NBER WORKING PAPER SERIES

TEACHERS, RACE AND STUDENT ACHIEVEMENT IN A RANDOMIZED EXPERIMENT

Thomas S. Dee

Working Paper 8432
http://www.nber.org/papers/w8432

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
August 2001

I would like to thank Ronald Ehrenberg, William Evans, Alan Krueger, Jens Ludwig and Diane Whitmore and seminar participants at Swarthmore, CUNY and the 2000 NBER Summer Institute for helpful comments. I would also like to thank Ben Keys for excellent research assistance and the Smith Richardson Foundation for financial support. The views expressed herein are those of the author and not necessarily those of the National Bureau of Economic Research.

Teachers, Race and Student Achievement in a Randomized Experiment
Thomas S. Dee
NBER Working Paper No. 8432
August 2001
JEL No. I2

ABSTRACT

Recommendations for the aggressive recruitment of minority teachers are based on hypothesized role-model effects for minority students as well as evidence of racial biases among non-minority teachers. However, prior empirical studies have found little or no association between exposure to an own-race teacher and student achievement. This paper presents new evidence on this question by evaluating the test score data from Tennessee's Project STAR class-size experiment, which randomly matched students and teachers within participating schools. Empirical results based on these data confirm that the racial pairings of students and teachers in this experiment were independently given. Models of student achievement indicate that a one-year assignment to an own-race teacher significantly increased the math and reading achievement of both black and white students by roughly three to four percentile points.

Thomas S. Dee
Department of Economics
Swarthmore College
Swarthmore, PA 19081
and NBER
tdee1@swarthmore.edu

1. INTRODUCTION

The economic literature on the policy determinants of student achievement has focused largely on the possible benefits of new educational resources that reduce class size and improve teacher salaries and training. However, the literature from other social sciences provides provocative evidence that several other contextual factors might also substantively influence the relationship between teachers, students and student achievement. In particular, the conventional wisdom among educators is that minority students are more likely to excel educationally when matched with teachers who share their race or ethnicity. The frequent calls for aggressive recruitment of under-represented minority teachers are typically motivated by the specific claim that such teachers are better-equipped to deal with the special needs of at-risk minority students and that they provide more effective role models (e.g. U.S. Department of Education, 1997; Graham, 1987; Ladson-Billings, 1994; NCTAF, 1996). There is also corresponding evidence that the racial pairings of teachers and students influence how teachers allocate their time in the classroom as well as their expectations and evaluations of students (e.g. Ferguson, 1998; Casteel, 1998; Zimmerman et. al., 1995; Ehrenberg, Goldhaber and Brewer 1995). Yet, relatively few studies have attempted to identify the relationship between exposure to own-race teachers and subsequent levels of student achievement. And what evidence is available suggests that there is actually little association between student achievement and the racial match between teachers and students (Ehrenberg, Goldhaber and Brewer 1995, Ehrenberg and Brewer 1995).

However, the appropriate specification for econometric models of student achievement is a controversial issue. For example, the contentious literature over whether "money matters" in models of student achievement has in part focused on specification issues like functional form and the role of omitted, endogenous or poorly measured variables (e.g. Burtless, 1996). In a recent contribution to this literature, Krueger (1999) examined data from the Project STAR experiment in order to address some of these specification issues as well as the relationship between class size and test scores. Tennessee's Project STAR (Student Teacher Achievement

Ratio) was a large-scale randomized experiment on the achievement benefits of small class sizes. It began in the 1985-86 school year with a group of over 6,000 students from 79 participating schools. The experiment continued through the third grade and ultimately included over 11,000 students.[1] A key feature of the experimental design was the random assignment of both students *and* teachers to small classes, regular-sized classes and regular-sized classes with teacher aides within each school. Project STAR was not designed to evaluate the relationship between own-race teachers and student achievement. Nonetheless, this experiment provides a novel and potentially compelling opportunity to do so since the putatively random pairings of students and teachers should circumvent the non-random and possibly confounding assignments inherent in conventional data on student achievement.[2]

