Disentangling the Racial Test Score Gap: Probing the Evidence in a Large Urban School District

Leanna Stiefel
Amy Ellen Schwartz
Ingrid Gould Ellen

Abstract

We examine the size and distribution of the gap in test scores across races within New York City public schools and the factors that explain these gaps. While gaps are partially explained by differences in student characteristics, such as poverty, differences in schools attended are also important. At the same time, substantial within-school gaps remain and are only partly explained by differences in academic preparation across students from different race groups. Controlling for differences in classrooms attended explains little of the remaining gap, suggesting little role for within-school inequities in resources. There is some evidence that school characteristics matter. Race gaps are negatively correlated with school size—implying small schools may be helpful. In addition, the trade-off between the size and experience of the teaching staff in urban schools may carry unintended consequences for within-school race gaps.

INTRODUCTION

Despite decades of research investigating racial test score disparities and continuing policy initiatives aimed at reducing these gaps, progress has been slow and disappointing. While some convergence in test scores took place during the 1970s and 1980s, the gaps have remained basically unchanged since the 1990s—white and Asian students continue to perform significantly better than black or Hispanic students on achievement tests. As one of many examples, white fourth and eighth graders score significantly higher than black or Hispanic students on the National Assessment of Education Progress (NAEP) in both reading and mathematics (U.S. Department of Education, 2003).[1]

Such disparities in test scores are troubling not only for the underlying educational inequality that they suggest,[2] but also because of the link between performance on tests and outcomes later in life. For instance, because performance on elementary and secondary school tests presage scores on college admission tests, early disparities are likely to translate into disparities in college attendance, financial aid, and performance. Moreover, several researchers have found that test scores are cor-

[1] The gaps are typically 20 to 30 points on a test ranging from 0 to 500. In addition, these same patterns emerge for other grades and for tests in other subjects as well (U.S. Department of Education, 2000).
[2] Or "inadequate" outcomes, to use the concept and term from recent state school finance court cases. See, for example, *Campaign for Fiscal Equity, Inc., v. State of New York,* 2001.

Journal of Policy Analysis and Management, Vol. 26, No. 1, 7–30 (2006)

Published by Wiley Periodicals, Inc. Published online in Wiley InterScience (www.interscience.wiley.com)
DOI: 10.1002/pam.20225

related with future labor market outcomes, such as earnings and wages (Murnane, Willett, & Levy, 1995; Neal & Johnson, 1996).[3] As Jencks and Phillips (1998) argue, reducing or eliminating these gaps in performance "would probably do more to promote [the goal of racial equality] than any other strategy that commands broad political support" (p. 4).

Policy initiatives aimed at reducing gaps have typically focused on eliminating disparities across school districts or schools. Important examples are school finance reform efforts, which, for more than 30 years, have sought to reduce disparities in spending across school districts, motivated in part by the argument that reducing such resource differences would translate into reductions in performance gaps. Perhaps more important, desegregation efforts have been motivated by the belief that integrating schools would reduce disparities in the resources and quality of schools attended by black and white students and this equalization would, in turn, reduce the performance gap.

In contrast, the recent federal No Child Left Behind Act looks for solutions *within* schools, by holding schools *per se* accountable for the racial and ethnic gaps in performance of their students and demanding reductions in these gaps over time. The legislation, however, is noticeably silent on how such reductions can be accomplished, failing to identify particular policies or programs that would be effective in reducing the race gaps within schools. Unfortunately, this silence reflects an absence of knowledge; relatively little research exists that examines within-school disparities in performance and assesses the prospects for school-level policies and programs to change them.

In this paper, we aim to take a step toward filling that gap, through examining the performance of New York City's elementary and middle school students on standardized tests and measuring and exploring the size and distribution of the racial and ethnic gaps in performance. Our analyses differ from the previous literature in two key ways. First, our data enable us to explore more thoroughly the role of various factors at the school and classroom level in shaping racial disparities and to explain the variation across schools in within-school gaps. Specifically, because our data provide a complete census of students, and link those students to their schools and classrooms, we have information on a much larger number of students per classroom and per school than is typically available.[4] Second, and equally important, we are able to complement analyses of the black-white test score gaps, standard in the literature, to explore disparities and differences across four racial/ethnic groups: whites, blacks, Hispanics, and Asians. This widened focus reflects the growing importance of Hispanics and Asians in New York and the country as a whole.[5]

[3] Neal and Johnson (1996) find that "the average marginal effect of a standard deviation of test score on log wages is roughly .2 for both men and women. Since the black mean test score for each sex is about a standard deviation lower than the corresponding white mean, the test score gaps account for large portions of the black-white log wage gaps of –.18 and –.24 found for women and men, respectively" (p. 875).

[4] Although some national databases available from the National Center for Educational Statistics (NCES), such as the National Educational Longitudinal Study (NELS) or the Early Childhood Longitudinal Study (ECLS), also sample students within their schools and sometimes within their classrooms, our datasets offer distinct advantages. They are more recent than the NELS data, and unlike ECLS, our data include eighth graders as well as younger, elementary school students. In addition, as noted already, in contrast to both of these databases, our data are a complete census of students, schools, and classrooms within a large district, which allow us to more thoroughly explore the effects of factors at all three of these levels.

[5] Despite high levels of segregation in the city, most schools educate students belonging to at least two groups, making such within-school analyses of test score gaps possible. Also note that in our data, Hispanic background is recorded as a separate, mutually exclusive racial category in the New York City data; as a result, we will sometimes describe Hispanic ethnicity as a race in the paper.

The New York City school system provides an excellent opportunity to study racial achievement gaps. It is the largest school system in the country, with more than 1.1 million students, and its student body is extremely diverse, including large numbers of students of each racial/ethnic group. During the last decade, blacks and Hispanics each represented more than one-third of the K–8 students, Asians represented about one-tenth, and whites around one-sixth. While New York City is unique is many ways, it is important to note that the racial test score gaps are similar to those found elsewhere in the United States.[6]

To preview our results, we find significant disparities in performance between race groups overall, some of which are explained by differences in student characteristics, such as poverty, and some of which are explained by differences in schools attended. Put differently, controlling for student characteristics and school fixed effects yields partial reductions in the size of the race gaps. It is interesting that measured factors do a better job of explaining Hispanic-white gaps than black-white gaps, reflecting, in large part, the important role that English language proficiency plays in explaining the performance of Hispanic students. In addition, our measured factors explain more of the gaps in elementary school (fifth grade) than in middle school (eighth grade).

At the same time, substantial within-school gaps remain and there is wide variation in these gaps across schools. In some schools, there is little or no difference between races; in others, substantial gaps exist. These within-school gaps are partly, but not fully, explained by differences in family background and academic preparation across students from different race groups, as well as by prior test scores, suggesting that there may be school features and policies that are associated with smaller gaps.

Notably, controlling for differences in classrooms attended—using classroom fixed effects—explains little of the remaining gap, suggesting little role for within-school inequities in resources. Our analyses, however, suggest some additional school characteristics that may shape racial gaps. In particular, race gaps are, in some cases, negatively correlated with school size, implying that small schools may be able to provide a more equitable education. Also, the trade-off between the size and experience of the teaching staff in urban schools may carry unintended consequences for within-school race gaps. That said, the magnitudes of the estimated effects of these potential changes are not large; our results reveal no quick fixes that are easily implemented by schools to resolve the existing gaps in a short time period. New research on how to make schools effective, or on programs or market structures that do so, may yield different results in the future, but our analyses suggest little scope for substantial reduction in race gaps with policies we have identified. Instead, we may need to look to the methods and attitudes of individual teachers, or perhaps look outside schools altogether, as James Coleman et al. (1966) suggest and Richard Rothstein (2004) has argued more recently.

SCHOOLS AND THE RACIAL TEST SCORE GAP

Researchers have offered a long list of potential explanations for test score gaps, including differences across racial groups in family background (Barton, 2003;

[6] For example, Clotfelter, Ladd, and Vigdor (2006) review a number of studies that document the gap in state tests between black and white students in elementary or middle schools and find that in Texas the gap in math is around .7; in Pasadena, California, it is around .5; and in North Carolina, it is between .77 and .83, depending on the grade. These results, on state tests, are similar to ones in New York City. See Table 1.

