Harming the Best: How Schools Affect the Black–White Achievement Gap

Eric A. Hanushek
Steven G. Rivkin

Abstract

Sizeable achievement differences by race appear in early grades, but substantial uncertainty exists about the impact of school quality on the black–white achievement gap and particularly about its evolution across different parts of the achievement distribution. Texas administrative data show that the overall growth in the achievement gap between third and eighth grades is larger for students with higher initial achievement and that specific teacher and peer characteristics explain a substantial share of the widening. The adverse effect of attending school with a high black enrollment share appears to be an important contributor to the larger growth in the achievement differential in the upper part of the test score distribution. This evidence reaffirms the major role played by peers and school quality, but also presents a policy dilemma. Teacher labor market complications, current housing patterns, legal limits to desegregation efforts, and uncertainty about the overall effects of specific desegregation programs indicate that effective policy responses will almost certainly involve a set of school improvements beyond simple changes in peer racial composition and the teacher experience distribution. © 2009 by the Association for Public Policy Analysis and Management.

Perhaps no other social policy issue has been as important or as stubborn to deal with as racial gaps in economic outcomes. Black–white differences in academic attainment, occupation, and earnings, while showing some improvement over the past quarter century, remain large. Much of the policy effort aimed at reducing these gaps focuses on public elementary and secondary schools. This emphasis hinges on the widespread beliefs that school and peer characteristics disadvantage blacks relative to whites and that appropriate school interventions can raise achievement and improve future life outcomes. This paper investigates the first of these beliefs through an examination of the changes in the black-white achievement gap as students progress through school.[1]

Our findings suggest that the increase in the achievement gap across grades is larger for blacks with higher initial achievement and that this is due primarily to stronger deleterious effects for initially high-achieving blacks of attending schools

[1] The Hispanic–white achievement gap is also large but has followed a somewhat different pattern over time compared with the black–white gap. The heterogeneity of the Hispanic population in Texas, including differences in English proficiency and immigrant status, requires consideration of a number of factors in addition to race and ethnicity, and we intend to investigate Hispanic student academic achievement in future work.

Journal of Policy Analysis and Management, Vol. 28, No. 3, 366–393 (2009)

Published by Wiley Periodicals, Inc. Published online in Wiley InterScience
(www.interscience.wiley.com)
DOI: 10.1002/pam.20437

with a high black enrollment share. There is little evidence, however, that Hispanic enrollment share significantly affects the achievement of either non-Hispanic blacks or whites, indicating that it is the black rather than the minority enrollment share that is the key peer demographic.

Differences by initial achievement in both the growth in the achievement gap and relationship to school racial composition are striking and carry important implications for the future education and earnings distributions. The expanding racial achievement gap is fueled by relatively constant gaps at the bottom of the black and the white distributions and a dramatically increasing gap at the top. Given the relationship between cognitive skills and economic outcomes, the truncation at the top of the black achievement distribution does not bode well for the future expansion of the number of blacks who complete college and graduate school and who enter high-prestige occupations and positions of power.

We begin with a description of the evolution of achievement differences by initial achievement level. We follow this with an investigation of the contributions of school and peer characteristics to these changes, paying particular attention to the possibility that the importance of specific factors may differ by initial achievement. We conclude with consideration of the policy implications of our findings.

ECONOMIC MOTIVATION

Evidence consistently finds that cognitive skills explain a substantial share of black–white gaps in school attainment and in wages, and this has motivated aggressive policies to raise the quality of education for blacks.[2] The landmark decision in *Brown v. Board of Education* that attacked racial segregation of schools was the modern beginning of concerted federal, state, and local actions directed at improving black achievement.[3] Along with subsequent court cases, *Brown* ushered in a profound change in both school and peer characteristics, while contemporaneous increases in school spending, brought on in part by school finance litigation, further raised the resources devoted to black students in the public schools.

Nonetheless, substantial racial disparities have persisted, and Table 1 provides a stark picture of the black–white differences in academic, economic, and social outcomes that have survived the schooling and social policy interventions of the post-*Brown* period.[4] Among men and women 20 to 24 years old, blacks are far less likely to complete or be in the process of completing college, far less likely to work, and far more likely to be in prison or other institution. The rates of incarceration and nonemployment for young black men paint a particularly dire picture. Overall, these persistent gaps raise questions about the efficacy of schooling interventions, including desegregation and school finance redistribution.

REVIEW OF ACHIEVEMENT GAP RESEARCH

A number of different strands of literature, often linked directly to governmental policy initiatives, investigate the sources of racial achievement gaps. However, when combined, they tend to present a confusing and uncertain picture—particularly about the impact of schools.

[2] O'Neill (1990) and Neal and Johnson (1996) provide evidence on wage differences, and Rivkin (1995) provides evidence on differences in educational attainment. A more general review of the economic impacts of cognitive skills on individual earnings is found in Hanushek and Woessmann (2008).
[3] *Brown v. Board of Education*, 347 U.S. 483 (1954).
[4] Neal (2006) documents black–white gaps in both quantity and quality of schooling and shows evidence that convergence of earlier periods slowed or stopped in the 1980s and 1990s.

Table 1. Distribution of 20- to 24-year-olds by school status, employment status, years of schooling, and institutionalization status in 2000 (percentages by gender and race).

				High School Graduate						
		High School Dropout		Attending School		Not Attending school		College Graduate		
	Institutionalized	Not Employed	Employed	Not Employed	Employed	Not Employed	Employed	Not Employed	Employed	Total Observations
Men										
Blacks	14.1%	10.3%	6.7%	12.7%	13.2%	15.0%	23.3%	1.2%	3.8%	10,459
Whites	2.7%	4.2%	9.5%	13.6%	22.5%	6.3%	29.0%	2.3%	10.0%	53,820
Women										
Blacks	0.9%	10.3%	5.6%	15.6%	20.0%	17.2%	21.3%	2.2%	7.0%	10,728
Whites	0.3%	6.4%	4.7%	13.9%	26.8%	9.8%	19.5%	3.1%	15.5%	50,664

Note: Row percentages add to 100 percent.

Source: Author calculations from Census 2000 Public Use Microdata Sample (PUMS).

The original Coleman Report (Coleman et al., 1966) was mandated by the Civil Rights Act of 1964 and motivated by concerns about racial justice in the schools. It described a situation where the gap in student performance, measured in terms of grade-level equivalents, expanded across grades. Its analysis, however, did not suggest that the achievement differences existing in 1965 had much to do with the schools but rather were driven by family and peers.[5] Its companion report (U.S. Commission on Civil Rights, 1967) went further, focusing almost exclusively on the role of racial concentration in the schools as a primary cause of the existing achievement gaps.

Efforts to understand the ultimate impacts of desegregation and other interventions have not been very successful. Several different policy thrusts have targeted the racial achievement gap, including expanded desegregation efforts, but the numerous contemporaneous policy innovations in combination with other social, demographic, and economic changes have inhibited efforts to understand the impacts of specific programs and policies.[6] Much of the early research focused on the immediate achievement effects of court-induced desegregation. Although many of these studies used small-scale random assignment experiments (see, for example, Cook, 1984, and the review in Schofield, 1995), their concentration on immediate outcomes combines reactions to the introduction and form of judicial policy with the impacts of varying racial concentrations and school attendance patterns. Moreover, a lack of statistical power due to small sample sizes may lead to a failure to uncover even educationally important impacts. A related strand of research attempted to isolate the effects of racial concentration by examining longer-term effects of attending schools with few whites, though data limitations raise concerns about the possibility of omitted variables bias (see, for example, Crain & Mahard, 1978).

Researchers have also examined test score changes over time in an effort to identify desegregation effects (see, for example, Armor, 1995).[7] The National Assessment of Education Progress (NAEP) and other sources show a lessening of the black–white gap during the 1980s and early 1990s and stable or rising scores thereafter, and these trends have been the subject of various analyses.[8] In the end, the limited and often contradictory statistical evidence raises doubts about the efficacy of desegregation for addressing racial inequalities in achievement and longer-run outcomes.[9]

Compensatory education programs, which are relatively more important for black students, represent a second contemporaneous policy initiative with potential implications for achievement gaps. With the Elementary and Secondary Education Act (ESEA) of 1965, the federal government entered into the funding and operations of

[5] This analysis was subjected to considerable criticism; see, for example, Bowles and Levin (1968) or Hanushek and Kain (1972).

[6] The U.S. Supreme Court ruling in *Brown* was slow to be implemented, as many Southern districts resisted desegregation and as the court repeatedly refined and sharpened its message of 1954. As a result, the most significant changes in school desegregation occurred in the late 1960s and during the 1970s (see Welch & Light, 1987; Clotfelter, 2004).

[7] A more complete review and assessment of earlier studies can be found in Linn and Welner (2007).

[8] While there is some variation at different ages of NAEP testing, both reading and math score gaps rise somewhat from their low point in the late 1980s or early 1990s and then stabilize thereafter. For varying analyses at different places in the trend, see Congressional Budget Office (1986), Grissmer et al. (1994), the collection of research in Jencks and Phillips (1998), and Hanushek (2001).

[9] Earlier optimism about narrowing gaps (Jencks & Phillips, 1998) largely dissipated with new evidence that the black–white achievement gap stayed constant or even grew during the 1990s (National Center for Education Statistics, 2005). In terms of the specific policies that have been pursued, direct evidence on the benefits of school desegregation remains limited. Review of the evidence surrounding desegregation actions provides limited support for positive achievement effects (Schofield, 1995); Guryan (2004) finds that desegregation reduced the probability of dropping out of high school, though data limitations and methodological concerns raise questions about the findings. Accumulated evidence does not provide strong support for the belief that higher expenditure typically leads to substantial improvements in the quality of instruction, particularly with regard to higher pay for teachers with a master's degree or substantial experience (Hanushek, 2003).

schools more forcefully than ever before, initiating compensatory education programs designed to bring up performance of disadvantaged students.[10] Nonetheless, assessments of compensatory education programs under ESEA raise questions about its effectiveness, thus also raising questions about how such programs affect achievement gaps.[11]

Research on "summer fall back" supports the notion that increasing achievement gaps through the schooling years are largely not a product of the schools *per se*. This work, relying on summer and fall testing of achievement, suggests that learning during the school year might on average be the same for blacks and whites, but that the amount of learning during the summer months heavily favors white students (Alexander, Entwisle, & Olson, 2007; Downey, von Hippel, & Broh, 2004; Heyns, 1978), a finding consistent with the belief that family and peers are the primary contributors to the achievement gap.