This study presents such test-score evaluations by relying on the recently released Project STAR Public Access Data. However, this study also presents evidence on whether the racial pairing of students and teachers actually satisfies the supposedly randomized experimental design. This is a relevant concern since attrition from the experiment and classroom reassignments ("treatment crossover") could have compromised the integrity of the experimental design (Krueger 1999, Hanushek 1999). The empirical results presented here suggest that these issues are not problematic in this context. In particular, auxiliary regressions indicate that the within-school variation in exposure to an own-race teacher is uncorrelated with other important, student traits (e.g. small class assignment, age, free-lunch status).[3]

---

[1] Not surprisingly with an experiment of this scale and scope, there are a number of important issues with regard to its conduct that could threaten the resulting inferences (most notably, potentially non-random attrition and reassignment). Krueger (1999) addressed these concerns in the context of the class-size results and found that the conventional findings linking small classes with higher levels of student achievement were robust. This study presents similar evidence on whether these violations may be confounding in this context.

[2] Furthermore, because these students are "treated" and observed in their early stages of their formal schooling, the possible effects of own-race teachers may be more easily detectable in these data. Prior studies have evaluated college matriculation or test score gains among older students (Hess and Leal, 1997; Ehrenberg et al., 1995; Ehrenberg and Brewer, 1995).

[3] Furthermore, concerns about the possible biases due to reassignment and attrition are also addressed here in two other ways. One is by evaluating test score equations that include imputed data for students who left the experiment (Krueger 1999). The other is by generating 2SLS estimates of the effect of own-race

In brief, the results of the test score evaluations indicate that exposure to an own-race teacher did generate substantive gains in student achievement for both black and white students. More specifically, these results suggest that a year with an own-race teacher increased math and reading scores by three to four percentile points.[4] Notably, the estimated achievement gains associated with an own-race teacher exist for nearly all groups of students defined by race and gender. However, the estimated effects of assignment to an own-race teacher appear to vary in intriguing ways with respect to other student, teacher and classroom characteristics (e.g., free-lunch status, teacher experience and class size). The remainder of this study is organized as follows. Section 2 briefly summarizes and discusses the prior evidence on the educational implications of the match between the race, gender and ethnicity of students and teachers. Section 3 discusses the Project STAR data in more detail. Section 4 presents initial evaluations and evidence on the quality of the experimental variation in the racial pairings of students and teachers in Project STAR. Section 5 presents a broader set of empirical evidence on the achievement effects of exposure to an own-race teacher and assesses whether these inferences might be confounded by the unobserved dimensions of teacher quality.[5] Section 6 summarizes and speculates briefly about the research and policy implications of these results.

2. TEACHERS AND RACE

Three concerns have dominated discussions of educational policy regarding teachers in recent years. First, a shortage of teachers has been observed since the 1980's. The dearth of

---

teachers where the instrumental variable is a measure of the teacher race a student would have had in the absence of treatment crossover (i.e. the "intent to treat"). The uniformity of the results based on these models suggests that experimental violations are not confounding in this context.

[4] It is difficult to quantify the benefits of these test score gains (Krueger, 1999). However, a comparison with other estimated effects suggest these effects are sizable. More specifically, these estimated effects are often comparable to those associated with a small-class assignment (around four percentile points) and are relatively large in comparison to the observed black-white test gap (around 8 percentile points as well as the test differences between students who do and do not receive free lunches (around 12 percentile points).

[5] The randomized pairing of students and teachers only assures that the *students*' unobserved propensity for achievement is uncorrelated with their teacher's race. However, unobserved teacher quality could still

teachers has been exacerbated in recent years by the retirement of older "baby-boomer" teachers while the "echo" of the baby boom has simultaneously put upward pressure on school enrollments (U.S. Department of Education, 1997). Second, there is evidence that the academic achievement of the college graduates choosing to become teachers has been in decline (e.g., Murnane et. al., 1991). The third, frequently cited concern is a decline in the proportion of teachers who are minorities. Minorities have been historically underrepresented among teachers but, with the pattern of projected retirements and the expected relative growth of minority enrollments, this ostensibly problematic situation is expected to worsen. These stylized facts regarding teachers have motivated recommendations for renewed efforts at recruiting and retaining teachers, particularly those who are racial and ethnic minorities (NCTAF, 1996; U.S. Department of Education, 1997; Graham, 1987; Ladson-Billings, 1994). A maintained assumption underlying these recommendations is that minority teachers would be particularly adept at educating the growing population of minority students. Given that minority students are more likely to be at-risk for academic failure, it is perhaps not surprising that less attention has been paid to whether minority teachers might be less effective teachers of non-minority students.[6]