Conley, 2001; Scholz & Levine, 2004) and differences in social pressures and economic prospects faced by children of different races, regardless of neighborhood (Fordham & Ogbu, 1986, and Steele & Aronson, 1995, on "acting white" and stereotype threat; and Bertrand & Mullainathan, 2004, on persistence of labor market discrimination).

Some have pointed to differences in resources and quality between schools as contributing to racial disparities. The underlying notion is that school segregation results in fewer resources, less effective teachers, and/or lower performing peers in schools with larger shares of minority students (Clotfelter, Ladd, & Vigdor, 2005; Lankford, Loeb, & Wyckoff, 2002; Orfield, 2001).

More recently, researchers have begun to focus inside schools, highlighting differences in the educational experiences of students within the same schools. Students may be sorted into different classrooms, and high-quality teachers may be disproportionately placed in classrooms with white students. We might also see disparities within the same classrooms because teachers treat minority students differently due to lower expectations or some other mechanism (see Figlio, 2005). More generally, specific school features or policies may be particularly harmful to minority students or especially helpful to white students—such as tracking, or some feature of school configuration, like school size.

While there is a substantial literature aimed at quantifying the extent to which differences in socioeconomic background or individual characteristics explain the magnitude of race gaps in test scores (Hedges & Nowell, 1998, 1999), relatively few previous studies include information on schools. A notable exception, Cook and Evans (2000) use NAEP data on 13-year-olds between 1970 and 1988 and estimate that the within-school disparities are important—25% of the reduction in the black/white test score gap in that period can be attributed to shifts in family and school characteristics, while 75% can be attributed to reductions within schools. That is, three-quarters of the drop in the test score gap occurred within schools, driven, perhaps, by changes in tracking policies, enrollment in advanced and remedial classes, teacher attitudes, and/or discrimination.

In a more recent paper, using the kindergarten through first grade cohort from the ECLS, Fryer and Levitt (2004) find that the black-white test score gap is not significant for children entering kindergarten once socioeconomic status and several other background factors are controlled.[7] By the end of first grade, however, the adjusted gap increases to 0.2 standard deviations. These authors speculate that this increase could be because blacks attend lower quality schools, since the gap becomes insignificantly different from zero once school fixed effects are included. No observable school quality variables, however, explain much of the variation in test scores, and neither Asian nor Hispanic test scores exhibit similar patterns.

Todd and Wolpin (2004) examine the contributions of home and school inputs to racial test score gaps, using the National Longitudinal Survey of Youth (NLSY)-Child Sample and a fully specified production function model with child fixed effects. While they find almost no role for schools in narrowing the gaps, their data do not allow precise matching of students to schools, because school variables are measured at county or state levels.

Bali and Alvarez (2003) as well as Myers, Kim, and Mandala (2004) use large administrative databases to decompose racial test score gaps into their individual

[7] Reardon (2003), using the same ECLS dataset, finds similar results.

and school attributes. Neither study finds a large role for schools. Bali and Alvarez conclude that the small role that does exist affects all racial groups similarly, and Myers and colleagues conclude that although an unexplained portion of the race gap continues to exist beyond the effect of student, school, neighborhood, and home environment, evidence still exists to suggest that school policies such as increasing attendance, limiting school changes, and attention to gifted and special education programs can help reduce gaps.

Yet another type of evidence on the role of schools in affecting racial test scores emerges from school voucher experiments, such as those evaluated by Peterson and his co-authors. Howell, Wolf, Campbell, and Peterson (2002) report that African-American students who were randomly selected to receive vouchers to switch from public to private schools gained more on tests than their counterparts who did not receive vouchers and remained in public schools. This effect was not evident for students of any other race or ethnicity and reasons for its existence could not be determined, although the result does seem to indicate that differences across schools play some part in driving racial disparities.

A few authors have looked specifically at differences in treatment across races within schools. Boozer and Rouse (2001), for instance, provide evidence that in the early 1990s, black students attended larger classes than whites within the same schools. Ferguson (1998) presents evidence showing that blacks are disproportionately placed in lower-performing classes. Despite some fierce criticism levied at such ability grouping, however, he argues that its consequences remain unclear. Case studies and evidence based on suspensions indicate that minority students may be more harshly disciplined than their non-minority peers (Ferguson, 2000).

Taken together, then, the previous results suggest that socioeconomic status explains a significant amount of the test score gap, that quality differences in the schools attended by students of different races may contribute to gaps, and that within-school gaps remain, but their genesis is unclear. In this study, we depart from past studies to analyze the variation in within-school racial gaps. Our key question is the extent to which schools and school policies contribute to, and can ultimately help to ameliorate, racial and ethnic disparities in performance.

ANALYTICAL STRATEGIES

We employ several related regression strategies to study the role of schools in explaining racial test score gaps among elementary and middle school students in New York City. Elementary school analyses focus on fifth grade students. Middle school analyses focus on eighth grade students. The core of our analysis is based on the following model of individual student test scores:

$$\text{Test}_{ij} = \beta_0 + \beta_1\text{Black}_i + \beta_2\text{Hispanic}_i + \beta_3\text{Asian}_i + \beta_4\,\text{SES}_{ij} + \beta_5\,\text{Ed}_{ij} + u_j + \varepsilon_{ij} \qquad (1)$$

where i indexes individuals and j indexes schools, Test_{ij} is the normalized score on a citywide math or reading test; SES_{ij} is a vector of socioeconomic and demographic variables representing the student's poverty status, age, gender, immigrant status, and whether English is the primary language spoken at home; Ed_{ij} is a vector of individual schooling characteristics capturing English language proficiency, including an indicator for whether a student took a Language Assessment Battery (LAB) and his/her score on that test, participation in special education, mobility across schools, and attendance; u_j is a vector of school fixed effects and

ε_{ij} is an error term with the usual properties.[8] We estimate the model using cross-sectional data and, as described in greater detail below, wherever possible, we estimate each model for both the fifth grade sample and the eighth grade sample.

In this model, coefficients on the race variables capture the mean difference in performance between students of the corresponding race and white students, controlling for other factors in the equations. Thus, β_1 captures the regression-adjusted mean difference in performance between black and white students, β_2 the regression-adjusted mean difference in performance between Hispanic and white students, and β_3 the regression-adjusted mean difference between Asian and white students. As described below, we estimate alternative specifications of this model, beginning with a parsimonious specification with no controls or fixed effects—only the race variables—in order to capture the raw disparities in the test scores between races. A second specification includes control variables only. In a third specification, we include school fixed effects and the coefficients capture the mean regression-adjusted, within-school differences between the students of the indicated race group and white students—that is, the within-school race gap.[9]

Of course, these mean differences may conceal considerable variation across schools in the magnitude of the gaps within schools. Some schools might have large disparities between the performance of their white and black students, others might have parity. Before turning to our examination of the potential causes of the disparities, we explore the distribution of racial gaps across schools. Are the overall gaps a reflection of large gaps in a small number of schools, or are the gaps approximately the same in all schools? To explore this, we estimate specifications with a set of race-specific school fixed effects that yield, for each school, measures of the gap in performance between students of different races.

In the rest of the paper, we explore possible explanations for these within-school gaps. We first consider whether the observed disparities are due to differences in the prior academic performance of students, by introducing the lagged test score as an independent variable to create a value-added specification. Are the gaps merely a reflection of differences between the students' "last year" or did the school exacerbate (or ameliorate) the gap with differences in value added this year?

We then examine the possibility that within-school gaps are explained by differences in the classrooms attended by students of different races. Is within-school segregation, and corresponding inequity in resources or peers, for example, driving the differences in outcomes? One way to approach this would be to add variables on the characteristics of classrooms to Equation 1. An alternative way, adopted here, is to substitute classroom fixed effects for the school fixed effects. The classroom fixed effects capture all of the classroom specific attributes, including the characteristics of the teachers, class size, average test scores, and racial composition. (Note that this specification is only estimated for the fifth grade students

[8] The models include attendance as an explanatory variable despite its potential endogeneity. While it is possible that higher performance may influence attendance, this seems unlikely to be important in this context for two reasons. First, parents largely make attendance decisions for younger children and it is not clear how student performance would influence those decisions. Second, our measure captures average attendance over the whole school year, which largely precedes the student or family receipt of the test score data, typically available in June. Thus, we interpret attendance as capturing a range of home and family characteristics.