Other recent research generally provides additional support for that view. For example, Fryer and Levitt (2004, 2005) find that a substantial racial achievement gap exists at entry to school and increases with age, but that the majority of the increase occurs within schools and is not explained by quantifiable school characteristics.[12] Clotfelter, Ladd, and Vigdor (2006) document a large third-grade mean achievement gap in North Carolina that does not, on average, increase with schooling, but find that the black distribution appears to become more compressed over time. Our past work, on the other hand, highlights meaningful achievement impacts of specific peer and teacher inputs whose distributions differ substantially by race, suggesting possible school-based explanations of at least a portion of the black–white achievement differential.[13]

Analyses of the average achievement gap may miss important school and peer effects that vary across the achievement distribution. For example, discussions of peer pressures on blacks— the "acting white" literature—raises the possibility that high-achieving black students may face very different pressures than lower-achieving blacks.[14] In addition, academic preparation relative to the median or mean student in the school likely affects the extent to which the curriculum approaches the ideal level of challenge for a student. These and other considerations suggest the possibility of differential effects of desegregation and other school policies and factors across the initial achievement distribution.

TEXAS SCHOOLS PROJECT DATA

The UTD Texas Schools Project (TSP) data set provides a unique stacked panel of school administrative data that allows us to track the universe of Texas public

[10] Federal involvement in school accountability under the No Child Left Behind Act of 2001 is, in fact, the latest reauthorization of the ESEA statutes.

[11] Before each reauthorization, the federal government hired an outside evaluator to judge the effectiveness of the compensatory funding. Invariably, these studies found little success in terms of achievement (see Vinovskis, 1999). These findings dovetail with direct analyses of how school resources tend not to be related to student outcomes (for example, Hanushek, 2003). Notwithstanding the evidence, Title 1 funding under ESEA has continued to be the primary federal education program targeted at disadvantaged students.

[12] Note that Murnane et al. (2006) cannot replicate either the basic school patterns of the achievement gap or the influence of measured family background on the gaps when they go to a different, but in some ways richer, database. Neal (2006) finds little evidence of a growing gap past entry to school and discounts the role of schools in either creating or ameliorating any gaps.

[13] Hanushek, Kain, and Rivkin (2004) investigate the effects of student mobility; Rivkin, Hanushek, and Kain (2005) investigate the effects of teacher experience; and Hanushek, Kain, and Rivkin (forthcoming) investigate the effects of racial composition.

[14] See, for example, Austen-Smith and Fryer (2005), Cook and Ludwig (1997), Fordham and Ogbu (1986), Fryer (2006), McWhorter (2000), and Ogbu (2003).

elementary and middle school students as they progress through school.[15] For each cohort, there are over 200,000 students in over 3,000 public schools. Unlike many data sets that sample only small numbers from each school, these data enable us to create accurate measures of peer-group characteristics. We use data on four cohorts for grades three (the earliest grade tested) through eight. The most recent cohort attended eighth grade in 2002, while the earliest cohort attended eighth grade in 1999.

The student data contain a limited number of student, family, and program characteristics including race, ethnicity, gender, and eligibility for a free or reduced-price lunch (the measure of economic disadvantage). The panel feature of the data, however, is exploited to account implicitly for a more extensive set of background characteristics through the use of a value-added framework that controls for prior achievement. Moreover, students who switch schools can be followed as long as they remain in a Texas public school.[16]

Beginning in 1993, the Texas Assessment of Academic Skills (TAAS) was administered each spring to eligible students enrolled in grades three through eight. The tests, labeled criteria referenced tests, evaluate student mastery of grade-specific subject matter. This paper presents results for mathematics. Because the number of questions and average percent right varies across time and grades, test results are standardized to a statewide mean of zero and variance equal to 1 for each grade and year. Because these tests cannot be used to measure knowledge growth with age, they provide no information on absolute racial differences. If the variance in knowledge grows with age and time in school, as we believe most likely, any deterioration in the relative standing of blacks on the achievement tests would understate the increase in knowledge inequality.

The student database is linked to detailed information on teachers, including grade and subject taught, class size, years of experience, highest degree, race, gender, and student population served. Although individual student–teacher matches are not possible, students and teachers are uniquely related to a grade on each campus. Students are assigned the average class size and the distribution of teacher characteristics for teachers in regular classrooms for the appropriate grade, school, subject, and year.

We also exclude students with any missing grade-appropriate test observations from the regression analysis, making two aspects of the data important for the subsequent comparisons. First, differential rates of special education placements that do not call for testing directly affect the achievement comparisons because of the higher incidence among blacks. Second, students retained in grade at any point between grades three and eight cannot be directly included in the regression analysis because of the testing problems introduced by grade repetition. The pattern of grade retention and exclusion from tests, shown in Appendix Table A1,[17] also has

[15] A description of the project can be found at http://www.utdallas.edu/research/tsp/Index.htm.

[16] Given the high rate of school switching, particularly among lower income and minority students, the possibility of following movers is an important asset of the data, particularly when studying achievement gaps. In contrast, such mobility presents a serious sampling problem for survey data sets including the Early Childhood Longitudinal Survey (ECLS) that has been used in recent work on the racial achievement gap (Fryer & Levitt, 2004, 2005; Reardon, 2008). The descriptive analysis the ECLS data in Hanushek and Rivkin (2006) shows that blacks who move between grades 1 and 3 in the ECLS sample exhibit larger test score growth than stayers, a pattern contrary to that observed in most other longitudinal data sets. This highlights the difficulty of generating a representative sample of a mobile population. More generally, sample selection problems in survey data almost certainly grow in magnitude with age.

[17] All appendices are available at the end of this article as it appears in JPAM online. Go to the publisher's Web site and use the search engine to locate the article at http://www3.interscience.wiley.com/cgi-bin/jhome/34787.

implications for the description and analysis of racial achievement gaps.[18] Specifically, because of the racial differential, the simple test data understate the full extent of racial achievement differences, particularly differences in the shares of students in the lower tail of the distribution. As described below, however, we can obtain a bound on the impact of grade retention on the pattern of racial achievement gaps across grades.

RACIAL ACHIEVEMENT DIFFERENCES IN TEXAS

We begin by describing growth in the black–white achievement gap with age for the Texas public schools, considering both the mean achievement differential and differences across the achievement distribution. Virtually all analyses tracing racial patterns in school attainment or in achievement rely on repeated cross sections.[19] Such analyses, however, can be misleading because of underlying changes in the cohorts of students, and they cannot provide a direct comparison of the achievement trajectories of black and white students with similar scores in third grade, a central comparison of interest for this analysis.

Achievement Differences by Grade

To trace the evolution of achievement differences for grades 3, 5, and 8, we focus on a sample of students who progress with their class for six consecutive years, though we augment this cohort analysis by imputing scores to those retained in grade. As noted above, because blacks are more likely to be retained in grade than whites, a focus on only complete cohorts understates any growth in the achievement differential. Because the tests are not vertically scaled across grades, scores for students retained in grade are not comparable to scores for students who progress to the next grade.[20] We address this problem by moving retained students back to their original cohorts for the remaining years they are in the data and then imputing test scores for these years that place the retained students in the bottom of the distribution.[21] The assumption that all students retained in grade score in the bottom percentile appears reasonable to us: Fewer than 3 percent of blacks and 2 percent of whites are retained in any of these grades, and limited violations of this assumption should have little or no effect on any of the percentile comparisons.

Table 2 reports the overlap of the test distributions, characterized by the share of whites with test scores below the 10th, 25th, 50th, 75th, and 90th percentiles of the black distribution. The top panel excludes those retained in grade, while the bottom includes these students in the appropriate grade for their third-grade cohort. Several marked patterns in the distribution emerge from the table. First, less than a quarter of white third graders score below the median black, and half of white third graders score above the 75th percentile of the black distribution. By eighth grade this overlap becomes even smaller. Second, consistent with the trends in Clotfelter,

[18] As Hanushek and Rivkin (2006) show, the achievement gap of students who progress with their class understates black–white differences in academic progress at any point in time given the higher rates of grade retention and missing tests due primarily to special education classification (particularly for boys) among blacks. For test taking and placement in special education, see Hanushek, Kain, and Rivkin (2002).

[19] Clotfelter, Ladd, and Vigdor (2006) is an exception that follows several intact cohorts in North Carolina.

[20] Vertically scaled tests are calibrated to give comparable scores across different forms of tests such as reading tests designed for different grade levels. Specifically, the expected score should be the same for a student taking a vertically scaled test at two different grade levels of the test. It is not clear how accurate tests claiming to be vertically scaled actually would be for students retained in grade, but there is no available analysis of this.

[21] We thank an anonymous referee for this suggestion.

Table 2. Overlap of achievement distributions for black and white students at specific percentiles of the black achievement distribution, by grade and inclusion of students retained in grade.

		Proportion of White Students Scoring Below Given Percentile of Black Test Distribution				
	Grade	10th	25th	50th	75th	90th
Excluding students retained in grade						
	3	0.00	0.08	0.23	0.51	0.74
	5	0.02	0.07	0.21	0.47	0.71
	8	0.02	0.07	0.19	0.42	0.69
Including students retained in grade[a]						
	3	0.02	0.07	0.22	0.46	0.69
	5	0.02	0.07	0.20	0.43	0.67
	8	0.03	0.07	0.18	0.39	0.62

[a] Students retained in grade are put at the bottom of the distribution for their initial cohort and race.

Ladd, and Vigdor (2006) for North Carolina, where a narrowing at the bottom comes with a widening at the top of the distribution, our data show that the increase in the gap occurs mainly in the upper half of the distributions. Both the 10th and 25th percentile scores from the black distribution remain in roughly the same place in the white distribution from third to eighth grade regardless of whether students retained in grade are included in the sample. In contrast, the share of whites scoring below the median black score falls by 4 percentage points, the share of whites scoring below the 75th percentile black score falls by roughly twice that much, and the share of whites scoring below the 90th percentile black score falls by slightly more than the decline observed at the median. Finally, including the students retained in grade worsens the overall picture. With retained students, less than 40 percent of white students fall below the 75th percentile of the black distribution by grade eight; the reduction of the black 75th percentile score following the imputation of very low scores to blacks retained in grade more than offsets the effect of imputing very low scores for the smaller number of whites retained in grade.

Division by Initial Achievement

The comparison of repeated cross sections does not describe the experiences for individual students, and a number of different dynamics could produce the same observed patterns. Our approach to learning more about student experiences divides students into quartiles on the basis of third-grade achievement scores and examines changes in the achievement gaps for blacks and whites who begin in the same quartile. Categorizing students by early achievement is complicated, however, because achievement tests measure actual knowledge with error. As a result, two individuals with identical knowledge can obtain very different scores if one is lucky and the other unlucky in guessing or one was exposed to specific vocabulary words or mathematics problems while the other was not. If such "random errors" are uncorrelated over time, two students with the same growth in true knowledge will tend to experience different test score gains. Specifically, the lucky student with the error-inflated third-grade score will tend to exhibit a lower test score gain than the unlucky student with the artificially low third-grade score.