The prior literature offers at least two general explanations for why the racial pairing of students and teachers might exert an important influence on student achievement. These explanations are not mutually exclusive. One class of explanations involves what could be called "passive" teacher effects. These effects are simply triggered by a teacher's racial presence and not by explicit teacher behaviors. For example, one frequently cited reason for the relevance of a teacher's race is that, by its mere presence, a teacher's racial identity generates a sort of role-model effect that engages student effort, confidence and enthusiasm (e.g., King, 1993; Clewell and Villegas, 1998). For underprivileged black students, the presence of a black teacher may

---

impart biases to the extent it varies systematically with teachers' race. Several types of evidence suggest this sort of omitted variable bias is of limited empirical relevance.

[6] However, since this study presents models of achievement among white and black students separately, it provides evidence on this issue.

encourage them to update their prior beliefs about their educational possibilities. Similarly, students may feel more comfortable and focused in the presence of an own-race teacher regardless of the teacher's actual behavior. While the existence of such role-model effects is frequently assumed in commentaries on educational policy, there is actually little in the nature of direct empirical support (Cizek, 1995). Another possibly relevant sort of passive teacher effect (and one for which there is empirical support) is "stereotype threat" (Steele 1997, 1998). This refers to the possibility that, in situations where students perceive stereotypes might attach (e.g. black students with white teachers), students respond with an apprehension and detachment that retards their educational achievement. An alternative class of explanations for the educational benefits of own-race teachers points to "active" teacher effects: race-specific patterns of behavior among teachers. In particular, it may be that in allocating class time, in interacting with students and in designing class materials, teachers are more oriented towards students who share their racial or ethnic background. There is rather extensive empirical evidence of such racial biases among teachers.[7] For example, prior studies have indicated that black students with white teachers receive less attention, are praised less and scolded more than their white counterparts.[8] Unfortunately, the reduced-form test score models presented here cannot meaningfully distinguish between these two general hypotheses for why own-race teachers might matter. However, a later section speculates about whether the observed heterogeneity in the test-score effects of own-race teachers by student, teacher or classroom traits informs these distinctions.

While there is a seeming consensus on the importance of race in student-teacher interactions, there is surprisingly little clear evidence on the presumed implications for student achievement. For example, in a recent study, Hess and Leal (1997) found that the share of minority faculty in urban school districts is positively correlated with college matriculation rates. But Hess and Leal (1997) correctly noted that partial correlations linking own-race teachers with

[7] Ferguson (1998) discusses this evidence and assesses its implications for the black-white test score gap.

improved student outcomes may be very misleading. More specifically, they suggested that a high proportion of minority faculty can proxy for important but unobserved district-specific determinants of student achievement. Similarly, Ehrenberg and Brewer (1995), in a study examining the classic "Coleman Data" from the 1960's, demonstrated that black teachers are associated with improved test scores gains among black students. However, they also recognized the ambiguity of these partial correlations and actually found that these effects are not robust in models that correct for the simultaneous determination of teacher characteristics.[9] Ehrenberg, Goldhaber and Brewer (1995) reconsidered these questions using more recent data from the National Education Longitudinal Study of 1988 (NELS-88). As in prior educational studies, they found evidence that subjective teacher evaluations of students are often higher when student and teacher race coincide.[10] However, they found almost no evidence that the racial pairings of students and teachers influenced the test score gains among NELS-88 respondents. Since the public-use NELS-88 data do not include geographic identifiers (Dee, Evans and Murray, 1999), they did not attempt to address the endogeneity of teacher characteristics. It is important to note that the lack of a partial correlation between racial pairing and student achievement, such as that reported by Ehrenberg, Goldhaber and Brewer (1995), could also reflect a negative bias imparted by omitted or endogenous regressors. For example, if minority faculty sought out or were more likely to be assigned to at-risk minority students, naïve estimates of their impact on student outcomes would understate the true effects.[11] In the absence of compelling instrumental

---

[8] See Casteel (1998) for recent evidence on teacher-student interactions as well as a brief overview of this literature.

