates	-0.552	-0.571	-0.539	-0.570	-0.561	-0.547	-0.468	-0.504	-0.517	-0.499	-0.510	-0.520
Gap adjusted for covariates and school F.E.	-0.530	-0.528	-0.503	-0.478	-0.486	-0.454	-0.446	-0.470	-0.478	-0.440	-0.460	-0.465
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Other samples												
Raw gap, cohort without repeaters	-0.725	-0.756	-0.738	-0.782	-0.786	-0.769	-0.659	-0.704	-0.701	-0.711	-0.698	-0.715
	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)	(0.004)	(0.003)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)
Raw gap, repeated cross-sections	-0.805	-0.818	-0.788	-0.824	-0.830	-0.821	-0.729	-0.766	-0.763	-0.774	-0.766	-0.794
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)

Sample sizes: Fixed cohorts 345,482 for math, 343,423 for reading; fixed cohorts without repeaters 314,921 for math, 313,310 for reading; repeated cross-sections vary in size from 471,726 for reading in 4th grade to 488,764 for math in 8th grade.

Source: North Carolina Education Data Center; author's calculations
7/27/07

Appendix Table 3b. Alternative Gap Measures: Hispanic-White, Grades 3 to 8, 1995-1999 Combined Cohorts

	Math, by Grade						Reading, by Grade					
	3	4	5	6	7	8	3	4	5	6	7	8
Fixed cohorts												
Raw gap, mean	-0.516 (0.011)	-0.473 (0.012)	-0.449 (0.012)	-0.456 (0.012)	-0.427 (0.012)	-0.384 (0.012)	-0.562 (0.012)	-0.515 (0.012)	-0.460 (0.012)	-0.440 (0.012)	-0.391 (0.012)	-0.397 (0.012)
Raw gap, median	-0.539 (0.005)	-0.500 (0.012)	-0.498 (0.002)	-0.505 (0.000)	-0.455 (0.019)	-0.450 (0.005)	-0.616 (0.005)	-0.533 (0.003)	-0.471 (0.008)	-0.503 (0.012)	-0.469 (0.026)	-0.442 (0.018)
Proportion scoring below white median	0.712	0.689	0.684	0.680	0.679	0.665	0.714	0.701	0.660	0.677	0.650	0.659
Raw gap, demeaned scale score / 10	-0.581 (0.013)	-0.494 (0.012)	-0.448 (0.012)	-0.475 (0.012)	-0.465 (0.013)	-0.425 (0.013)	-0.549 (0.012)	-0.489 (0.011)	-0393 (0.010)	-0.416 (0.011)	-0.336 (0.010)	-0.337 (0.010)
Gap adjusted for covariates	-0.112	-0.042	-0.008	-0.029	0.024	0.070	-0.137	-0.078	-0.003	-0.008	0.067	0.064
Gap adjusted for covariates and school F.E.	-0.102 (0.011)	-0.025 (0.011)	0.009 (0.011)	0.020 (0.011)	0.055 (0.011)	0.101 (0.011)	-0.127 (0.011)	-0.064 (0.011)	0.011 (0.011)	0.023 (0.011)	0.084 (0.011)	0.084 (0.011)
Other samples												
Raw gap, cohort without repeaters	-0.495 (0.012)	-0.450 (0.012)	-0.433 (0.012)	-0.439 (0.012)	-0.414 (0.012)	-0.366 (0.012)	-0.542 (0.012)	-0.497 (0.012)	-0.431 (0.012)	-0.433 (0.012)	-0.374 (0.012)	-0.379 (0.012)
Raw gap, repeated cross-sections	-0.532 (0.009)	-0.541 (0.009)	-0.587 (0.008)	-0.627 (0.008)	-0.650 (0.007)	-0.634 (0.007)	-0.583 (0.010)	-0.610 (0.009)	-0.629 (0.008)	-0.677 (0.008)	-0.682 (0.008)	-0.761 (0.007)

Sample sizes: See Appendix Table 3a.

Appendix Table 4a. Regression Estimates, Math Scores, 1995-99 Combined Cohorts

Variable	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
Race/ Ethnicity (white omitted)						
Black	-0.530*	-0.528*	-0.503*	-0.478*	-0.486*	-0.454*
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Hispanic	-0.102*	-0.025*	0.009	0.020	0.055*	0.101*
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
Asian	0.071*	0.186*	0.239*	0.273*	0.318*	0.384*
	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)
Am. Indian	-0.299*	-0.296*	-0.285*	-0.262*	-0.262*	-0.237*
	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)
Multiracial	-0.145*	-0.139*	-0.119*	-0.133*	-0.116*	-0.091*
	(0.016)	(0.016)	(0.017)	(0.016)	(0.016)	(0.016)
Male	-0.012*	-0.010*	-0.015*	-0.015*	-0.031*	-0.026*
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Age in years,	-0.068*	-0.123*	-0.156*	-0.202*	-0.187*	-0.218*
May of grade 3	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Subsidized lunch	-0.181*	-0.191*	-0.194*	-0.206*	-0.205*	-0.203*
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Parental education						
No HS degree	-1.048*	-1.107*	-1.130*	-1.068*	-1.150*	-1.149*
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
HS, no college	-0.559*	-0.601*	-0.632*	-0.634*	-0.676*	-0.665*
degree	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Year of Cohort (1995 omitted)						
1996	-0.006	-0.003	-0.004	0.000	-0.007	-0.012*
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
1997	0.001	-0.006	-0.002	0.005	-0.002	-0.017*
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
1998	0.008	-0.008	-0.007	0.005	0.001	0.011*
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
1999	0.013*	-0.001	-0.002	0.005	0.047*	0.028*
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
N	345,482	345,482	345,482	345,482	345,482	345,482
R^2	0.257	0.276	0.280	0.297	0.307	0.300

Note: Equations estimated with school fixed effects, standard errors in parentheses; * denotes significance at the 5% level.