Such regression to the mean seriously complicates black–white comparisons by achievement level because of race differences in the actual initial skill distributions. The intuition of the problem is straightforward. Assume that students can be classified as either high skilled or low skilled as of third grade and that blacks are less likely to be high skilled than whites because of lower income or related family factors. In addition, assume that tests measure skill with error and that the distribution of the errors is identical for blacks and whites. Given these assumptions, the following hold: (1) the probability that a student who is truly low skilled is erroneously observed to be high skilled is the same for blacks and whites; and (2) the probability that a student observed to be high skilled is actually low skilled is higher for blacks than for whites, because blacks are less likely to be truly high skilled than whites.

Therefore, a black observed to be high skilled will have a more positive error on average than a white observed to be high skilled (because the observed score is true plus error). If the error distribution is identical for blacks and whites in the subsequent grade and independent of the error in the early grade, blacks initially observed to have high skill will be more likely to be observed subsequently in the low-skill category than whites initially observed to be high skilled. The average test score gain of blacks will be lower than the average gain for whites for students initially observed to be high skilled. Thus, even if the growth in true achievement for blacks and whites in the top category is the same, measurement errors and regression to the mean lead the observed test score growth of blacks in the top category to be below that for whites.

Just the same holds for students initially observed to be low skilled. Whites are more likely to be erroneously placed in the low category due to a negative error (as opposed to true low skill) and more likely to be observed in the higher-skilled category in the second year than blacks even in the absence of any race differences in true skill acquisition. Thus, expected gains for blacks are lower than expected gains for whites with the same low skill observed initially and the same growth in true knowledge.

This pattern, described more formally in Appendix B,[22] invalidates the simple categorization of students on the basis of initial mathematics test scores as a method to match blacks and whites with identical initial skills. To address this problem, we first use a test in a different subject to categorize students by initial skill level, based on the assumptions of positive correlations across subjects in true skill and of no correlation in the test measurement errors across subjects. If these assumptions are true, this approach severs the link between initial category and expected difference in the error realizations for the initial and subsequent periods. However, the extent of teaching to the test, the probability of cheating (particularly by teachers), and the likelihood a student is ill (different subject tests are administered on adjacent days) are probably similar across subjects, raising the possibility of correlated errors and potentially invalidating this approach to categorizing students.

Table 3 reports the evolution of the black–white achievement gap for all students combined and by initial test score quartile.[23] Note first the very large gap increases between third and fourth grades in the top half of the distribution. These jumps in observed gaps raise concerns that the errors in the mathematics and reading tests

[22] All appendices are available at the end of this article as it appears in JPAM online. Go to the publisher's Web site and use the search engine to locate the article at http://www3.interscience.wiley.com/cgi-bin/jhome/34787.

[23] The determination of the group cutoffs provides an additional issue to resolve. Because the distribution for whites has less dispersion than that for blacks and because there are far fewer blacks in the sample, we chose to divide the sample on the basis of the 25th, 50th, and 75th percentile of third-grade test scores computed only over the sample of blacks. This leaves blacks split almost evenly into the four categories, while whites are concentrated in the highest quartile.

Table 3. Texas public school mean black–white mathematics test score gap for intact cohorts by third-grade reading test quartile.[a]

	Grade				Observations	
	3	4	5	8	Blacks	Whites
Overall gap	0.59	0.62	0.65	0.70	89,563	344,833
Third-grade reading quartile						
Lowest	0.49	0.52	0.57	0.54	22,491	30,258
2nd	0.31	0.38	0.42	0.46	21,840	52,284
3rd	0.23	0.34	0.36	0.45	21,863	85,961
Highest	0.20	0.34	0.38	0.48	22,676	173,582

[a] The third-grade reading test score quartile is determined by the distribution of scores by blacks. All TAAS test scores are standardized scores with mean zero and standard deviation of one in each grade and year; all students must stay in appropriate grade and have valid mathematics test scores for grades three through eight.

are positively correlated, a condition that would introduce upward bias into our estimate of the change in the achievement gap. Given both *a priori* concerns about correlated third-grade errors across subjects and the observed pattern, we focus on quartile differences between grades 4 and 5 and between grades 5 and 8. It is also important to note that within-quartile gaps in third grade are not meaningful; it is only the changes with age that provide useful information.

Several clear patterns emerge. First, there is a large gap in mean test scores as of third grade that rises from 0.59 standard deviations to 0.65 standard deviations in fifth grade and to 0.70 standard deviations in eighth grade, a very different pattern than that observed in Clotfelter et al. (2008). Second, the increases in the racial gap between grades 4 and 5 are similar across the initial achievement distribution. Third, the racial achievement gap increases much more between grades 5 and 8 for students in the top two initial achievement quartiles than for those in the lower two quartiles (and where the gap declines for students in the lowest initial quartile). Overall, between grades 4 and 8 the gap increases by 0.02 standard deviations for students in the lowest initial achievement group, by 0.08 for students in the second group, by 0.11 for students in the third group, and by 0.14 for students in the top initial achievement quartile.[24]

The changes reported in Table 3 highlight race differences in achievement trajectories for students who begin with very similar scores, information not provided by the repeated cross sections showing the shares of whites outscoring specific percentiles in the black test score distribution. The finding that increases in gaps are largest for students at the top of the initial test score distribution raises the possibility that segregation and school factors exert larger adverse effects on initially high-achieving blacks than on blacks further down the initial achievement distribution.

THE IMPACT OF TEACHER AND PEER EFFECTS ON ACHIEVEMENT GAPS

We now turn to whether specific teacher and school variables account for the observed growth in the achievement gap during elementary school, paying particular attention to differences across the initial achievement distribution. Our primary goal

[24] Note that the changes in Table 3 are measured in standard deviation units, while those in Table 2 reflect changes in the share of whites scoring below a particular percentile of the black distribution.

Table 4. Key characteristics of elementary schools and peers by race and initial achievement quartile, measured in fourth grade.

	Initial Reading Quartile (Third Grade)			
	Lowest	2nd	3rd	Highest
Blacks				
Black schoolmates	38.8%	38.2%	37.9%	40.0%
Hispanic schoolmates	21.9%	21.2%	20.4%	19.7%
Teachers with zero years of experience	8.7%	8.3%	8.1%	8.5%
Teachers with one year of experience	8.2%	8.0%	7.8%	7.5%
Average class size	19.4	19.5	19.6	19.7
Whites				
Black schoolmates	9.3%	9.0%	8.7%	8.4%
Hispanic schoolmates	19.4%	18.1%	16.9%	16.0%
Teachers with zero years of experience	6.8%	6.4%	6.1%	6.0%
Teachers with one year of experience	6.7%	6.5%	6.2%	5.9%
Average class size	19.5	19.5	19.6	19.6

is to assess whether schools have a discernible impact on the pattern and growth of the racial achievement gap. As noted, there remains an active debate over the contribution of schools to the racial achievement gap, with prior work even suggesting uncertainty about whether the racial gap expands or contracts during the school-age years. In addition, since the earliest work on achievement gaps in Coleman et al. (1966), pinpointing any specific school influences has proven difficult.

For schools to have an impact on the racial achievement gap (positively or negatively), significant school factors either must have differential incidence by race or must differentially impact black and white students, or both. Most attention has gone to the first case—school factors that are unevenly distributed by race—and this has motivated both policy and analysis.[25] Our investigation is guided by previous work into the determinants of achievement differences among students. We focus most attention on teacher experience and the racial concentration of schools on the basis of prior work that has shown both to be significant determinants of achievement and to be distributed differently by race. We also consider the effects of proportion Hispanic and class size, though these are less likely candidates for explaining racial achievement difference in Texas because of the small race differences in their distributions.[26]

The potential import of these factors is readily apparent in Table 4, which describes their distribution in fourth grade. Blacks are more likely than whites to have teachers with little or no experience and on average attend school with a much

[25] A traditional approach to assessing the potential importance of school factors involves the decomposition of the achievement gap into a within- and a between-school component, where the between-school component is assumed to provide an upper bound on the impact of schools. Simple nonparametric decompositions of the gap based on a regression with school fixed effects suggests that roughly two-thirds of the changes in the gap occurs between schools in Texas (not reported), a somewhat different finding than that reported for similar decompositions in Fryer and Levitt (2005). However, because of the central importance of family locational decisions, student heterogeneity is likely to contribute significantly to between-school differences, and thus such decompositions do not provide compelling evidence on the role of schools.

[26] Note that by concentrating on these two factors we ignore other components of schools, such as school leadership, that are likely distributed more favorably for whites than blacks. We return to this as follows.

higher black enrollment share, regardless of initial achievement quartile. On the other hand, there is a relatively small difference in the percent of Hispanic classmates for blacks and whites, and the average class sizes are virtually identical. It is also interesting that these characteristics exhibit little systematic variation across quartiles of the achievement distribution for blacks and only modest variation for whites—indicating that any heterogeneity of outcomes by quartile must come from differential impacts on high or low achievers.

Basic Empirical Model

Identification of the effects of teacher and peer characteristics on achievement is difficult primarily because the distribution of peer and teacher variables is not an accident but rather an outcome of government, teacher, and family choices. The endogeneity of teacher and peer variables increases the probability that there are a number of confounding factors that must be accounted for in order to isolate exogenous variation in these variables. Because data limitations impede efforts to control for all family, teacher, and school influences on learning, it is unlikely that statistical adjustments relying only on measured family and school factors will account fully for school, community, and family influences related to both achievement and the school and peer variables of interest.

Social scientists have become increasingly aware of the difficulty of identifying the causal impact of different potential policy instruments through simple observational data. A combination of unmeasured but correlated influences on outcomes plus self-selection into many samples has led to increasing care with estimation in situations like this. The use of random assignment of treatments—common not only in many medical situations such as drug trials but also in social situations such as the provision of family welfare programs—has gained increasing favor because it explicitly protects against many such problems through experimental design, lessening the potential for biased estimation of policy impacts.[27]

We follow a different line of attack. Our approach takes advantage of the repeated test score observations for each student and multiple grades and years of data for each school by employing panel data methods to account for student, family, teacher, and school factors that might otherwise bias the estimates. Specifically, we include prior mathematics test scores in order to account for student heterogeneity and multiple levels of school, grade, and year fixed effects to account for unobserved grade and time-invariant school characteristics. The objective is to ensure that any variation in teacher and peer characteristics that remains once the observed variables and fixed effects have been accounted for is not systematically related to any unmeasured factors affecting achievement, that is, to the error term in statistical models. The underlying idea is to emulate the key aspect of random assignment design in a situation where we cannot randomly assign students to different teachers and peers.