[9] But they also find that the evidence of lower gain scores among white students with black teachers is more robust. However, in general, the quality of their identification strategy for endogenous teacher characteristics may be suspect since it relies on variables that could presumably influence student achievement (e.g. family traits, county or SMSA variables and starting teacher salary).

[10] Actually, they exhaustively study the influence of interactions between race, gender and ethncity. However, the focus of this study is race alone. Almost all Project STAR students and teachers are black or non-Hispanic whites and almost all teachers are female (Krueger, 1999).

[11] This is similar to the potential bias in conventional class-size studies: if at-risk students are more likely to be assigned to small classes, naïve evaluations can understate the impact of the smaller classes. Ehrenberg et. al. (1995) also note that the use of gains in student test scores (as opposed to test score levels) may exacerbate measurement error and lead to attenuated estimates. Krueger (1999) demonstrates that

variables, the uncertain biases inherent in inferences based on observational data represent a seemingly intractable problem for evaluating the educational impact of own-race teachers. For this reason, the putatively random pairings of Project STAR students and teachers provide a unique opportunity to identify the possible links between student and teacher race and student outcomes.

3. TENNESSEE'S PROJECT STAR

In the spring of 1985, the Tennessee Legislature authorized $3 million for the first year of a four-year study of class size that began with kindergarten students that fall (Word et. al., 1990). In the first year of the study, 79 schools (and over 6,000 students) participated. Over the four-year study, roughly 11,600 students participated with about 2200, 1600 and 1200 entering in the first, second and third grades respectively (Krueger, 1999).[12] The participating schools were drawn from around the state and, by legislative mandate, included inner-city and suburban schools from larger metropolitan areas (e.g. Knoxville, Nashville, Memphis and Chattanooga) as well as rural schools and urban schools from smaller towns. Recognizing that schools around the state differed in substantive ways that are inherently difficult to quantify, a within-school experimental design was chosen. This implied that smaller schools were necessarily excluded. Participating schools had to have enough students in a given grade so that three class types, a small class of 15 students and two regular-sized classes of 22 (one with a teacher's aide), could be formed (Mosteller, 1995). Students and teachers within participating schools were randomly matched to three class types. It was originally intended that, once assigned, a student would keep their class type through third grade so that cumulative effects could be identified. However, over the course of the four-year study, this pure experimental design was potentially compromised by both class-type reassignment and student attrition. The next section discusses these issues in

---

econometric models of gain scores can be misleading since they obscure the one-time benefit of an educational determinant.

more detail and presents evidence on the extent to which they might confound these evaluations. For more detailed discussions of the Project STAR experiment in general, see Word et. al. (1990), Mosteller (1995) and Krueger (1999).

The empirical results presented here are based on the Project STAR Public Access Data. Given the very limited number of Hispanic, Asian and American Indian Project STAR participants, these data were edited to include only those observations from black and white non-Hispanic students with black and white non-Hispanic teachers. The implied reductions in sample size were quite modest. For example, among the 6,325 kindergarten students this eliminated only 95 observations.[13] A small number of observations with missing data on key observed characteristics (e.g. age, gender, free lunch status) were also deleted (e.g. five of the kindergarten students). Observations were also omitted simply because test score data were unavailable (largely due to student absenteeism). For example, test score data were available for only about 5900 kindergarten students. The test scores available in these data are the scaled scores from the Stanford Achievement Tests (SAT) in math and reading.[14] Similar to Krueger (1999), the test outcomes modeled here are the percentile ranks based on these scores. More specifically, since the two tests differed across grades, percentile ranks specific to each grade were computed for each subject test. Pooling the individual observations over the four years leads to 23,883 observations on the math test and 23,544 on the reading test (Table 1). Aggregating the data in this way is useful since it increases statistical precision and allows us to consider the effects of cumulative exposure to an own-race teacher. The key observed student characteristics available in these data include a school identifier, class type assignment, student race, gender and age (here represented by a binary indicator for a birth year prior to 1980) as well as an indicator for whether the student received free lunches in their entry year (Table 1). The free-lunch variable is

---

[12] The number of first-grade entrants was fairly high since kindergarten was not required.