[9] Note that this carries the implication that the test scores of students in racially homogeneous schools do not contribute to the estimate of the race gap—the absence of within-school variation in race means there is no within-school gap. At the extreme, if all schools were racially homogenoeus, the estimates of within school gaps would be zero by construction.

because students in the eighth grade "travel" from one subject area class to another during the day, with varying degrees of constancy in their classmates.)

Finally, we turn our attention to school-level variables. To do so, we estimate the impact of school-level variables on the value-added of students in each school-race group using a hierarchical linear model or, put differently, a random coefficients model. As an example, we estimate the relationship between race-group performance within a school and school size.

DATA AND SAMPLE

This paper utilizes a dataset provided by the New York City Department of Education (NYCDOE) that includes individual-level socioeconomic and educational information on every fifth and eighth grade student in the New York City public school system in the 2000–01 school year, as well as an identifier for the schools and classrooms in which they are registered, and lagged test score data.[10]

For our analyses, we exclude students in full-time special education programs, of "other or Native American" race, who do not take an exam in the current year, or who have missing or unknown birthplace data.[11] All together these exclusions eliminate fewer than 7,600 (11,500) fifth (eighth) graders or around 9.7% (17%) of the 2000–01 fifth (eighth) grade sample.[12]

As shown in Table 1, our 2000-01 sample includes 70,638 fifth graders and 55,921 eighth graders.[13] Among the fifth graders (eighth graders), 11% (11.4%) are Asian, 36.1% (36.3%) black, 37.4% (34.6%) Hispanic, and 15.5% (17.7%) white. These fifth graders are distributed across 667 schools with fifth grades (elementary schools), while the eighth graders are spread across 278 schools with eighth grades (middle schools).[14] The fifth graders are distributed across nearly 2,900 individual classrooms.

Table 1 also presents the mean characteristics of students and their schools by race in fifth and eighth grades, revealing significant differences in individual demographic and educational characteristics across race groups, on average. Black and Hispanic students are more likely to be poor (as measured by their eligibility for the free lunch program), with white students having the lowest poverty rates by far. On the other hand, black and Hispanic students are less likely to be in the near-poor group, as measured by eligibility for reduced price lunch. A larger share of Asians and Hispanics live in a home where a language other than English is regularly spoken, have been assessed for English proficiency (that is, have taken the Laboratory Assessment Battery of English Language or "LAB"), and are found to be eligible for programs to address limited English proficiency (scored less than 40th percentile on the LAB).[15] Notably, Asians

[10] Results for the 1997–98 yields qualitatively similar results to the 2000–01 results shown.
[11] In the period of this study, most students in full-time special education did not take the citywide tests.
[12] Regressions of the determinants of test-taking in fifth grade and in eighth grade in 2000–01, which include all the variables in our most complete model, show almost no differences in test-taking across races.
[13] The smaller number of eighth compared to fifth graders mostly reflects student mobility out of New York City schools during middle school as well as some grade retention between fifth and eighth grades. School attendance is compulsory to age 16 in New York State, so the numbers do not reflect significant numbers of dropouts.
[14] Almost 100 schools include both fifth and eighth grades and therefore are included in the analysis of fifth graders as well as of eighth graders.
[15] The table reports the proportion of all students who scored less than the 40th percentile of scores nationwide, including students who did not take the LAB, all of whom are counted as scoring above 40%. This is done to provide a measure of Limited English Proficiency (LEP) status among all students. LEP status allows students access to special English language instruction.

Table 1a. Mean characteristics of New York City public school students and their schools, by race, 5th grade, 2000–01.

			5th Grade		
	All	Asian	Black	Hispanic	White
Students					
Age	10.56	10.42	10.61	10.60	10.44
Female	0.51	0.48	0.51	0.51	0.50
Foreign-born	0.13	0.32	0.08	0.11	0.15
Free lunch	0.76	0.65	0.84	0.87	0.38
Reduced-price lunch	0.08	0.14	0.08	0.07	0.11
Took English language LAB	0.06	0.04	0.01	0.13	0.01
<= 40th percentile LAB	0.04	0.02	0.01	0.10	0.01
LAB percentile	29.84	38.40	29.26	28.52	42.55
Other than English at home	0.43	0.75	0.06	0.76	0.29
Same school as last year	0.85	0.90	0.83	0.83	0.92
Attendance percentage	0.94	0.97	0.93	0.93	0.94
Part-time special ed.	0.09	0.05	0.08	0.10	0.09
Number of students	70,638	7,778	25,478	26,424	10,958
Percentage	100	11.0	36.1	37.4	15.5
Schools					
Enrollment	958	1,004	899	1,038	856
Teacher/pupil ratio	0.08	0.07	0.08	0.08	0.07
% licensed teachers	84.36	92.34	81.25	81.23	93.18
% teachers with master's degree	75.27	82.86	71.47	73.08	83.57
% teachers with more than 2 years in the school	63.72	69.66	60.92	62.00	69.80
Number of schools[i]			664		

Notes: The full sample consists of 667 schools, but three schools are missing teacher variables. Thus, there are 664 in this table.
Sample includes all students with reading scores.
LAB percentile is calculated based on those students who took LAB test.
3.5% of fifth graders are missing free or reduced price lunch data. Percentages in this table are calculated based on non-missing data.

(but not Hispanics) are much more likely to be foreign-born. Asians are also less likely to participate in part-time special education.

As shown in the bottom panel of Table 1, the schools attended by students differ by race as well, due, in part, to segregation in the city's schools. Most strikingly, black and Hispanic students attend schools with significantly less experienced and qualified teachers as compared to the schools attended by their white and Asian counterparts. Importantly, schools attended by blacks and Hispanics also have somewhat higher teacher-pupil ratios, compared to whites and Asians. This apparent tradeoff between the number and "quality" of teachers in schools with larger populations of poor and minority students has been observed in previous studies, but it is unclear why the tradeoff emerges and what its implications are (Rubenstein, Schwartz, Stiefel, & Bel Hadj Amor, 2006). Finally, it is worth noting that Hispanic and Asian students attend larger elementary schools on average than white and black students. One explanation is that they may be more likely to live in growing, immigrant communities. Alternatively, this may be the result of differential

Table 1b. Mean characteristics of New York City public school students and their schools, by race, 8th grade, 2000–01.

	8th Grade				
	All	Asian	Black	Hispanic	White
Students					
Age	13.52	13.41	13.56	13.57	13.42
Female	0.51	0.49	0.53	0.51	0.49
Foreign-born	0.19	0.42	0.14	0.17	0.20
Free lunch	0.74	0.64	0.81	0.87	0.38
Reduced-price lunch	0.10	0.16	0.10	0.07	0.12
Took English language LAB	0.05	0.05	0.01	0.11	0.02
<= 40th percentile LAB	0.04	0.04	0.01	0.09	0.01
LAB percentile	22.56	26.99	22.31	20.78	36.25
Other than English at home	0.42	0.71	0.07	0.76	0.31
Same school as last year	0.89	0.93	0.85	0.90	0.94
Attendance percentage	0.92	0.96	0.92	0.91	0.93
Part-time special ed.	0.08	0.04	0.08	0.10	0.08
Number of students	55,921	6,368	20,312	19,335	9,906
Percentage	100	11.4	36.3	34.6	17.7
School					
Enrollment	1,221	1,385	1,129	1,224	1,288
Teacher/pupil ratio	0.08	0.07	0.08	0.08	0.07
% licensed teachers	76.52	83.05	72.04	73.98	86.52
% teachers with master's degree	71.88	78.28	68.15	69.66	79.69
% teachers with more than 2 years in the school	60.53	66.97	56.98	58.31	67.96
Number of schools[i]			277		

Notes: The full sample consists of 278 schools, but one school is missing teacher variables. Thus, there are 277 in this table.
Sample includes all students with reading scores.
LAB percentile is calculated based on those students who took LAB test.
10.44% of 8th graders are missing free or reduced price lunch data. Percentages in this table are calculated based on non-missing data.

access to the small alternative schools that are gaining popularity in New York City and elsewhere.