Equation (1) highlights the key identification issues that must be addressed in the absence of random assignment. Here achievement (A_{iGsy}) for student i in grade G and school s in year y is modeled as a function of student, family, school, and peer factors at that time:

$$A_{iGsy} = \alpha_{iGy} + \beta X_{iGsy} + \delta S_{iGsy} + \lambda P_{iGsy} + e_{iGsy} \tag{1}$$

where P is peer composition (proportion black and proportion Hispanic), S is school quality (including teacher quality) in grade G, X is a vector of flows of

[27] Campbell and Stanley (1963).

contemporaneous family background inputs during grade G, α is an individual-specific intercept particular to grade G in year y that captures the prior effects of family, neighborhood, and school experiences and characterizes the knowledge and skills that each student separately brings at entry to grade G, and e is a stochastic term capturing other unmeasured influences.

If P and S were uncorrelated with e and α, simple cross-sectional regressions would yield unbiased estimates of the effects of peer and school characteristics. But the complications inherent in the sorting of students, teachers, and administrators among schools—combined with existing evidence about the importance of each—strongly suggest that typically available variables contained in X will not account adequately for potentially confounding factors and that other methods and variables must be used to identify the effects of the peer and school characteristics.[28]

Incorporating School and Grade Fixed Effects

A key element of our empirical strategy is the use of an array of fixed effects in the statistical models to eliminate the most difficult measurement and selection issues. Fixed effects are simply a complete set of separate intercepts for a given factor; thus, for example, a "school fixed effect" would involve a separate intercept for each school in the sample. Implicitly, a fixed effect eliminates variations in outcomes and in explanatory factors between the different observations—that is, between schools in the example. As such, school fixed effect models use just within-school variation over time, and any average differences across schools are subsumed in the separate intercepts.

Our estimation actually takes this idea further. The variation used to identify the parameter estimates can be illustrated by considering the impact of proportion black on achievement for a single school. (In a more general case with multiple schools, the coefficients would reflect the average of these within-school relationships across the sample).[29] With multiple years of data for any single grade in the given school, we could use cohort-to-cohort differences in achievement and racial composition (that is, differences across years) to identify the proportion black effect. School-by-grade fixed effects use this type of variation to identify the coefficients. These comparisons across cohorts within the single school implicitly hold constant any stable differences across years in elements of curriculum, teachers, and leadership, whether readily measured or not. A finding that the achievement of blacks tended to fall as the black enrollment share increased would provide evidence that a higher proportion black adversely affects achievement. Yet, while this approach eliminates stable grade- and school-specific factors, it does not account for differences in other unobserved school and neighborhood determinants of achievement, including changes in the neighborhood environment or school principal, that could be related to changes in proportion black.

Alternatively, with multiple grades of data for a single year, we could use grade differences in achievement and in proportion black within the school to identify the racial composition effect. School-by-year fixed effects use this type of variation to identify the coefficients. Again, a finding that the achievement of blacks tended to fall as the black enrollment share increased would provide evidence that a higher proportion black adversely affects achievement. As was the case with the school-by-grade fixed effect model, this is not without potential problems. While eliminating time-varying school and neighborhood factors, this approach does not account for changes

[28] A more formal development of the model and estimation is found in Hanushek and Rivkin (2008).
[29] We use observations of multiple schools in the estimation to account for any grade-specific, statewide changes in policy, curriculum, or the difficulty of tests. Grade-by-year fixed effects control for such grade-year factors.

in achievement and proportion black that evolve systematically with age, as would be the case if the black enrollment share tended to increase (perhaps due to higher rates of white flight to private schools in higher grades) and achievement growth tended to be lower as black students progress through school.

Data for multiple years and grades also permit the inclusion of both school-by-grade and school-by-year fixed effects, in which case the racial composition effects are identified by deviations of black student concentrations from the school average for each grade and year. If the deviations from the grade and year averages are unrelated to other factors that systematically affect achievement, this approach can be viewed as a "quasi-experiment" in which the deviation of proportion black from the school average for a particular grade and year is randomly assigned to school-grade combinations each year.

The availability of multiple schools not only generates a large number of "quasi-experiments"; it also enables us to control for systematic changes over time at the state level, including changes in test difficulty for each grade. Ultimately, the racial composition variation that remains after eliminating the systematic elements of parental choice of schools, of specific leadership, of neighborhood conditions, and the like results from what we assume to be "random" differences across cohorts.

Additional Strategies to Minimize Bias

If, however, the remaining variation is related to other determinants of achievement, inclusion of the multiple levels of school fixed effects will not control fully for confounding factors, and additional actions must be taken. Within-school differences in racial composition come from (1) persistent cohort differences that appear upon cohort entry into school, and (2) student mobility–induced year-to-year changes in racial composition for a given cohort. Consideration of these sources of within-school variation in school proportion black suggests three factors that potentially contaminate the estimates if not incorporated into the estimation: (1) student heterogeneity, (2) adverse effects of mobility and student turnover, and (3) unobserved changes in teacher quality.

Because differences in family background and student circumstances may be associated with both sources of variation, it is necessary to account directly for student heterogeneity. We include indicators for economic disadvantage and receipt of special education. In addition, we include prior-year test score as a measure of academic readiness at the start of the school year in order to account for the cumulative influences of prior family, school, and peer factors on current achievement.[30] The lagged achievement model does not, however, capture all possible effects of differences in unobserved ability, and a key identifying assumption is that remaining student heterogeneity is orthogonal to the deviations in racial composition from grade and year averages for each school. This assumption would be violated, for example, if we were using classroom racial composition and students were being purposefully sorted among classrooms in a manner that introduced a relationship between racial composition and unobserved ability.[31] This type of sorting is particularly relevant when considering race and achievement because of observed tendencies to segregate students within schools (Clotfelter, Ladd, & Vigdor, 2003). We circumvent this problem by using grade average rather than classroom variation in proportion black and other peer and teacher characteristics, meaning that only grade average differences among and within cohorts are used to identify the estimates.

[30] Boardman and Murnane (1979) and Todd and Wolpin (2003) also highlight the importance of unobserved ability and the cumulative nature of learning. See our more formal development in Hanushek, Kain, and Rivkin (forthcoming).
[31] Rothstein (2008) raises this issue in his critiques of the estimation of teacher quality.

Mobility-induced changes, though frequently ignored in research based on cohort comparisons, may also introduce potentially serious problems.[32] Hanushek, Kain, and Rivkin (2004) show that blacks are much more likely to change schools than whites and thus to contribute disproportionately to year-to-year changes in school racial composition. In order to purge the contaminating influences of mobility, we control directly for the effects of moving on school changers with a vector of variables that allow for different effects by timing, number, and type of move.[33] Moreover, the evidence shows that movers tend to have lower prior achievement, another source of student heterogeneity that must be accounted for by lagged achievement.

Finally, unobserved changes in teacher quality may contaminate the estimates if they are related to deviations in measured peer and school factors from their average for each grade and year. Similar to the case for unobserved student skill, the purposeful allocation of students among classrooms provides the primary source of association between unobserved teacher quality and racial composition, and the use of grade aggregate peer and student characteristics avoids problems introduced by any such purposeful sorting.[34]

Remaining Concerns with the Analytical Strategy

A significant question given the inclusion of the array of fixed effects is whether adequate variation remains in the relevant explanatory factors in order to obtain reliable estimates of their impacts. Table 5 provides direct evidence on this for proportion black. For both elementary and middle schools, the standard deviation of proportion black is over 20 percentage points, but most of this comes from variations across schools—variation that will incorporate parental residential and school choices and a variety of other things that are difficult to consider explicitly. Eliminating this variation between schools (by including school fixed effects and the vector of other variables) reduces the relevant standard deviation to 3.5 percent in elementary schools and 2.8 percent in middle schools. The inclusion of school-by-grade and school-by-year fixed effects further reduces the residual variation to 1.7 percent and 1.4 percent for elementary and middle schools, respectively. The implication of these calculations is that roughly 95 percent of values of residual proportion black relevant for estimation of the full model will fall within a band of about 7 percentage points (that is, plus or minus two standard deviations). The precision with which the coefficients can be estimated depends directly on the relevant variation in each of the explanatory factors, implying that it is very difficult to separate the effect of racial concentration if the variation in it is too limited. It is important that the standard errors on the coefficient estimates for the key variables also depend on the sample size, so that the large samples that come from having all of the

[32] An identifying assumption in a number of studies that make use of cohort differences is that either raw cohort differences or differences remaining following the removal of school specific trends over time are not correlated with confounding factors. This approach, which builds on the intuition that students close in age in the same school have many similar experiences, has been used in a variety of circumstances (for example, Ehrenberg & Brewer, 1995; Ferguson & Ladd, 1996; and more recently generalized by Hoxby, 2000).

[33] Indicator variables differentiate both among those moving during the summer, school year, or at least twice in the same year and among within district changes, district changes within geographic region, and moves across regions.

[34] A second way in which teacher differences could enter would be through parental mobility induced by specific teacher assignments or classroom assignments. If parents selectively withdraw students from a school because of the particular classroom assignment, a variant of selectivity bias could be introduced. We do not believe, however, that this is likely to be important because classroom assignments are often made immediately prior to the beginning of the school year, limiting alternatives for parents. Further, the typical family has more than one child in a school, making individual student placement only part of the decision problem.

Table 5. Standard deviation of percent black students after removal of school, grade, and year fixed effects.[a]

	Standard Deviation	
	Elementary School	Middle School
Overall (no fixed effects)	23.3	20.3
School fixed effects	3.5	2.8
School-by-grade fixed effects	3.3	2.6
School-by-grade and school-by-year fixed effects	1.7	1.4

[a] Calculations come from regressions of percent black on the identified fixed effects and covariates. The quantities are the residual standard deviation of percent black.

students in multiple cohorts for the entire state of Texas can help to detect effects even when the residual variation is small. Nonetheless, the inclusion of multiple fixed effects can exacerbate measurement error even as it mitigates omitted variables bias, and this must be kept in mind when examining the pattern of estimates.