[13] Only 30 kindergarten students were neither black nor white, 3 more had missing race/ethncity data. All of kindergarten teachers for whom data were available were identified as white or black. However, teacher race was missng for 62 kindergarten students.

particularly useful since it provides the only available information on the students' socioeconomic background.

The available teacher information includes race, years of experience, education and merit pay status. Notably, teacher gender is not included on the public-use data. In all likelihood, this was intended to preserve confidentiality since almost none of the teachers are male.[15] This does limit the generalizability of this study somewhat since gender may generate some heterogeneity in the race-based interactions between students and teachers. Each student's exposure to an own-race teacher is represented in two ways in this study. The main approach is simply to identify whether each student had an own-race teacher in their current academic year. On average, 94 percent of white students and 45 percent of black students had an own-race teacher (Table 1). However, some models will address the cumulative effects of student exposure to an own-race teacher through the use of an unrestrictive set of binary indicators for one to four years of exposure (with no exposure as the reference).

An important specification issue in this study involves the potential bias in estimated teacher race effects due to the unobserved dimensions of teacher quality.[16] Therefore, the available data on observed teacher traits typically associated with quality (education, experience, merit pay status) can facilitate important robustness checks. Each teacher's education is measured here by a binary indicator for having some type of graduate degree (e.g. M.S., M.A., Ed.S. or Ph.D.) with the reference category including those who only have a bachelor's degree. Roughly 38 percent of students were assigned to a teacher with a graduate degree. Teaching experience is measured in years (a quadratic term is also included in regression models). On average, students were assigned to teachers with nearly 12 years of experience. Another binary indicator identifies teachers recognized by Tennessee's contemporaneous merit pay plan, the Career Ladder

---

[14] The tests were given in late March or early April of each of the four study years.
[15] Krueger (1999, appendix table) reports that none of the kindergarten or first-grade teachers are male while one and three percent of the second-grade and third-grade teachers are.
[16] The empirical relevance of this concern is discussed in more detail in Section 5.

Evaluation System (Dee and Keys 2001). Roughly 90 percent of students were assigned to teachers who participated in Tennessee's career ladder. Whether such merit pay programs can systematically identify and reward good teachers is actually a controversial issue (Murnane and Cohen 1986, Ballou 2001). However, Tennessee's program was considered a relatively sophisticated one since it blended pecuniary and professional rewards and relied on several teaching evaluation instruments (including classroom observation).[17] Nonetheless, its usefulness as a proxy for teacher quality appears to be somewhat limited. Dee and Keys (2001) find that Tennessee's merit pay program had only mixed success in rewarding teachers who raised the average level of classroom achievement.

4. THREATS TO VALIDITY

Project STAR arguably provides a unique and compelling opportunity for making reliable inferences about the determinants of student achievement. However, a number of factors may attenuate the generalizability of inferences based on these data. For example, Krueger (1999) notes the possibility that the class-size effects in Project STAR simply reflect a phenomenon specific to Tennessee. That caveat regarding the "external validity" of this experiment is likely to be much more appropriate in the context of this study since the links between own-race teachers and student achievement may reflect Tennessee-specific cultural factors that to some degree will not generalize to other schools. Other threats to external validity include the lack of male teachers in the study and the lack of small schools in the experiment. But, perhaps even more important are the existing threats to the "internal validity" of the inferences based on this experiment. As noted earlier, like any social experiment, Project STAR had some notable and potentially problematic complications in its execution. For example, because of parental complaints, students in the regular-sized classes were randomly reassigned to

[17] Brandt (1995) praised Tennessee's approach as "perhaps the country's most comprehensive experiment in summative evaluation."

regular-sized classes with and without teacher aides at the beginning of first grade.[18] Assignments to small classes were generally unaffected by this re-randomization. However, roughly 10 percent of students were also moved between small and regular class assignments, largely because of complaints or behavioral problems. Furthermore, attrition from the study was fairly high. This attrition could reflect conventional family mobility, grade repeating or advancing as well as a non-random response to a student's class or teacher assignment.