Student performance on standardized tests is reported for a citywide test in reading (CTB/McGraw Hill Test of Basic Skills or New York State English Language Assessment) and mathematics (California Achievement Test, the CAT; or New York State Math Assessment). To make tests comparable across grades and years, we convert them to z scores.[16] The means for each grade are therefore 0 by construction. The appendix to this paper briefly describes the content of these two tests.

Table 2 shows the coefficients of models with only the race dummy variables and no controls, estimated for fifth and eighth graders in the 2000–01 school year. (As

[16] Z scores are calculated by subtracting the mean scaled score for all students in the grade and year from each student's scaled score and dividing by the standard deviation of scaled scores. The New York City tests used during 2000–01 are either tests used nationwide in some other districts or constructed for New York State specifically. In both cases, scale scores are normally distributed.

Table 2. New York City racial test score gaps, difference from mean white student test score, based on uncontrolled regressions, 2000–01.

	5th Grade Reading	5th Grade Math	8th Grade Reading	8th Grade Math
Black	–0.726***	–0.805***	–0.776***	–0.844***
	(0.011)	(0.011)	(0.012)	(0.011)
Hispanic	–0.712***	–0.724***	–0.785***	–0.747***
	(0.011)	(0.010)	(0.012)	(0.011)
Asian	-0.091***	0.179***	–0.054***	0.203***
	(0.013)	(0.015)	(0.017)	(0.016)
Constant[i]	0.535***	0.542***	0.554***	0.537***
(White mean)	(0.009)	(0.009)	(0.010)	(0.009)
Observations	70,638	72,004	55,921	57,787
R^2	0.09	0.14	0.12	0.16

Notes: Constant term represents mean white student test score. The other coefficients capture difference from mean white student test score.
Robust standard errors in parentheses.
* significant at 10%; ** significant at 5%; *** significant at 1%.

in Table 1, "white" is the omitted category.)[17] Black and Hispanic students do worse than white students on both reading and math tests, with differences ranging from 0.71 to 0.79 standard deviations in reading and from 0.72 to 0.84 standard deviations in math. (For data showing comparable differences on other tests, see Campbell, Harnbo, & Mazzeo, 2000; Hedges & Nowell, 1999; and Neal & Johnson, 1996). Differences between Asians and whites are considerably smaller. In general, Asian students perform slightly worse than or the same, on average, as whites in reading, but somewhat better in math, and the differences never exceed 0.20 standard deviations. Disparities are similar but somewhat greater for eighth grade students as compared to fifth grade students.

It is worth emphasizing that Table 2 shows only the differences in the mean test scores, masking the wide variation in test scores and considerable overlap among groups. Nearly one-fifth of black and Hispanic children in both fifth and eighth grades scored as well or better on reading tests than the average white student, for instance. Similarly, approximately 20 percent of white students performed worse than the average black or Hispanic student in both fifth and eighth grades. Overall, however, white and Asian students are clearly overrepresented at the top of the entire distribution of student scores and underrepresented at the bottom. (See Hedges & Nowell, 1999, for similar findings on other tests for black and white high school seniors.)

[17] Differences are consistent across other school years in New York City and are also consistent with national figures from the NAEP and other data.

REGRESSION RESULTS

Individual Characteristics and Schools

Table 3 presents two sets of regression results.[18] Columns 1 and 3 show estimation results for models including individual student characteristics for fifth and eighth graders, respectively; columns 2 and 4 add school fixed effects. As shown, the regression-adjusted racial test score gaps are considerably smaller than the unadjusted mean differences shown in Table 2. To be specific, controlling for the socioeconomic and school characteristics reduces the estimated black-white gap by about a quarter of a standard deviation (to –.479 and –.548 in fifth and eighth grades); analogous reductions in the Hispanic-white gap are even larger—somewhat over a third of a standard deviation (to –.332 and –.445 in fifth and eighth grades).

These results highlight the importance of individual characteristics in determining test score outcomes. The difference in performance between poor students (measured by free lunch eligibility) and non-poor students approaches .5 standard deviations, making the "poverty gap" almost as large as the race gap in some instances. The significantly higher incidence of poverty among black and Hispanic students, noted earlier, means that some of the origin of the race gap is outside the purview of schools. Nonetheless, our results indicate that a significant race gap remains, even after differences in poverty status are controlled. That said, partly as a result of the importance of the language variables, individual characteristics explain significantly more of the within-school Hispanic-white gap than the black-white gap. Also notice that not all of the individual characteristics are beyond the reach of schools. Students who scored below the 40th percentile on the LAB test, for example, are eligible for programs to address limited English proficiency, and these students perform considerably better than otherwise similar students. Thus, Bilingual Education or English as a Second Language programs may help to reduce the Hispanic-white gaps.[19]

As shown in columns 2 and 4, adding school fixed effects reduces the size of the black-white test scope gap and the Hispanic-white test scope gap by over a tenth of a standard deviation more (to –.300 and –.417 for blacks in fifth and eighth grades, and to –.166 and –.295 for Hispanics in fifth and eighth grades). In addition, the Asian-white gaps are, in these models, either positive or no different from zero. Thus, differences in schools explain some of the racial test score gaps. That said, the coefficients on the black and Hispanic variables remain negative and statistically significant even after school fixed effects are included, suggesting that significant within-school gaps remain for black and Hispanic students.

The Racial Test Score Gaps Within Schools

We turn next to an exploration of within-school race gaps. Is the race gap pervasive and consistent across schools, or are some schools able to achieve parity? As described above, our analysis proceeds by estimating a set of race-specific school fixed effects—essentially four fixed effects for each school, one for each race group, rather than the single school fixed effect estimated previously. For these analyses, we limit the sample to schools with at least five students of each race, in order to

[18] We present results for reading tests only. Results for math scores are qualitatively similar.

[19] It is possible that some of the other individual schooling characteristics can be influenced by school policies to raise test scores. For example, the positive coefficients on attendance and stability suggest that school programs aimed at improving attendance and/or stability might raise student test scores.

Table 3. Regression results for reading test scores, New York City public school students, 2000–01.

	5th Grade		8th Grade	
	(1)	(2)	(3)	(4)
Socioeconomic Characteristics				
Black	–0.479***	–0.300***	–0.548***	–0.417***
	(0.011)	(0.014)	(0.012)	(0.014)
Hispanic	–0.332***	–0.166***	–0.445***	–0.295***
	(0.011)	(0.013)	(0.012)	(0.013)
Asian	–0.012	0.024	–0.000	0.065***
	(0.013)	(0.014)	(0.015)	(0.016)
Free lunch	–0.464***	–0.322***	–0.483***	–0.333***
	(0.010)	(0.011)	(0.012)	(0.012)
Reduced-price lunch	–0.246***	–0.169***	–0.252***	–0.180***
	(0.014)	(0.014)	(0.016)	(0.016)
Age	–0.185***	–0.172***	–0.197***	–0.174***
	(0.007)	(0.007)	(0.007)	(0.007)
Female	0.114***	0.111***	0.276***	0.260***
	(0.007)	(0.006)	(0.007)	(0.007)
Other than English at home	–0.004	–0.014	0.025**	–0.005
	(0.009)	(0.009)	(0.010)	(0.010)
Foreign-born	0.038***	0.040***	–0.024**	0.002
	(0.011)	(0.010)	(0.010)	(0.009)
Individual Schooling Characteristics				
Took LAB	–1.968***	–1.868***	–1.741***	–1.603***
	(0.075)	(0.074)	(0.070)	(0.071)
LAB <= 40th percentile	0.497***	0.498***	0.474***	0.419***
	(0.052)	(0.052)	(0.057)	(0.058)
LAB percentile	0.026***	0.025***	0.022***	0.020***
	(0.001)	(0.001)	(0.001)	(0.001)
Part-time special education	–0.484***	–0.520***	–0.655***	–0.655***
	(0.014)	(0.014)	(0.011)	(0.011)
Same school as last year	0.114***	0.087***	0.211***	0.164***
	(0.010)	(0.012)	(0.011)	(0.012)
Attendance	0.017***	0.013***	0.019***	0.016***
	(0.001)	(0.001)	(0.000)	(0.000)
Constant	0.606***	0.644***	1.017***	0.905***
	(0.105)	(0.105)	(0.112)	(0.109)
School fixed effects	No	Yes	No	Yes
Number of fixed effects		667		278
Number of students	70,638	70,638	55,921	55,921
R^2	0.24	0.29	0.33	0.40

Notes: Robust standard errors in parentheses.
* significant at 10%; ** significant at 5%; *** significant at 1%.
Models include dummy variables to indicate whether free lunch and reduced price lunch are non–missing.

avoid calculating gaps based on too few students. Given the high degree of segregation in the city's schools, this limitation reduces our numbers of schools considerably. For example, at the fifth grade level, there are only 168 schools (down from 667) with at least five white and five black students. At the eighth grade level, the reduction is less severe (from 278 to 111 schools for white and black students, for example), due to their larger sizes and the weaker tie between neighborhood residence and school assignment.