A related concern is the possibility that some of the pathways through which racial composition could effect achievement are eliminated by the multiple levels of school fixed effects. For example, if higher proportion black schools tend to have lower teacher quality due to greater difficulty attracting and retaining teachers, this would introduce a negative relationship between achievement and proportion black that would be eliminated by the fixed effects. Whether this is an empirically important phenomenon is unclear given the difficulty of isolating the proportion black effect on achievement with cross-sectional data and the general difficulty of identifying the effect of proportion black on teacher hiring and retention, given the correlation among proportion black, proportion poor, and other determinants of working conditions. Nonetheless, it is plausible to believe that any estimates we obtain are lower bounds on the true impacts of racial concentration in schools.

Estimation Results

The analysis focuses on the effects of racial composition and initial teacher experience; class size is also included in all specifications, though the small differences by race rule out a sizeable role for class size in explaining growth in the achievement gap unless there is a substantial differential impact by race. We also considered other factors, including teacher education and average teacher experience.[35] Consistent with prior work, however, no teacher education variable or measure of experience beyond the initial years was a significant determinant of achievement, and their exclusion from the analysis had virtually no effect on the other coefficients. As noted earlier, the effects of all variables are allowed to differ by race and achievement quartile in order to allow for heterogeneous effects.

Table 6 reports estimates by race and initial achievement quartile and robust standard errors clustered by school for three specifications that progressively add school-by-grade and school-by-year fixed effects estimated separately for elementary

[35] Rivkin, Hanushek, and Kain (2005) find that teacher experience is important in the first two years of a teaching career (but not thereafter) and that class size has small effects in earlier grades. These patterns are consistent with a number of other high-quality recent works including Rockoff (2004), Boyd et al. (2006), and Kane, Rockoff, and Staiger (2008). Hanushek, Kain, and Rivkin (forthcoming) find increased concentration of black students has a particularly deleterious effect on black achievement. This finding is consistent with Guryan (2004), Angrist and Lang (2004), and Hanushek and Raymond (2005).

Table 6. Estimated effects of racial composition and teacher experience on math achievement by race and initial test score quartile.

	Elementary School			Middle School		
Fixed Effects	(1)	(2)	(3)	(4)	(5)	(6)
School-by-grade	no	yes	yes	no	yes	yes
School-by-year	no	no	yes	no	no	yes
1. Proportion Black						
Blacks						
Bottom quartile	0.01	−0.01	−0.02	−0.16***	−0.24***	−0.10
Second quartile	−0.02	−0.04	−0.05	−0.12***	−0.21***	−0.08
Third quartile	−0.04***	−0.06**	−0.07	−0.16***	−0.24***	−0.11
Top quartile	−0.14***	−0.16***	−0.15**	−0.19***	−0.28***	−0.16**
Whites						
Bottom quartile	−0.10***	−0.09***	−0.11	−0.04*	−0.14***	−0.03
Second quartile	−0.05***	−0.04	−0.04	−0.07***	−0.18***	−0.07
Third quartile	−0.02*	0.00	−0.01	−0.04***	−0.13***	−0.02
Top quartile	−0.03***	0.01	0.00	−0.03***	−0.11***	0.00
2. Proportion Hispanic						
Blacks						
Bottom quartile	0.09***	0.09***	0.10*	−0.07***	−0.02	0.04
Second quartile	0.04**	0.04	0.05	−0.04**	0.00	0.06
Third quartile	0.05***	0.05*	0.06	−0.05***	−0.02	0.04
Top quartile	0.01	0.00	0.01	−0.07***	−0.04	0.01
Whites						
Bottom quartile	0.05***	0.01	0.02	−0.01	0.00	0.06
Second quartile	0.05***	0.02	0.03	−0.02*	0.00	0.06
Third quartile	0.06***	0.02	0.03	−0.01*	0.01	0.06
Top quartile	0.04***	0.00	0.02	−0.01**	0.02	0.07
3. Proportion teachers with 0 years of experience						
Blacks						
Bottom quartile	−0.09***	−0.12***	−0.10**	−0.07***	−0.04**	−0.07***
Second quartile	−0.09***	−0.10***	−0.09***	−0.06***	−0.05***	−0.08***
Third quartile	−0.13***	−0.14***	−0.12***	−0.02	−0.02	−0.05**
Top quartile	−0.16***	−0.15***	−0.09***	−0.01	−0.01	−0.04**
Whites						
Bottom quartile	−0.20***	−0.17***	−0.21***	−0.07***	−0.06***	−0.09***
Second quartile	−0.13***	−0.12***	−0.16***	−0.03***	−0.02**	−0.05***
Third quartile	−0.09***	−0.07***	−0.11***	−0.03***	−0.03***	−0.06***
Top quartile	−0.05***	−0.04***	−0.08***	−0.02***	−0.02***	−0.04***
4. Proportion teachers with 1 year of experience						
Blacks						
Bottom quartile	−0.10***	−0.08***	−0.09**	−0.11***	−0.08***	−0.12***
Second quartile	−0.03	−0.01	−0.03	−0.03**	0.00	−0.04
Third quartile	0.03*	0.04**	0.00	−0.04**	−0.01	−0.05**
Top quartile	0.10***	0.10***	0.03	−0.03**	0.00	−0.05**
Whites						
Bottom quartile	−0.07***	−0.06***	−0.08**	0.00	0.00	−0.03
Second quartile	−0.06***	−0.05***	−0.07***	−0.04***	−0.03***	−0.06***
Third quartile	−0.03***	−0.03**	−0.04**	−0.03***	−0.03***	−0.05***
Top quartile	−0.03***	−0.02**	−0.04**	−0.01**	−0.01	−0.03***
Observations		658,723			727,504	

Note: Asterisks indicate significance at the following levels: * (0.1), ** (0.05), and *** (0.01). Significance levels are calculated using robust standard errors clustered by school. All specifications include black and female indicators, indicators for a transition to junior high, subsidized lunch eligibility, special education participation, and a non-structural move (all fully interacted with black and initial achievement quartile), and a full set of grade-by-year variables.

and middle schools.[36] In addition to the reported estimates, all specifications include indicators for subsidized lunch eligibility, participation in special education, female, a family-initiated move, and a transition to middle school (in the middle school regressions), the share of students who are new to the school, and the share of teachers with two years of experience, all fully interacted by initial achievement quartile and race.

The first specifications (column 1 for elementary schools and column 4 for middle schools) are provided for comparison purposes with alternative cross-sectional analyses and with the more preferred specifications that include fixed effects. Because between-school differences in racial composition and teacher experience are associated with family background and other determinants of achievement, these estimates do not identify causal effects. Therefore, we focus on the fixed effect estimates reported in columns 2 and 3 and columns 5 and 6.

The estimates in the top panel of Table 6 reveal substantial differences in the effects of proportion black by initial achievement quartile, race, and schooling level. We focus first on elementary schools. Here the inclusion of the school-by-grade and school-by-year fixed effects has little impact on the magnitude of the estimates, though it does increase the standard errors (reduce the precision and significance of the estimates), as expected.[37] For blacks, all proportion black coefficients are small and insignificant, with the exception of the top quartile. That coefficient equals − 0.15 in the full model and is significantly different from zero and from each of the other proportion black coefficients at the 1 percent level. For whites, by comparison, the proportion black effect is almost always insignificant except for the bottom quartile in the fixed effects models. Because of the stability of the estimate across specifications, we take these estimates as suggesting that a higher proportion black does adversely affect achievement for whites in the bottom quartile of the distribution (even though it does not quite reach the 10 percent significance level in the full model of column 3).

The middle school estimates reveal both similarities and differences with the pattern of elementary school coefficients. As is the case with elementary school students, the most adverse proportion black effect for blacks occurs for students initially in the top achievement quartile, regardless of specification. In the full model of column 6, the hypothesis that the proportion black effect of − 0.16 for students in the top quartile is equal to the effects for students in the bottom, second, and third quartiles is rejected at the 5 percent significance level for each comparison.

Despite these similarities, there are also differences between the elementary and middle school proportion black coefficients for blacks that merit discussion, including the much larger fluctuation in magnitudes across specifications. For students in each initial achievement quartile, the addition of school-by-grade fixed effects increases the magnitude of the proportion black coefficient, while the further addition of school-by-year fixed effects reduces the magnitude below that observed in the first column. Preliminary work suggested that middle school transition difficulties for students in very high proportion black elementary schools contributed to the observed pattern. During the school transition, these students tended to experience a relatively large test score decline accompanied by a decrease in proportion black due to the combining of multiple elementary schools into a single middle school. This introduces a spurious positive relationship between achievement and proportion black that school-by-grade fixed effects eliminate. In terms of the decrease in magnitude following the addition of school-by-year fixed effects, this could result from the reduction in omitted variables bias, exacerbation of measurement error induced attenuation bias, or some combination of the two. The roughly doubling of

[36] Only black and white non-Hispanic students who remain with their cohort and have nonmissing test scores for grades 3 through 8 are included in the sample. A small number of observations are excluded because of missing information on teachers.
[37] Note that standard errors for the estimated coefficients have not been published because of space/complexity concerns in the table. The full results are available from the authors.

the standard errors (not shown) is similar to the pattern observed for the elementary school, but the substantial decrease in coefficient magnitudes clearly differs from the stability of the elementary school estimates. This lends some support to the belief that time-varying confounding factors bias estimates produced by the specifications lacking school-by-year fixed effects.

The fluctuations in the proportion black estimates for whites exceeded those observed for blacks, and none of the coefficients in the full model are significant at any conventional level for whites. Similar to the case for blacks, the increase in standard error magnitudes for the middle school estimates is quite similar to that observed for the elementary school estimates. Again, this suggests that confounding factors may bias downward the estimated effect of proportion black in specifications without school-by-year fixed effects.

In contrast to the effects observed for proportion black, estimates reported in the second panel reveal that there is little or no evidence in support of the hypothesis that a higher proportion Hispanic adversely affects achievement for either whites or blacks. Only two of the coefficients from either of the fixed effect models are significant at even the 10 percent level (both pertain to blacks in the lowest initial achievement quartile), and both are positive. This strongly suggests that it is proportion black rather than “minority concentration” that exerts an adverse achievement effect for some students.[38]

The remaining panels of Table 6 reveal a strong negative relationship between achievement and the share of teachers with little or no experience. The negative effects of rookie teachers hold for both whites and blacks across the initial achievement distribution. Although there is no systematic ordering of effect magnitudes by initial achievement quartile for blacks, a strong pattern does emerge for whites. It is particularly pronounced in elementary school, where the negative effect of a new teacher produced by the full model equals − 0.21 for those in the bottom initial quartile, − 0.16 for those in the second quartile, − 0.11 for those in the third quartile, and − .08 for those in the bottom quartile. All coefficients are significant at the 1 percent level, and *F*-tests reveal that the hypothesis of equal achievement effects of new teachers is rejected for all quartile pairings at the 5 percent level for all but comparisons between the bottom and second quartile effects. The middle school ordering is weaker and the magnitudes are smaller, but the adverse effect of a new teacher remains highest for students in the bottom initial achievement quartile. The estimated effects of teachers with only one year of prior experience reported in the bottom panel are uniformly smaller in magnitude for both black and white elementary school students regardless of initial quartile, but no such systematic differential emerges for middle school.