Krueger (1999) finds that the class-size effects are robust to corrections for these experimental deficiencies.[19] And non-random attrition or reassignment would seem less likely to be problematic in this context than in a high-profile study of class-size effects. When parents chose a school, they presumably had fairly sound prior expectations for the conditional probability that their child would be assigned an own-race teacher. And, since teachers and students would be reassigned in the next academic year, the racial pairings in a given year do not provide a particularly strong incentive for attrition. In contrast, a student's assignment to small or large classes was intended to persist through the third grade. Nonetheless, it is important to consider whether Project STAR's deviations from an ideal experimental design might confound this study's inferences, which link assignments to an own-race teacher with higher test scores. Notably, the likely direction of biases introduced by non-random attrition or class reassignment is uncertain a priori. They could plausibly result in reduced-form results that overstate or understate the true effect of an own-race teacher. For example, if the decision to move a child to an own-race teacher partly reflects unobserved family or parental priors that harm student achievement (e.g. ignorance, intolerance, poor socialization), the results presented here will understate the academic benefits of an own-race teacher. In contrast, to the extent that parents who tend to provide strong academic support for their child are also more likely to seek out an own-race

[18] Initial evaluations based on the kindergarten students indicated that the addition of teacher aides had no impact on student achievement.

teacher, the results presented here will overstate the benefits of such teachers. Clearly, the latter of these possibilities is most relevant given the pattern of results reported here.

These concerns are evaluated here in several ways. One is by presenting results based only on the kindergarten data which were relatively unaffected by these experimental problems (Word et. al., 1990; Krueger, 1999). A second approach is based on considering ad-hoc regressions that evaluate the randomness of the within-school racial pairings by assessing the association between observed student traits and assignment to an own-race teacher. These approaches are combined in the kindergarten results presented in Table 2. The first column of Table 2 reports the results of an OLS regression where the dependent variable is a binary indicator for whether the kindergarten student had an own-race teacher.[20] The regressors in this model are five basic student traits and school fixed effects. These results indicate that black students are substantially less likely to have an own-race teacher. This is to be expected since relatively few black teachers are available among the participating schools. However, if the matching of students and teachers were indeed random, we should find no within-school association between the other observed student traits and exposure to an own-race teacher. The results in Table 2 indicate that this is so. For example, poorer students (that is, those receiving free or reduced-price lunches) are less likely to have an own-race teacher. However, the estimated effect is both small and statistically indistinguishable from zero. Similarly, gender, age and a small-class assignment all exhibit small and statistically weak relationships with an own-race teacher assignment. These four variables are jointly insignificant determinants as well (p-value of 0.35). But the test score results in Table 2 indicate that assignment to an own-race

---

[19] The effects of sample attrition were addressed through the use of imputed test scores. The implications of reassignment were addressed through the use of IV estimates that employ a student's original assignment as an instrumental variable for current class size. Similar robustness checks are presented here.

[20] Throughout this study, standard errors that allow for classroom-specific heteroscedasticity are reported. Classroom identifiers were not reported in the public-use data. However, I effectively identified them by concatenating school, grade, class type assignment, merit pay status and teacher education. The validity of this classroom identifier was confirmed by my ability to replicate the class size distribution reported by Krueger (1999, Table 3). The correspondence of these distributions was exact except for the apparent typo

teacher is associated with higher achievement in both math and reading. Kindergarten students with an own-race teacher had math scores that were a statistically significant 3.6 percentile points higher. Among these kindergarten students, the increase in reading scores (nearly 3 percentile points) is not statistically distinguishable from zero.

The kindergarten results in Table 2 provide important evidence that assignment to own-race teachers appears to have been independently given and that this assignment increased math achievement. However, most of the results in this study will instead exploit the pooled data since they can generate more precision as well as identify cumulative effects. The increase in sample size will also allow us to estimate some models separately by race and gender. The key variable in most models based on the pooled data will again be a binary indicator for whether the student had an own-race teacher in a given year. Other models will exploit a binary measure of whether a student would have had an own-race teacher if they hadn't changed their classroom assignment. This type of "intent to treat" variable provides a plausible instrumental variable for the actual race of a student's teacher. The measure of intended exposure to an own-race teacher was constructed by matching each student to the race of the teacher they would have had if they remained within their entry school and classroom type assignment. However, because of data limitations, this variable is not an exact measure of intent to treat. First, we do not know the exact class they would have attended in the absence of treatment crossover, only the school and classroom type (small, regular, regular-sized with aide). Therefore, this variable takes on the appropriate fractional value in the few cases (15 percent) where students from a given entry school and classroom type could have had a black or white teacher in subsequent years.[21] Second, as Krueger (1999) notes, the classroom assignments that we observe are the actual ones, not necessarily the intended ones (which are not available in the data). However, Krueger (1999)

---

the number of students in 22-student regular-sized classes with aides. Krueger (1999) reports 329 students while I identified 330 students in 15 separate classes.