Figures 1–4 show the distribution of raw within-school test score gaps across schools, based on regressions with race-specific school fixed effects (with no other controls). The distributions show that in most schools, black and Hispanic students in fifth and eighth grade achieve lower test scores than their white peers. But there are a significant number of schools where minority students do better than whites, or at least do equally well.

What drives these within-school gaps? One explanation is that the within-school race gaps simply reflect unmeasured differences between populations of different races.

To test for this possibility, we introduce prior test scores for each student in models of determinants of performance. Table 4 shows these results for the same models estimated in Table 3, but with prior test scores (from the previous grade and year for each student) included. Columns 1 and 4 show results with only individual controls, columns 2 and 5 show results with school fixed effects. After including prior test scores, the coefficients on the black and Hispanic dummy variables fall considerably. Within-school gaps drop by roughly one-fifth of a standard deviation for blacks and over one tenth of standard deviation for Hispanics (columns 2 and 5 in Table 4 compared to columns 2 and 4 in Table 3 or to –.106 and –.201 for blacks in fifth and eighth grades and to –.044 and –.165 for Hispanics in fifth and eighth grades). Still, gaps remain significant, especially for both black and Hispanic students in eighth grade.

Thus, while differences in academic preparation across racial groups help to explain some of the within-school racial gaps, significant disparities remain. Another explanation is that some segregation exists across classrooms within schools, which leads students of different races to experience dramatically different educational environments even within the same school walls—say, differences in teachers, class size, etc.

To test for this possibility, we substitute classroom fixed effects for school fixed effects, in the fifth grade model.[20] The classroom fixed effects control for all differences in classroom-specific variables—again, such as teachers or class size—that are experienced equally by all students within a single classroom. Results are shown in the third column of Table 4.[21] The results suggest that controlling for classroom fixed effects, once prior test score is taken into account, does little to explain racial test score gaps. In fact, the estimates of within-school racial gaps are almost identical to those for within-classroom racial gaps. This suggests that, while classroom sorting may exist, there is little evidence there are other unobserved differences in classrooms that drive the within-school race gaps, beyond the differences in student prior performance and other individual characteristics. That is, the evidence does

[20] In eighth grade, students change classes and we have data only on their homeroom assignments, which are of less use than would be their academic assignments.

[21] The sample size falls very slightly when we include classroom fixed effects because we are unable to assign 95 students to individual classrooms. Results of the school fixed-effects regressions re-estimated with this slightly smaller sample are nearly identical.

Table 4. Regression results for reading test scores, New York City public school students, 2000–01, value-added models.

	5th Grade			8th Grade	
	(1)	(2)	(3)	(4)	(5)
Prior test score	0.616***	0.608***	0.540***	0.656***	0.622***
	(0.004)	(0.004)	(0.005)	(0.005)	(0.005)
Socioeconomic Characteristics					
Black	–0.191***	–0.106***	–0.099***	–0.250***	–0.201***
	(0.009)	(0.012)	(0.012)	(0.009)	(0.011)
Hispanic	–0.104***	–0.044***	–0.044***	–0.227***	–0.165***
	(0.009)	(0.011)	(0.011)	(0.009)	(0.010)
Asian	0.004	0.016	0.011	–0.007	0.040***
	(0.011)	(0.012)	(0.011)	(0.012)	(0.012)
Free lunch	–0.149***	–0.100***	–0.076***	–0.202***	–0.151***
	(0.008)	(0.009)	(0.009)	(0.010)	(0.010)
Reduced-price lunch	–0.072***	–0.051***	–0.042***	–0.098***	–0.075***
	(0.011)	(0.011)	(0.011)	(0.013)	(0.012)
Age	–0.029***	–0.028***	–0.032***	–0.080***	–0.079***
	(0.006)	(0.006)	(0.006)	(0.005)	(0.005)
Female	0.012**	0.012**	0.011**	0.206***	0.200***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Other than English at home	0.009	0.009	0.013*	0.033***	0.011
	(0.007)	(0.008)	(0.008)	(0.007)	(0.007)
Foreign-born	0.054***	0.057***	0.064***	0.006	0.021***
	(0.009)	(0.009)	(0.009)	(0.008)	(0.007)
Individual Schooling Characteristics					
Took LAB	–1.223***	–1.171***	–1.175***	–0.856***	–0.827***
	(0.074)	(0.073)	(0.073)	(0.067)	(0.066)
LAB <= 40th percentile	0.398***	0.394***	0.408***	0.274***	0.249***
	(0.049)	(0.048)	(0.049)	(0.051)	(0.051)
LAB percentile	0.020***	0.019***	0.020***	0.013***	0.013***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Part-time special education	–0.156***	–0.177***	–0.145***	–0.275***	–0.297***
	(0.012)	(0.012)	(0.012)	(0.009)	(0.009)
Same school as last year	0.017*	–0.045***	–0.049***	0.081***	0.055***
	(0.009)	(0.011)	(0.011)	(0.010)	(0.010)
Attendance	0.007***	0.006***	0.005***	0.011***	0.010***
	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)
Constant	–0.619***	–0.530***	–0.319***	–0.285***	–0.237***
	(0.091)	(0.092)	(0.092)	(0.089)	(0.088)
School fixed effects	No	Yes	No	No	Yes
Classroom fixed effects	No	No	Yes	No	No
Number of fixed effects		667	2,892		278
Number of students	70,638	70,638	70,076	55,921	55,921
R^2	0.50	0.52	0.56	0.60	0.63

Notes: Column 3 includes classrooms with at least five students, reducing the number of students by 562 and the number of schools by four.
Robust standard errors in parentheses.
* significant at 10%; ** significant at 5%; *** significant at 1%.
Models include dummy variables to indicate whether free lunch, reduced price lunch, and prior test score are non-missing.

not point to inequities in teachers and class sizes that drive differences in value-added test score gaps.

In sum, once we control for prior test scores and school assignment, racial gaps diminish significantly. Nonetheless, black and Hispanic students continue to perform worse than their white and Asian peers, even after controlling for their performance on tests in the prior year. The academic performance of black and Hispanic students, in other words, is improving more slowly than the performance of their white and Asian peers who attend the same schools, and this difference is not explained by classroom assignment.

There is variation in the size of these racial gaps, however. In some schools, racial gaps are minimized or even reversed. In the next section, we exploit this variation to explore whether there is something we can identify about the characteristics of these "successful" schools that allows them to show more progress between grades for black or Hispanic students compared to white students.

School Characteristics and Within-School Racial Test Score Gaps

If the within-school gaps are not explained by differences across classrooms, do differences in school-level characteristics such as school size play a role? To answer this question, we estimate a set of hierarchical linear models (using random coefficients specification) to examine the relationship between school characteristics and the performance of students by race group.[22] As shown in Table 5, regressions include school variables as well as individual characteristics, including past performance, and the results for the individual-level variables are quite similar to those estimated in the previous models.[23]

Five school-level variables are included. School size (enrollment) is included, following the popular notion that small schools differ from larger schools. Four variables capture the size and character of the teaching workforce: the teacher-pupil ratio, indicating the availability of teachers as a resource; and the percentage of teachers with master's degrees, the share with more than two years in the school, and the percentage who are licensed.