Interpretive Issues

The estimation does not provide answers to all of the interpretive issues that one might desire from a policy viewpoint. Most important, we do not have a good understanding of the underlying causes of the observed effects of peers and teacher experience. With teacher experience, for example, the lower average effectiveness of inexperienced teachers may reflect shortcomings in skills that could be overcome by better pre-service training or by more effective mentoring. Or they may reflect skills that can only come through on-the-job learning. Or they may reflect the fact that ineffective new teachers are more likely to exit teaching, causing average teacher quality to increase with experience. More likely, they capture some combination of these and other causes. Without knowledge of the source, the remaining policy

[38] The effect of proportion Hispanic on Hispanic children may well be quite different, but, as noted, analysis of this is beyond the scope of this paper.

options, as discussed below, will revolve around lessening the exposure of students to rookie teachers.

The bigger interpretive questions surround the underlying source of the impact of racial concentration on black achievement. A variety of alternative pathways could produce this result. One suggestion has been that it reflects teacher expectations and curricular adjustments, that is, that teachers lessen the rigor of the curriculum as the black enrollment share increases. An alternative view centers on cultural norms and is captured by the "acting white" hypothesis—that peer pressure on black students operates to downgrade efforts to learn more and to reach higher achievement levels.[39] A final concern, however, is that it simply reflects some other peer characteristics such as having lower-achieving and lower-income classmates (characteristics that are correlated with race). The alternative explanation, that differences in school or teacher quality drive the observed proportion black effect, is less plausible given that the estimates are identified solely by within-school and -year variation in proportion black; such variation is far less likely to be systematically related to differences in teacher or school quality than variation across schools or even changes over time in school average proportion black.

Data limitations make it difficult to determine directly the extent to which the proportion black effect is driven by income or academic preparation of peers. The available income variation provides a crude measure of family resources, and methodological complications impede efforts to isolate the impact of academic preparation.[40] However, the fact that the inclusion of proportion Hispanic has little impact on the proportion black coefficients despite the negative correlation of proportion Hispanic with both family income and initial academic achievement does suggest that the proportion black does not simply serve as a proxy for class or academic preparation. Nonetheless, the research design required to untangle the influences of these various factors probably involves more direct qualitative studies within schools, where behavioral differences might be observed. But, much like with the teacher experience finding, this uncertainty leads us to concentrate attention on the overall impacts of proportion black and the potential benefits of altering the distribution of school racial composition.

IMPLICATIONS OF SCHOOL EFFECTS

The combination of race differences in the school averages of proportion black and teacher experience and of the finding that these variables significantly affect achievement indicates that schools account for at least a portion of the growth in the racial achievement gap. Moreover, because the magnitudes of the effects differ by initial achievement, the impacts of these variables will vary by initial achievement as well.

To illustrate their differential importance across the initial achievement distribution, we use the estimates from the full fixed effect model reported in Table 6 (columns 3 and 6) to simulate the cumulative impact of setting the racial composition and teacher distribution faced by black students to the average in all Texas public schools. These simulations assume an annual depreciation of knowledge of 0.3

[39] Fordham and Ogbu (1986); Cook and Ludwig (1997); McWhorter (2000); Ogbu (2003); Austen-Smith and Fryer (2005); Fryer (2006).

[40] The identification of the effects of peer achievement is very difficult because of the simultaneous interactions of all students (see Manski, 2000). The problems are particularly difficult here where the estimation involves interaction of peer variables with initial achievement category. Our prior work, however, does not suggest that much of the estimated overall effect of black concentration is the result of commensurate changes in peer achievement (Hanushek, Kain, & Rivskin, forthcoming).

(which is consistent with the underlying statistical models, where the coefficient on lagged achievement score for blacks is roughly 0.7).[41]

Other than for the lowest initial achievement group that experienced a very small increase in the gap, the simulations reported in Table 7 suggest that reducing average proportion black and the shares of teachers with little or no experience to the state averages for black and white students would eliminate between 15 and 20 percent of the growth in the achievement differential between grades four and eight.[42] Indeed, the impact of schools is likely larger than we show here. We have relied on just the within-school variations in the limited number of factors considered to obtain our estimates, but these are undoubtedly just a portion of the differences in schools faced by blacks and whites. As Rivkin, Hanushek, and Kain (2005) show, easily quantifiable variables do not explain the bulk of the variance in teacher and school quality. If race differences in unobserved teacher and school quality also favor whites, the contribution of schools to the growth in the achievement gap would be correspondingly larger.

One hint at the potential magnitude comes from decomposing the growth in the achievement gap into between-school and within-school components. Although the sharp divergence in school attendance patterns by race and accompanying uncertainty about the appropriate method for weighting observations introduces some uncertainty, several different decomposition algorithms suggest that roughly two-thirds of the growth in the gap occurred between schools for all four achievement groups. Of course, family and community differences across schools also contribute to the between-school component, so this estimate should not be interpreted as the contribution of schools *per se*.[43]

THE POLICY DILEMMA

By any measure, black–white differences in schooling outcomes are a matter of enormous concern. The early progress toward racial convergence that followed *Brown v. Board of Education* and the civil rights legislation of the 1960s has slowed if not stopped over the past two decades (Neal, 2006). The implications of this slowdown for earnings inequality and the economic well-being of blacks have been magnified by the substantial increase in the return to skill experienced over the past 30 years. The differences in measured skills between blacks and whites are enormous. By age 17, the average black student is performing at around the 20th percentile of the white distribution.[44] This leads directly to large differences in college attendance and completion and ultimately to substantial black–white differences in lifetime earnings and occupational prestige.

A major thrust of governmental policy has been to deal with the black–white achievement gap through improving the quality of elementary and secondary education for disadvantaged students. But, as highlighted in Figure 1, the actions taken have been unsuccessful in closing the black–white achievement gap, which grows across grades and grows most for the initially highest achieving blacks in Texas. The main results of this study suggest that the existing distribution of new teachers and

[41] The estimation, as noted, includes each student's achievement from the prior grade to account for historical factors that influence initial levels of knowledge. The estimated coefficient on lagged achievement indicates how much of prior learning is retained. See Hanushek, Kain, and Rivkin (forthcoming) for a fuller discussion.
[42] Were the rate of depreciation to be much higher—for example, on the order of 70 percent, as suggested by Jacob, Lefgren, & Sims (2008)—the explanatory power of these three variables would be slightly more than half as large.
[43] In earlier grades, Fryer and Levitt (2005) estimate a smaller between-school share.
[44] See data on the National Assessment of Educational Progress, or NAEP (National Center for Education Statistics, 2005).

Table 7. Simulated effect on average achievement for blacks of reducing school proportion black and proportions of teachers with zero or one year of experience to the averages for blacks and whites, by initial achievement quartile.

	Equalizing Change in Characteristic		Annual Achievement Effect			
	Elementary	Middle	Elementary	Middle	Cumulative Effect Grades 4–8	% Reduction in Gap Growth
1. School proportion black						
Lowest quartile	0.066	0.057	0.0015	0.0059	0.0134	
Low middle	0.066	0.057	0.0031	0.0047	0.0113	
High middle	0.066	0.057	0.0044	0.0064	0.0154	
Highest quartile	0.071	0.060	0.0106	0.0097	0.0249	
2. Proportion of teachers with 0 years of experience						
Lowest quartile	0.0046	0.0056	0.0004	0.0004	0.0011	
Low middle	0.0046	0.0063	0.0004	0.0005	0.0013	
High middle	0.0051	0.0060	0.0006	0.0003	0.0008	
Highest quartile	0.0058	0.0072	0.0005	0.0003	0.0009	
3. Proportion of teachers with 1 year of experience						
Lowest quartile	0.0030	0.0028	0.0003	0.0003	0.0008	
Low middle	0.0035	0.0030	0.0001	0.0001	0.0003	
High middle	0.0039	0.0032	0.0000	0.0002	0.0004	
Highest quartile	0.0044	0.0032	−0.0001	0.0001	0.0003	
Total						
Lowest quartile					0.0153	77%
Low middle					0.0128	16%
High middle					0.0166	15%
Highest quartile					0.0261	19%

Note: Authors' calculations based on columns 3 and 6 of Table 6; total change in gaps from Table 3; depreciation in learning set to 0.3.

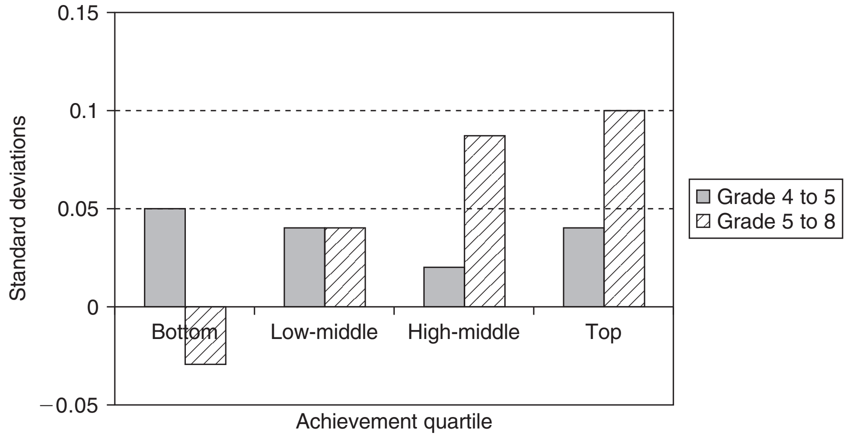

Figure 1. Change in black–white math gaps.

segregation inhibit black student progress, particularly high achievers in grade 3. Nonetheless, implications for policy remain uncertain.

At first blush, effective policy initiatives would seem straightforward. Because blacks are adversely affected by a higher likelihood of having a new teacher and racial segregation, a redistribution of students that reduces the average share of blacks' schoolmates who are black and the probability of having a teacher with little or no prior experience would reduce the achievement differential, particularly for initially high achievers.

Policy is not, however, so simple. To deal with racial concentrations, a variety of race-based policies have been applied since the *Brown* decision, but the permissible policies have been increasingly restricted and do not deal with the constraints emanating from housing and locational decisions. The evidence regarding teacher labor markets also suggests potentially high costs to policies aimed at altering the distribution of inexperienced teachers.