7/18/07

Appendix Table 4b. Regression Estimates, Reading Scores, 1995-99 Combined Cohorts

Variable	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
Race/ Ethnicity (white omitted)						
Black	-0.446*	-0.470*	-0.478*	-0.440*	-0.460*	-0.465*
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Hispanic	-0.127*	-0.064*	0.011	0.023*	0.084*	0.084*
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
Asian	-0.079*	-0.022	0.040*	0.050*	0.110*	0.120*
	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)
Am. Indian	-0.291*	-0.285*	-0.281*	-0.261*	-0.251*	-0.238*
	(0.015)	(0.015)	(0.016)	(0.015)	(0.015)	(0.015)
Multiracial	-0.095*	-0.090*	-0.066*	-0.060*	-0.061*	-0.054*
	(0.017)	(0.017)	(0.017)	(0.016)	(0.017)	(0.016)
Male	-0.166*	-0.151*	-0.139*	-0.146*	-0.163*	-0.165*
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Age in years,	-0.113*	-0.137*	-0.148*	-0.177*	-0.160*	-0.183*
May of grade 3	(0.003)	(0.003)	(0.004)	(0.003)	(0.003)	(0.003)
Subsidized lunch	-0.203*	-0.210*	-0.208*	-0.222*	-0.215*	-0.218*
	(0.004)	(0.004)	(0.004)	(0.003)	(0.004)	(0.004)
Parental education						
No HS degree	-1.077*	-1.101*	-1.156*	-1.061*	-1.147*	-1.141*
	(0.007)	(0.007)	(0.007)	(0.006)	(0.007)	(0.007)
HS, no college	-0.572*	-0.593*	-0.601*	-0.588*	-0.593*	-0.590*
degree	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Year of Cohort (1995 omitted)						
1996	-0.001	0.000	-0.009	-0.005	-0.020*	-0.012*
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
1997	0.010*	0.007	-0.005	0.008	-0.011*	-0.021*
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
1998	0.016*	0.004	-0.014*	0.014*	-0.011*	0.029*
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
1999	0.025*	0.016*	-0.015*	0.010*	0.043*	0.046*
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
N	343,423	343,423	343,423	343,423	343,423	343,423
R^2	0.256	0.277	0.280	0.291	0.288	0.297

Note: Equations estimated with school fixed effects, standard errors in parentheses; * denotes significance at the 5% level.

7/18/07

Appendix Table 5. Sample Sizes and Mean Values for Selected Variables, by Race/Ethnicity, 1999 Cohort

	White	Black	Hispanic	Asian	Am.Indian	Multiracial
N	49,617	24,498	2,025	1,163	1,191	912
Normalized grade 3 score						
Math	0.320	-0.463	-0.206	0.368	-0.179	0.072
Reading	0.307	-0.407	-0.297	0.179	-0.262	0.116
Male	0.510	0.487	0.521	0.506	0.490	0.493
Age May of grade 3	9.21	9.24	9.32	9.17	9.24	9.17
Parental Education						
College	0.355	0.118	0.089	0.408	0.126	0.240
HS, no college degree	0.591	0.809	0.488	0.440	0.777	0.708
No HS degree	0.053	0.073	0.422	0.153	0.097	0.052
Subsidized lunch	0.265	0.757	0.764	0.459	0.761	0.536
Ever exceptional	0.168	0.173	0.146	0.093	0.154	0.148
Ever limited English	0.002	0.002	0.489	0.301	0.005	0.013
District Group						
Largest Five	0.233	0.343	0.300	0.513	0.103	0.320
Urban coastal	0.085	0.133	0.080	0.037	0.016	0.126
Urban piedmont	0.090	0.123	0.130	0.108	0.021	0.130
Urban mountain	0.094	0.044	0.049	0.116	0.015	0.090
Rural coastal	0.058	0.079	0.056	0.006	0.006	0.041
Rural piedmont	0.238	0.241	0.271	0.081	0.791	0.206
Rural mountain	0.201	0.037	0.114	0.139	0.048	0.087

Sample: See Appendix Table 2 note.

7/19/07

Appendix Table 6. Linear Trends over Time in Achievement Gaps, Grades 3 and 8, 1995 to 1999 Cohorts

	3rd Grade		8th Grade	
	Math	Reading	Math	Reading
Black	0.008*	0.000	0.009*	0.014*
	(0.002)	(0.002)	(0.002)	(0.002)
Hispanic	0.006	-0.021*	0.026*	-0.009
	(0.008)	(0.008)	(0.008)	(0.008)
Asian	0.011	-0.009	0.020*	-0.010
	(0.009)	(0.009)	(0.009)	(0.009)
Amer. Indian	0.040*	0.027*	0.002	0.024*
	(0.008)	(0.008)	(0.008)	(0.008)

Note: Table shows the coefficients of interaction terms between race/ethnicity indicators and a linear time trend. Other variables include race/ethnicity indicators, linear time trend, age, and indicators for gender, subsidized lunch, parental education, district type, region, and school fixed effects. Standard errors in parentheses; * denotes significance at the 5% level.

Source: North Carolina Education Research Center; authors' calculations.

7/30/07