The results are intriguing. Perhaps most interesting is the relationship between school size and performance. Among fifth graders, whites, Hispanics, and Asian students do worse in schools with larger enrollments, but among eighth graders, black students appear hurt by larger enrollments, while students of other racial groups are unaffected. Thus, the "small school" reform movement in high schools might effectively move toward middle and elementary schools, though additional research is warranted. The result for eighth grade black students is particularly noteworthy, given that districts such as Chicago and New York that are most aggressively pursuing small high school reforms are ones with high proportions of black students. Perhaps large districts need to start with younger students and reduce the size of their middle schools, especially ones that are disproportionately black.

[22] Models are estimated to adjust the standard errors for clustering of the students in schools. We also estimated these models in OLS format, with interactions between race and school-level characteristics, with robust standard errors, clustered by school. The signs and magnitudes of the coefficients on school-level characteristics are similar between models, but the precision is greater in the OLS versions. Therefore we report the most conservative versions. The random coefficient models were estimated using SAS proc mixed, unconstrained covariance matrices.

[23] Three schools with fifth grades and one with eighth grade did not have data for teachers and were excluded from the analyses. These schools were small, with fewer than 110 students in either grade.

Table 5. Random coefficients (HLM) models for reading test scores, New York City public school students, 5th grade, 2000–01.

	5th Grade (1)	8th Grade (2)
Prior test score	0.609***	0.622***
	(0.003)	(0.003)
Socioeconomic Characteristics		
Free lunch	–0.106***	–0.148***
	(0.009)	(0.009)
Reduced-price lunch	–0.055***	–0.071***
	(0.012)	(0.012)
Age	–0.028***	–0.079***
	(0.005)	(0.005)
Female	0.013**	0.201***
	(0.005)	(0.005)
Other than English at home	0.011	0.005
	(0.008)	(0.007)
Foreign-born	0.055***	0.023***
	(0.009)	(0.007)
Individual Schooling Characteristics		
Took LAB	–1.184***	–0.835***
	(0.057)	(0.066)
LAB percentile	0.020***	0.013***
	(0.001)	(0.001)
LAB <= 40th percentile	0.398***	0.255***
	(0.045)	(0.055)
Part-time special education	–0.174***	–0.295***
	(0.010)	(0.010)
Same school as last year	–0.034***	0.056***
	(0.009)	(0.010)
Attendance	0.006***	0.010***
	(0.000)	(0.000)
School–Level Characteristics		
White		
Constant	–0.933***	–0.450*
	(0.214)	(0.231)
Enrollment	–0.057*	–0.064
	(0.033)	(0.045)
Teacher/pupil ratio	0.555	–2.085
	(0.944)	(1.394)
% teachers with master's degree	0.005***	0.000
	(0.002)	(0.003)
% teachers with more than 2 years in the school	–0.002**	0.001
	(0.001)	(0.001)
% licensed teachers	0.002	0.005*
	(0.002)	(0.003)
Black		
Constant	–1.000***	–0.575***
	(0.117)	(0.140)
Enrollment	–0.037	–0.058*
	(0.026)	(0.030)
Teacher/pupil ratio	–1.599***	–1.580**
	(0.618)	(0.761)

(continued)

Table 5. *(continued).*

	5th Grade (1)	8th Grade (2)
% teachers with master's degree	0.004***	0.000
	(0.001)	(0.002)
% teachers with more than 2 years in the school	0.000	0.002***
	(0.001)	(0.001)
% licensed teachers	0.002**	0.003**
	(0.001)	(0.001)
Hispanic		
Constant	–0.827***	–0.539***
	(0.113)	(0.135)
Enrollment	–0.073***	–0.034
	(0.021)	(0.028)
Teacher/pupil ratio	–1.201**	–0.649
	(0.576)	(0.703)
% teachers with master's degree	0.003***	0.000
	(0.001)	(0.002)
% teachers with more than 2 years in the school	0.000	0.002**
	(0.001)	(0.001)
% licensed teachers	0.001*	0.002
	(0.001)	(0.001)
Asian		
Constant	–0.806***	–0.386*
	(0.182)	(0.227)
Enrollment	–0.052*	–0.046
	(0.028)	(0.046)
Teacher/pupil ratio	–0.719	–3.109**
	(0.942)	(1.441)
% teachers with master's degree	0.001	0.000
	(0.002)	(0.003)
% teachers with more than 2 years in the school	0.000	0.003**
	(0.001)	(0.001)
% licensed teachers	0.004**	0.003
	(0.002)	(0.003)
Observations	70,532	55,902

Notes: * significant at 10%; ** significant at 5%; *** significant at 1%.
Models include dummy variables to indicate whether free lunch, reduced price lunch, and prior test score are non-missing.

The results for the teacher variables are mixed. At least one of the three variables capturing teacher characteristics is positively related to test scores for minority students, with the percent of teachers with master's degrees and the percent licensed positively related to the test scores of minority fifth graders. Among eighth graders, the percent of teachers with at least two years of experience appears to correlate with higher test scores. Meanwhile, the coefficients on the teacher-pupil variable suggest that black and Hispanic fifth graders and Asian eighth graders perform worse in schools with higher teacher-student ratios.

Caution is warranted in interpreting this last result, however. Previous research has documented that in New York City, the least skilled teachers disproportionately

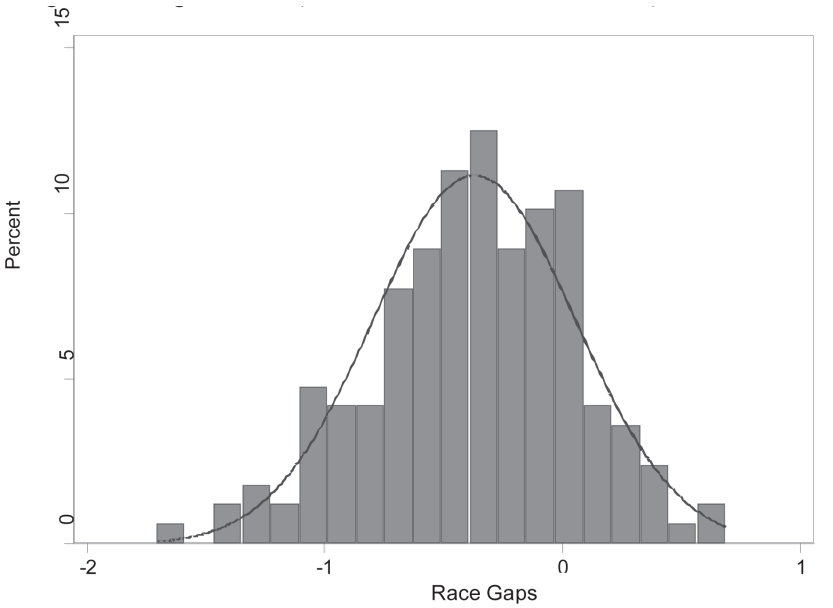

Figure 1. Difference in Black-White Within-School Racial Test Scores (Fixed Effects, No Controls), New York City Public School Students, 5th Grade Reading, 2000–01 (Schools with More Than 5 Students).

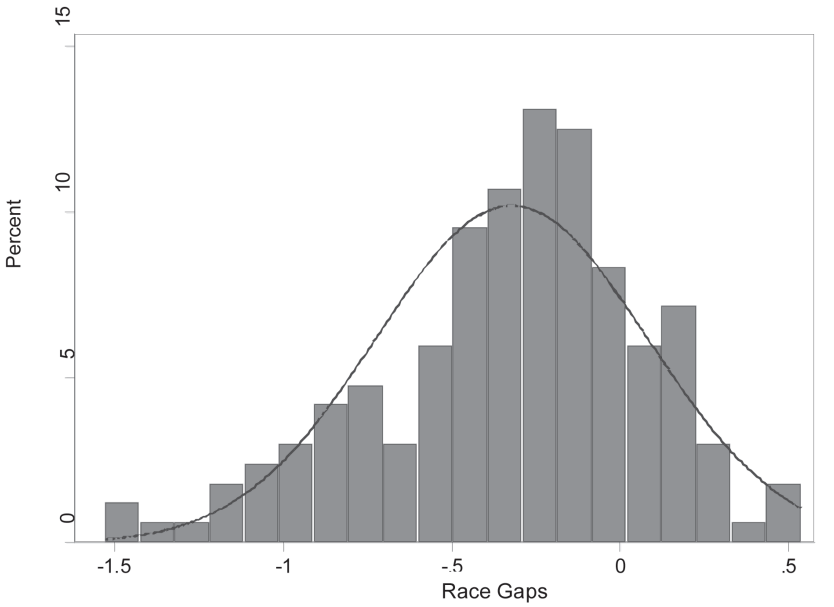

Figure 2. Difference in Hispanic-White Within-School Racial Test Scores (Fixed Effects, No Controls), New York City Public School Students, 5th Grade Reading, 2000–01 (Schools with More Than 5 Students).