Perhaps the most easily identified policies revolve around ensuring that black students do not draw a disproportionate share of beginning teachers, but the effects of any particular policy depend in large part on teacher reactions. It is important to note that the underlying factors leading to teacher choices of schools are not completely understood, although there is evidence that a combination of locational preferences, working conditions, leadership qualities, and ease of the teaching challenges contribute.[45] Because teachers cannot be required to teach in particular districts, urban districts may have a very difficult time raising salaries high enough to attract experienced teachers to high-poverty, high-proportion-black schools. Any changes in the process through which districts allocate teachers to schools that work to disadvantage experienced teachers are likely to induce exit from the district unless salaries are increased, perhaps substantially.[46]

[45] Hanushek, Kain, and Rivkin (2004) document that teachers generally move away from schools with high concentrations of black students and with low student achievement levels. Boyd et al. (2005) find similar patterns with the addition of the pull of a teacher's own location while growing up. School leadership appears to contribute importantly to teacher perceptions of working conditions. Moreover, because principals appear to have preferences similar to those of teachers, the most disadvantaged schools face ongoing leadership problems (Branch, Hanushek, & Rivkin, 2009). See also the reviews of literature on the impact of working conditions in Hanushek and Rivkin (2007) and Horng (forthcoming).
[46] Hanushek, Kain, and Rivkin (2004) estimate that just equalizing the teacher outflow between central city and suburban schools in Texas would require increasing the average salary of younger teachers by over 40 percent. On the other hand, data from North Carolina suggest that teachers are perhaps more sensitive to bonuses (Clotfelter et al., 2008).

Reducing racial concentrations in the schools in order to lessen the impact of peer composition is even more problematic. The recent U.S. Supreme Court decisions concerning schools in Seattle and Louisville, following a long period of movement away from active policies aimed at school desegregation, severely limited, if not curtailed, the use of race-based considerations in district assignment of students to schools.[47] As a result, proxies for the race distribution such as family income have been proposed, but—while there are differences across metropolitan areas—this approach does not seem to be a generally powerful alternative to reducing racial concentrations within districts (Reardan, Yun, & Kurlaender, 2001).

Nonetheless, this concentration on within-district policies is a bit of a red herring. As Rivkin and Welch (2006) report, housing patterns account for the bulk of school segregation, and prior court decisions limit inter-district desegregation programs.[48] In addition, our sample covers a period without much involuntary desegregation activity, and the relationship between achievement and racial composition might depend on both programmatic and historical factors that determine school attendance patterns in a given district. Consequently, active initiatives designed to substantially increase black exposure to whites might produce a different relationship between achievement and racial composition than we identify here.

The implication is that, although we identify specific school and peer factors that systematically affect racial achievement gaps, policy directed at just these factors is unlikely to be very successful. Instead, it appears that a more comprehensive set of policies aimed at improving the quality of schools attended by blacks will be required. The substantial teacher experience effects and other research on school quality point to the importance of improving the quality of instruction. Efforts to expand the pool of potential teachers in high proportion black schools might combine deregulation (relaxed licensing requirements for new teachers), salary increases, and improvements in working conditions.

It may also be possible to pursue other policies that ameliorate the impact of racial concentration, though such efforts are hindered by not fully understanding the underlying behavior or causes of the observed impact of racial concentration. Policies to address lowered expectations of teachers with high proportion black classrooms differ from those that would seek to lessen negative peer behavior such as equating high achievement with "acting white." If identified, however, more direct policies might be possible. For example, to the extent that the adverse impact of racial concentration results from cultural norms, it may be possible to counteract these, as some charter schools appear to have been able to do.[49] Such efforts may even involve interventions during preschool years that might foster academic and social development. Nonetheless, the research base both for identifying the underlying causes of the observed relationships or for instituting approaches to deal with the causes remains uncertain.

Finally, it is crucial to recognize that test score differences do not provide adequate information for those not in the test sample, which in this case includes students retained in grade and those excused from test taking because of a disability or other circumstance.[50] Given the much higher rate of special education classification and grade retention for blacks than for whites, and for black boys in

[47]*Crystal D. Meredith v. Jefferson County Board of Education* was consolidated in the United States Supreme Court decision of *Parents Involved in Community Schools v. Seattle School District No. 1*, 127 S.Ct. 2738 (2007). See also the discussion in Linn and Welner (2007).

[48] See *Milliken v. Bradley*, 418 U.S. 717, 744–746 (1974) and *Missouri v. Jenkins,* 515 U.S. 70 (1995).

[49] While the circumstances differ by state, charter schools tend to have *higher* racial concentrations than the traditional public schools from which they draw students (see, for example, Hanushek et al., 2007). Some charter schools, such as KIPP schools and others, direct attention at cultural issues as a significant part of their educational plan (see WestEd, 2007; Thernstrom & Thernstrom, 2003; Mathews, 2009). Nonetheless, the results are not uniformly good (see, for example, Bifulco & Ladd, 2007), so it is difficult to know what kinds of policies would ensure closing the achievement gap through these approaches.

particular, the achievement comparisons do not capture fully the gap in education progress and do not illustrate the educational difficulties of many at the lower end of the achievement distribution, including those who will likely experience poor academic, social, and labor market outcomes in the future.[51]

ERIC A. HANUSHEK is a Senior Fellow at the Hoover Institution of Stanford University and Chair of the Executive Committee of the Texas Schools Project at the University of Texas at Dallas.

STEVEN G. RIVKIN is Professor of Economics at Amherst College and Director of Research of the Texas Schools Project at the University of Texas at Dallas.

ACKNOWLEDGMENTS

Support for this work has been provided by the Packard Humanities Institute.

REFERENCES

Alexander, K. L., Entwisle, D. R., & Olson, L. S. (2007). Lasting consequences of the summer learning gap. American Sociological Review, 72, 167–180.

Angrist, J. D., & Lang, K. (2004). Does school integration generate peer effects? Evidence from Boston's Metco program. American Economic Review, 94, 1613–1634.

Armor, D. J. (1995). Forced justice: School desegregation and the law. New York: Oxford University Press.

Austen-Smith, D., & Fryer, R. G., Jr. (2005). An economic analysis of "acting white." Quarterly Journal of Economics, 120, 551–583.

Bifulco, R., & Ladd, H. F. (2007). School choice, racial segregation, and test-score gaps: Evidence from North Carolina's charter school program. Journal of Policy Analysis and Management, 26, 31–56.

Boardman, A. E., & Murnane, R. J. (1979). Using panel data to improve estimates of the determinants of educational achievement. Sociology of Education, 52, 113–121.

Bowles, S., & Levin, H. M. (1968). The determinants of scholastic achievement—an appraisal of some recent evidence. Journal of Human Resources, 3, 3–24.

Boyd, D., Grossman, P., Lankford, H., Loeb, S., & Wyckoff, J. (2006). How changes in entry requirements alter the teacher workforce and affect student achievement. Education Finance and Policy, 1, 176–216.

Boyd, D., Lankford, H., Loeb, S., & Wyckoff, J. (2005). The draw of home: How teachers' preferences for proximity disadvantage urban schools. Journal of Policy Analysis and Management, 24, 113–132.

Branch, G. F., Hanushek, E. A., & Rivkin, S. G. (2009, January 3–5). Principal turnover and effectiveness. Paper presented at the American Economic Association Annual Meeting, San Francisco, CA.

Campbell, D. T., & Stanley, J. C. (1963). Experimental and quasi-experimental designs for research. Boston: Houghton Mifflin Company.

Clotfelter, C. T. (2004). After Brown: The rise and retreat of school desegregation. Princeton, NJ: Princeton University Press.

[50]See Appendix Table A1 for the quantitative magnitudes of exclusions. All appendices are available at the end of this article as it appears in JPAM online. Go to the publisher's Web site and use the search engine to locate the article at http://www3.interscience.wiley.com/cgi-bin/jhome/34787.
[51] The low rates of test taking for blacks and to a lesser extent whites among those who participate in all five waves in the ECLS sample raise the possibility that the reported growth in achievement differences understates the actual increase during the early elementary school years.

Clotfelter, C. T., Glennie, E., Ladd, H. F., & Vigdor, J. L. (2008). Would higher salaries keep teachers in high-poverty schools? Evidence from a policy intervention in North Carolina. Journal of Public Economics, 92, 1352–1370.

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2003). Segregation and resegregation in North Carolina's public school classrooms. North Carolina Law Review, 81, 1463–1511.

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. (2006). The academic achievement gap in grades 3 to 8. NBER W12207. Cambridge, MA: National Bureau of Economic Research.

Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfeld, F. D., et al. (1966). Equality of educational opportunity. Washington, DC: U.S. Government Printing Office.

Congressional Budget Office. (1986). Trends in educational achievement. Washington, DC: Congressional Budget Office.

Cook, P. J., & Ludwig, J. (1997). Weighing the "burden of 'acting white'": Are there race differences in attitudes toward education? Journal of Policy Analysis and Management, 16, 256–278.

Cook, T. D. (Ed.). (1984). School desegregation and black achievement. Washington, DC: National Institute of Education, U.S. Department of Education.

Crain, R. L., & Mahard, R. E. (1978). Desegregation and black achievement: A review of the research. Law and Contemporary Problems, 42, 17–53.

Downey, D. B., von Hippel, P. T., & Broh, B. A. (2004). Are schools the great equalizer? Cognitive inequality during the summer months and the school year. American Sociological Review, 69, 613–635.

Ehrenberg, R. G., & Brewer, D. J. (1995). Did teachers' verbal ability and race matter in the 1960s? Coleman revisited. Economics of Education Review, 14, 1–21.

Ferguson, R. F., & Ladd, H. F. (1996). How and why money matters: An analysis of Alabama schools. In H. F. Ladd (Ed.), Holding schools accountable: Performance-based reform in education (pp. 265–298). Washington, DC: Brookings.

Fordham, S., & Ogbu, J. (1986). Black students' school success: Coping with the burden of "acting white." Urban Review, 58, 54–84.

Fryer, R. G., Jr. (2006). Acting white. Education Next, 6, 53–59.

Fryer, R. G., Jr., & Levitt, S. D. (2004). Understanding the black-white test score gap in the first two years of school. Review of Economics and Statistics, 86, 447–464.

Fryer, R. G., Jr., & Levitt, S. D. (2005). The black-white test score gap through third grade. NBER W11049. Cambridge, MA: National Bureau of Economic Research.

Grissmer, D. W., Kirby, S. N., Berends, M., & Williamson, S. (1994). Student achievement and the changing American family. Santa Monica, CA: Rand Corporation.

Guryan, J. (2004). Desegregation and black dropout rates. American Economic Review, 94, 919–943.