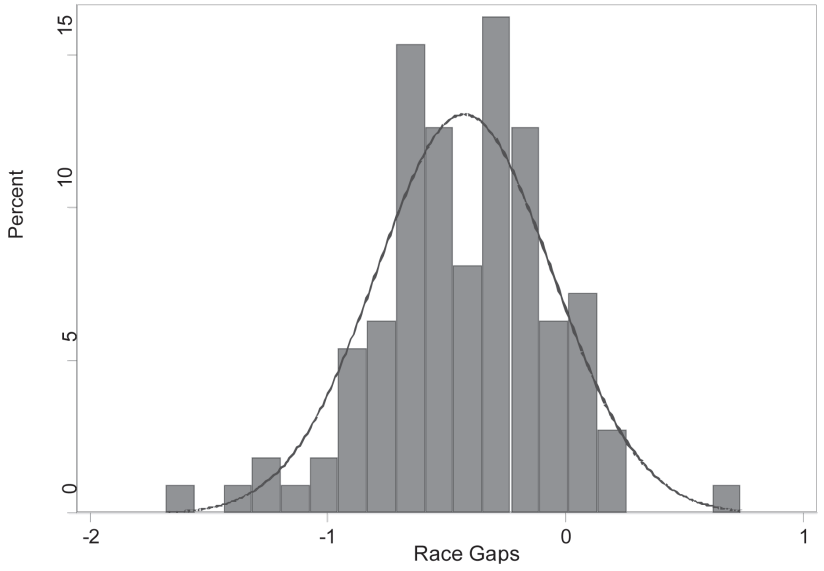

Figure 3. Difference in Black-White Within-School Racial Test Scores (Fixed Effects, No Controls), New York City Public School Students, 8th Grade Reading, 2000–01 (Schools with More Than 5 Students).

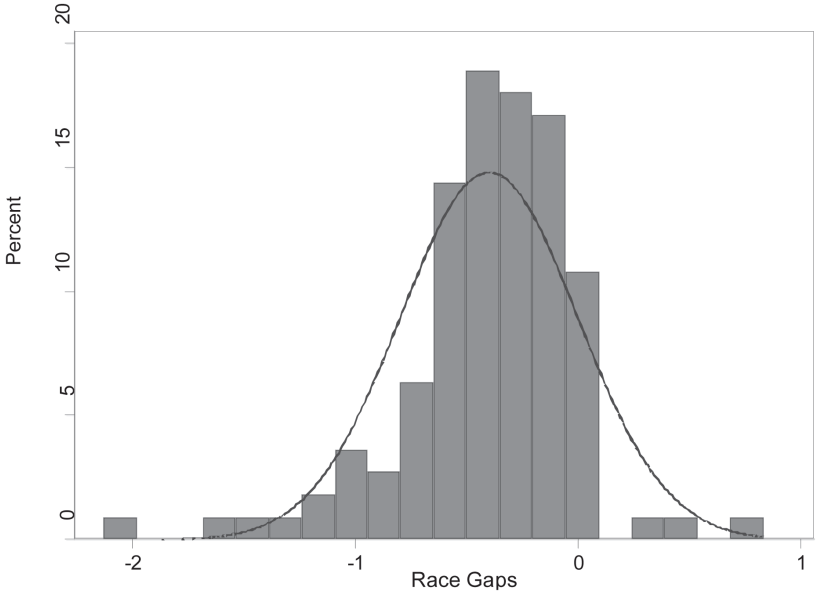

Figure 4. Difference in Hispanic-White Within-School Racial Test Scores (Fixed Effects, No Controls), New York City Public School Students, 8th Grade Reading, 2000–01 (Schools with More Than 5 Students).

teach in the schools with higher percentages of non-white students (Lankford, Loeb, & Wyckoff, 2002). Moreover, previous evidence suggests that there is some trade-off being made—schools with less experienced teachers have more teachers per pupil (Rubenstein et al., 2006). Thus, disentangling the independent effects of these factors is difficult. That said, it is clear that the mix of resources matter and that further work is needed to investigate the possibility that this trade-off between number of teachers and experience level, intentional or otherwise, may not serve all students well.

SUMMARY AND DISCUSSION

Consistent with other work, we find significant disparities between the test scores of white and black students and those of white and Hispanic students in New York City's elementary and middle schools. Much of these test score gaps are explained by racial and ethnic differences in poverty rates and differences in limitations in English proficiency, at least for the Hispanic-white gap. An even more important determinant of the race gap in any year is the performance of students in prior years, which is lower for black and Hispanic students. Put simply, much of the race gap "this year" is explained by the race gap "last year," a factor beyond the influence of this year's schools and teachers.

How much can schools themselves do? Is it possible that we could achieve some reduction in these gaps by redistributing children across schools? As an example, following the sanctions of the federal No Child Left Behind Act, we could close the worst 10% of schools in each of New York City's five boroughs (measured by the regression-adjusted value-added school fixed effects, say). If the students in these schools were placed in schools that performed at the average of their borough, our estimates suggest that the black-white gap would decline by .03 standard deviations and the Hispanic- white gap would decline by about .012 standard deviations in both fifth and eighth grades (to –.103 and –.198 for blacks in fifth and eighth grades and –.032 and –.153 for Hispanics in fifth and eighth grades).[24] While these are small reductions, if these relationships held every year from fifth to twelfth grade, and such changes were made each year, the overall black-white gap would decline by .24 standard deviations by high school graduation.

Of course, as emphasized in this paper, even if we were able to achieve such a re-sorting across schools, significant achievement gaps would still exist within schools. Our estimates suggest that even after controlling for individual characteristics, previous performance, and the school attended, within-school gaps of significant magnitudes remain, especially in 8th grade, where they are between .12 and .17 of a standard deviation for black and Hispanic students, respectively, compared to white students. In New York City, classroom sorting appears to play little direct role in these gaps, but this may be partly because there is so much segregation across schools in the city. We found that only about one-quarter of the 667 elementary schools (168 schools) enrolled five or more black students and five or more white students in the fifth grade. It is possible that racial sorting across

[24] To perform this simulation, we used the fixed effects from regressions in columns 2 and 5, Table 4. These ranged from between –.81 and .60 citywide. We identified the bottom 10% of schools by borough based on the borough distribution of fixed effects. For each student in one of the bottom 10% of the schools, we added to the actual student test score the difference between the average school fixed effect for the borough and the student's actual school fixed effect. We then calculated the new differences between white student scores and black and Hispanic student scores.

classrooms would be more important in school systems in which schools were more racially integrated, where parents might exert more pressure to separate their children or where some other mechanism might come into play in classroom assignments.[25]

How, then, can individual schools do better by their non-white students? One solution is to pursue some of the characteristics that are associated with higher average performance levels among black and Hispanic students. For example, the coefficients in our estimates on school enrollment are large for black and Hispanic students, often near negative .05. Thus, a reduction in enrollment of 100 students might increase average scores by around .05 standard deviations in a year. Note, however, that reducing school size and closing low performing schools and redistributing students to higher performing schools, as per our simulation, are incompatible strategies. Redistributing students will increase—not decrease—the enrollment in higher performing schools.

While the pressure of No Child Left Behind and other current reforms seem intended, at least in part, to reduce the race gaps in performance by pressing for changes within schools, our results do not provide any easy answers for how schools can accomplish these reductions. More research and experimentation is clearly needed to understand how and why scores on standardized tests and improvements in those scores continue to differ significantly for students of different races in the same schools and even the same classrooms. Of course, it is quite possible that other features of schools that we have not measured are important. As an example, particular curricula or whole school reforms, such as Success for All or Accelerated Schools, may be effective. Perhaps, as some have suggested, teachers have systematically different expectations about minority and white students and treat them differently in turn. Or perhaps meaningful reductions ultimately call for strategies that are beyond the scope of schools and school policies.