Hanushek, E. A. (2001). Black-white achievement differences and governmental interventions. American Economic Review, 91, 24–28.

Hanushek, E. A. (2003). The failure of input-based schooling policies. Economic Journal, 113, F64–F98.

Hanushek, E. A., & Kain, J. F. (1972). On the value of "equality of educational opportunity" as a guide to public policy. In F. Mosteller & D. P. Moynihan (Eds.), On equality of educational opportunity (pp. 116–145). New York: Random House.

Hanushek, E. A., & Raymond, M. E. (2005). Does school accountability lead to improved student performance? Journal of Policy Analysis and Management, 24, 297–327.

Hanushek, E. A., & Rivkin, S. G. (2006). School quality and the black-white achievement gap. NBER W12651. Cambridge, MA: National Bureau of Economic Research.

Hanushek, E. A., & Rivkin, S. G. (2007). Pay, working conditions, and teacher quality. Future of Children, 17, 69–86.

Hanushek, E. A., & Rivkin, S. G. (2008). Harming the best: How schools affect the black-white achievement gap. NBER W14211. Cambridge, MA: National Bureau of Economic Research.

Hanushek, E. A., & Woessmann, L. (2008). The role of cognitive skills in economic development. Journal of Economic Literature, 46, 607–668.

Hanushek, E. A., Kain, J. F., & Rivkin, S. G. (2002). Inferring program effects for specialized populations: Does special education raise achievement for students with disabilities? Review of Economics and Statistics, 84, 584–599.

Hanushek, E. A., Kain, J. F., & Rivkin, S. G. (2004). Disruption versus Tiebout improvement: The costs and benefits of switching schools. Journal of Public Economics, 88, 1721–1746.

Hanushek, E. A., Kain, J. F., & Rivkin, S. G. (forthcoming). New evidence about Brown v. Board of Education: The complex effects of school racial composition on achievement. Journal of Labor Economics.

Hanushek, E. A., Kain, J. F., Rivkin, S. G., & Branch, G. F. (2007). Charter school quality and parental decision making with school choice. Journal of Public Economics, 91, 823–848.

Heyns, B. (1978). Summer learning and the effects of schooling. New York: Academic Press.

Horng, E. L. (forthcoming). Teacher tradeoffs: Disentangling teachers' preferences for working conditions and student demographics. American Educational Research Journal.

Hoxby, C. M. (2000). Peer effects in the classroom: Learning from gender and race variation. NBER W7867. Cambridge, MA: National Bureau of Economic Research.

Jacob, B., Lefgren, L., & Sims, D. (2008). The persistence of teacher-induced learning gains. NBER W14065. Cambridge, MA: National Bureau of Economic Research.

Jencks, C., & Phillips, M. (Eds.). (1998). The black-white test score gap. Washington, DC: Brookings Institution.

Kane, T. J., Rockoff, J. E., & Staiger, D. O. (2008). What does certification tell us about teacher effectiveness? Evidence from New York City. Economics of Education Review, 27, 615–631.

Linn, R. L., & Welner, K. G. (Eds.). (2007). Race-conscious policies for assigning students to schools: Social science research and the Supreme Court cases. Washington: National Academy of Education.

Manski, C. F. (2000). Economic analysis of social interactions. Journal of Economic Perspectives, 14, 115–136.

Mathews, J. (2009). Work hard. Be nice. How two inspired teachers created the most promising schools in America. Chapel Hill, NC: Algonquin Books.

McWhorter, J. H. (2000). Losing the race: Self-sabotage in black America. New York: Free Press.

Murnane, R. J., Willett, J. B., Bub, K. L., & McCartney, K. (2006). Understanding trends in racial/ethnic achievement gaps during elementary school. In G. Burtless & J. R. Pack (Eds.), Brookings-Wharton papers on urban affairs, 2006 (pp. 97–127). Washington: Brookings.

National Center for Education Statistics. (2005). NAEP 2004: Trends in academic progress: Three decades of student performance in reading and mathematics. Washington, DC: U.S. Department of Education.

Neal, D. (2006). Why has black-white skill convergence stopped? In E. A. Hanushek & F. Welch (Eds.), Handbook of the economics of education (pp. 511–576). Amsterdam: North Holland.

Neal, D., & Johnson, W. R. (1996). The role of pre-market factors in black-white differences. Journal of Political Economy, 104, 869–895.

O'Neill, J. (1990). The role of human capital in earnings differences between black and white men. Journal of Economic Perspectives, 4, 25–46.

Ogbu, J. U. (2003). Black American students in an affluent suburb: A study of academic disengagement. Mahwah, NJ: Lawrence Erlbaum Associates.

Reardan, S. F. (2008). Differential growth in the black-white achievement gap during elementary school among initially high- and low-scoring students. IREPP Working Paper #2008-07. Stanford University: Institute for Research on Education Policy and Practice.

Reardan, S. F., Yun, J. T., & Kurlaender, M. (2001). Implications of income-based school assignment policies. Educational Evaluation and Policy Analysis, 28, 49–75.

Rivkin, S. G. (1995). Black/white differences in schooling and employment. Journal of Human Resources, 30, 826–852.

Rivkin, S. G., & Welch, F. (2006). Neighborhood segregation and school integration. In E. A. Hanushek & F. Welch (Eds.), Handbook of the economics of education (pp. 1019–1049). Amsterdam: North Holland.

Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. Econometrica, 73, 417–458.

Rockoff, J. E. (2004). The impact of individual teachers on student achievement: Evidence from panel data. American Economic Review, 94, 247–252.

Rothstein, J. (2008). Teacher quality in educational production: Tracking, decay, and student achievement. NBER W14442. Cambridge, MA: National Bureau of Economic Research.

Schofield, J. W. (1995). Review of research on school desegregation's impact on elementary and secondary school students. In J. A. Banks & C. A. M. Banks (Eds.), Handbook of research on multicultural education (pp. 597–616). New York: Macmillan Publishing.

Thernstrom, A., & Thernstrom, S. (2003). No excuses: Closing the racial gap in learning. New York: Simon & Schuster.

Todd, P. E., & Wolpin, K. I. (2003). On the specification and estimation of the production function for cognitive achievement. Economic Journal, 113, F3–33.

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

Vinovskis, M. A. (1999). Do federal compensatory education programs really work? A brief historical analysis of Title I and Head Start. American Journal of Education, 107, 187–209.

Welch, F., & Light, A. (1987). New evidence on school desegregation. Washington, DC: U.S. Commission on Civil Rights.

WestEd. (2007). K–8 charter schools: Closing the achievement gap. Washington, DC: U.S. Department of Education.

APPENDIX A: EXCLUSIONS FROM STATISTICAL ANALYSIS OF ACHIEVEMENT GAPS

Table A1. Distribution of Texas public school students by test and grade retention status, by race, gender, and grade.

	Grade					Grade				
	4	5	6	7	8	4	5	6	7	8
	Blacks					Whites				
Girls										
has test score	88.2%	89.3%	88.6%	89.3%	87.8%	93.3%	94.3%	94.1%	94.0%	92.5%
no test score:										
special education	8.4%	9.1%	9.2%	7.5%	7.5%	4.1%	4.1%	4.1%	3.7%	4.0%
other	2.5%	1.0%	1.1%	1.6%	3.9%	2.2%	1.2%	1.2%	1.5%	3.0%
Retained in grade	0.8%	0.6%	1.1%	1.6%	0.8%	0.4%	0.4%	0.5%	0.8%	0.5%
Boys										
has test score	80.3%	81.2%	79.9%	80.8%	79.1%	90.3%	91.2%	90.6%	90.5%	88.9%
no test score:										
special education	15.9%	16.8%	16.8%	14.4%	14.3%	7.0%	7.0%	7.0%	6.4%	7.1%
other	2.7%	1.1%	1.4%	1.8%	5.2%	2.1%	1.2%	1.3%	1.6%	3.2%
Retained in grade	1.2%	0.9%	2.0%	3.1%	1.7%	0.6%	0.6%	1.1%	1.6%	1.0%

Source: Author calculations from TSP data.

APPENDIX B: RELATIVE EFFECTS OF MEASUREMENT ERROR ON ACHIEVEMENT GAPS.

Understanding the impact of measurement error on the analysis of the earnings gap is easy to see in a simple stylized calculation comparing black and white achievement for two ability groups. The key element is that a common measurement error will lead to different relative errors (compared to true observations) when the underlying distributions between blacks and whites differ, and this will distort comparisons of growth in achievement for blacks and whites if not taken into account. This Appendix formalizes the text discussion of division of the students by initial achievement.

The top panel of Table B1 presents an assumed two-level distribution of actual skill for blacks and whites, where blacks are more concentrated in the lower category than whites. The bottom panel describes the resulting distribution of observed test scores, where P_{ij} is the probability that somebody with true ability in category i is observed in category j. Importantly, identical conditional probabilities of misclassification are assumed for blacks and whites, based on the assumption that the distribution of test errors does not differ by race.

Given the assumptions regarding race differences in knowledge and measurement error, the bottom panel illustrates that a higher proportion of whites than blacks are misclassified into the low observed skill category, while the opposite is true for the high observed skill category. A higher percentage of the blacks in the low-achievement category are classified correctly, while a higher percentage of the whites are misclassified in the low-achievement category because of negative errors in test measurement. In contrast, a higher percentage of the whites in the high-achievement category are correctly classified, while a higher percentage of the blacks are misclassified in this high-achievement category because of positive measurement errors. This implies that the average errors for blacks in both observed achievement category are larger than the average errors of whites. If we assume that the expected measurement error in each period is zero and that test errors are uncorrelated over time, these relative differences in observed categorization by race directly affect any estimates of achievement gains over time. The expected achievement gain in the next period is higher for whites than for blacks throughout the observed initial skill distribution, because the larger negative errors of whites and positive errors of blacks are expected to disappear in the subsequent period. Therefore, a finding that the gap grows over time in each category would be expected even if there was no increase in the true knowledge differential. For this reason, a good estimate of the relative achievement growth of blacks and white requires dealing with the measurement error in the tests that leads to differential classification errors across the achievement distribution.

Table B1. Simulated observed and actual test score distributions for blacks and whites.

Skill Category	Blacks	Whites
Actual Distributions of Initial Skills		
Low	0.6	0.4
High	0.4	0.6
Observed Test Distribution with Measurement Error		
Low	$0.6*P_{LL} + 0.4*P_{HL}$	$0.4*P_{LL} + 0.6*P_{HL}$
High	$0.6*P_{LH} + 0.4*P_{HH}$	$0.4*P_{LH} + 0.6*P_{HH}$

Note: P_{ij} = probability of being in actual category i but observed as category j.