ACKNOWLEDGMENTS

We thank the Spencer and Rockefeller Foundations for funding that helped support this research. Charles Parekh and Luis Chalico, doctoral candidates at the Wagner School, New York University, provided skilled and dedicated research assistance and Richard Arum, Sewin Chan, Dylan Conger, Caroline Persell, Katherine O'Regan, Beth Weitzman, John Yinger, and four anonymous reviewers contributed valuable advice. We thank seminar participants at New York University's Education Policy Workshop, Columbia University's Social Work School, Syracuse University's Maxwell School, and University of Connecticut's Public Administration program, as well as session participants at the 2004 annual meetings of the Association of Policy Analysis and Management and the American Education Finance Association, for helpful comments. All analyses and conclusions are the responsibility of the authors alone. Note: The authors are listed in reverse alphabetical order and contributed equally.

LEANNA STIEFEL is Professor of Economics at the Wagner Graduate School, New York University.

AMY ELLEN SCHWARTZ is Professor of Public Policy, Education, and Economics at the Institute for Education and Social Policy, Steinhardt and Wagner Schools, New York University.

[25] We thank an anonymous referee for pointing out this possibility.

INGRID GOULD ELLEN is Associate Professor of Policy and Planning at the Wagner Graduate School, New York University.

REFERENCES

Bali, V. A., & Alvarez, M. R. (2003). Schools and educational outcomes: What causes the 'race gap' in student test scores? Social Science Quarterly, 84, 485–507.

Barton, P. E. (2003). Parsing the achievement gap: Baselines for tracking progress. Princeton, NJ: Educational Testing Service.

Bertrand, M., & Mullainathan, S. (2004). Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. American Economic Review, 94, 991–1013.

Boozer, M., & Rouse, C. (2001). Intraschool variation in class size: Patterns and implications. Journal of Urban Economics, 50, 163–189.

Campaign for Fiscal Equity, Inc. v. State of New York, 187 Misc. 2d 1,719 N.Y.S. 2d 475 (N.Y. Sup. Ct. 2001).

Campbell, J. R., Harnbo, C. M., & Mazzeo, J. (2000). NAEP 1999 trends in academic progress: Three decades of student performance. NCES 2000-469. Washington, D.C.: U.S. Department of Education, National Center for Education Statistics.

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. (2005). Who teaches whom? Race and the distribution of novice teachers. Economics of Education Review, 24, 377–392.

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. (2006). The academic achievement gap in grades 3 to 8. NBER Working Paper 12207.

Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mead, A. M., Weinfeld, F. D., & York, R. L. (1966). Equality of educational opportunity. Washington, D.C.: U.S. Department of Health, Education and Welfare.

Conley, D. (2001). Decomposing the black-white wealth gap: The role of parental resources, inheritance, and investment dynamics. Sociological Inquiry, 71, 39–66.

Cook, M. D., & Evans, W. N. (2000). Families or schools? Explaining the convergence in white and black academic performance. Journal of Labor Economics, 18, 729–754.

Ferguson, A. A. (2000). Bad boys: Public schools in the making of black masculinity. Ann Arbor, MI: The University of Michigan Press.

Ferguson, R. F. (1998). Can schools narrow the black-white test score gap? In C. Jencks & M. Phillips (Eds.), The black white test score gap (pp. 318–374). Washington, D.C.: Brookings Institution Press.

Figlio, D. (2005). Names, expectations and the black-white test score gap. NBER Working Paper 11195.

Fordham, S., & Ogbu, J. (1986). Black students' school success: Coping with the burden of "acting white." Urban Review, 18, 176–206.

Fryer, R. G., & Levitt, S. D. (2004). Understanding the black-white test score gap in the first two years of school. Review of Economics and Statistics, 68, 551–60.

Hedges, L. V., & Nowell, A. (1998). Black-white test score convergence since 1965. In C. Jencks and M. Phillips (Eds.), The black white test score gap (pp. 149–181). Washington, D.C.: Brookings Institution Press.

Hedges, L. V., & Nowell, A. (1999). Changes in the black-white gap in achievement test scores. Sociology of Education, 72, 111–135.

Howell, W. G., Wolf, P. J., Campbell, D. E., & Peterson, P. E. (2002). School vouchers and academic performance: Results from three randomized field trials. Journal of Policy Analysis and Management, 21, 191–217.

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

Lankford, H., Loeb, S., & Wyckoff, J. (2002). Teacher sorting and the plight of urban schools: A descriptive analysis. Educational Evaluation and Policy Analysis, 24, 37–62.

Murnane, R. J., Willett, J. B., & Levy, F. (1995). The growing importance of cognitive skills in wage determination. Review of Economics and Statistics, 77, 251–266.

Myers, S. L., Jr., Kim, H., & Mandala, C. (2004). Effect of school poverty on racial gaps in test scores: The case of the Minnesota Basic Standards Test. Journal of Negro Education, 73, 81–98.

Neal, D. A., & Johnson, W. R. (1996). The role of premarket forces in black-white wage differences. Journal of Political Economy, 104, 869–895.

Orfield, G. (2001). Schools more separate: Consequences of a decade of resegregation. The Civil Rights Project. Cambridge, MA: Harvard University.

Reardon, S. F. (June 2003). Sources of educational inequality: The growth of racial/ethnic and socioeconomic test score gaps in kindergarten and first grade. Population Research Institute, Pennsylvania State University. Working paper 03-05R.

Rothstein, R. (2004). Class and schools: Using social, economic, and educational reform to close the black-white achievement gap. New York: Teachers College Press.

Rubenstein, R., Schwartz, A. E., Stiefel, L., & Bel Hadj Amor, H. (forthcoming). From districts to schools: The distribution of resources across schools in big city school districts. Economics of Education Review.

Scholz, J. K., & Levine, K. (2004). U.S. black-white wealth inequality. In K. M. Neckerman (Ed.), Social inequality (pp. 895–930). New York: Russell Sage Foundation Press.

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African-Americans. Journal of Personality and Social Psychology, 69, 797–811.

Todd, P. E., & Wolpin, K. I. (2004). The production of cognitive achievement in children: Home, school and racial test score gaps. Draft April 26.

U.S. Department of Education (2003). National Center for Education Statistics, National Assessment of Educational Progress (NAEP), Washington, D.C.

APPENDIX: CONTENT OF NEW YORK CITY MATH AND ENGLISH LANGUAGE TESTS

In 2000–2001, New York City administered the CTB-Mathematics Test to students in Grades 3, 5, 6, and 7, and the State Mathematics Assessment to students in grades 4 and 8. According to the NYC Department of Education Web site (http://www.nycenet.edu/daa/ctbm/index.html), "The problems and questions on the City CTB-Mathematics and the State Mathematics tests come from the same test series, the TerraNova published by CTB/McGraw-Hill . . . [and these tests] . . . can be used to follow the progress of students in mastering the mathematics standards from grades 3–8 . . . The City CTB-Mathematics Test . . . measures students' attainment of skills such as number relations; computation and numerical estimation; operation concepts; measurement; geometry and spatial sense; data analysis, statistics, and probability; patterns, functions, and algebra; and problem-solving and reasoning."

In English Language Arts, NYC administered the CTB-Reading Test to students in grades 3, 5, 6, and 7 and the State English Language Arts Test to students in grades 4 and 8. According to the NYC Department of Education Web site (http://www.nycenet.edu/daa/ctbr/index.html): "The city CTB-Reading and the state ELA tests . . . come from the same test series, published by CTB/McGraw-Hill [and] . . . can be used to follow the progress of students in mastering the English language arts standards from grade 3 to grade 8. The CTB-Reading Test . . . measures students' attainment of skills such as basic understanding, the analysis of text, evaluating and extending meaning, and identifying reading strategies . . . The fourth and eighth-grade state ELA Tests [measure] . . . skills such as understanding story events, drawing conclusions, making predictions, identifying the main idea. . . ."