[21] Also, in the few cases where the class type or grade was not observed for an entry school, students were assigned the mean value for the school and grade or just for the school.

compared intended and actual classroom assignments for kindergarten students from 18 schools and found that they differed for only 0.3 percent of students.

Table 3 presents some critical evidence on whether the within-school variation in these putatively random measures is independent of other student characteristics.[22] The results in the top panel relate the within-school variation in current exposure to an own-race teacher to other student traits. In all of these models, we again see small and statistically insignificant relationships between the current assignment to an own-race teacher and the observed student characteristics. Furthermore, as the p-values indicate, these variables are jointly insignificant as well. The bottom panel relates the *intended* assignment to an own-race teacher to these student traits. Again, these models indicate that, within schools and entry waves, there was no association between assignment to an own-race teacher and other student characteristics. These results provide an important validation of the exogeneity of the experimental assignment to an own-race teacher. However, the subsequent empirical models also examine the relevance of these issues by exploiting an intended assignment to an own-race teacher as an instrumental variable for their actual assignment and by evaluating reduced-form models that include imputed test scores for students who left the experiment or were absent when a test was given.

5. RESULTS

OLS and 2SLS estimates

The basic econometric model presented here relates Y_{isgc}, the grade and subject-specific percentile test rank for student i in school s, grade g and class c, to student, teacher and classroom traits and fixed effects for the grade, entry wave (kindergarten, grades 1 through 3) and the school of entry. More specifically, this model takes the following form:

$$Y_{isgc} = \mathbf{Z}_{isgc}\mathbf{\Pi} + \mathbf{X}_{sgc}\boldsymbol{\beta} + \alpha_g + (\alpha_f \times \alpha_s) + \varepsilon_{isgc}$$

where α_g, α_f and α_s are grade, entry-wave and school-of-entry fixed effects and ε_{isgc} is a mean-zero random error. Because randomization occurred in the school of entry upon the year of entry, a full set of interactions between α_f and α_s is included (Krueger and Whitmore, 2001). However, fixed effects specifications that exclude these interactions return similar results. The matrix, **Z**, includes variables that vary at the individual level such as race, gender, age and free lunch status. In the long form of this model, **Z** also includes several student-specific measures of peer group traits: the percent of classmates on free or reduced-price lunch, the percent who are black, the percent who are female and the percent who attended kindergarten.[23] The matrix, **X**, includes class-specific variables such an assignment to an own-race teacher and assignment to a small class. In the long form of this model, **X** also includes other class-specific measures: years of teaching experience and its square and binary indicators for whether the teacher has a graduate degree and for whether the teacher is in the merit pay program. The impact of introducing these controls is of interest since it will suggest whether the observed effects of teacher race simply reflect the systematic racial differences in the background of teachers. As noted earlier, since there is classroom-specific variation in class size and other unobserved determinants, classroom-specific heteroscedasticity is accommodated in this model through the use of Huber-White standard errors.

Tables 4 and 5 present the OLS and 2SLS estimates of the effect of current exposure to an own-race teacher on math scores in models broken out by the race and gender of the students.[24] The results in Table 4, which are for white males and females, indicate that

---

[22] As is described more formally in the next section, the pooled models include fixed effects for the grade and the interaction of fixed effects for grade of entry and school of entry. Kruger and Whitmore (2001) employ a similar specification.

[23] This last measure is subject to measurement error since we only know kindergarten attendance for those who were in the experiment at that time.

[24] The first-stage effect of an intended assignment to an own-race teacher is not reported. But, not surprisingly, the marginal effect of an intended assignment is quite large and statistically significant. The first-stage coefficient is typically around 0.9 and about 20 times larger than its standard error. Since the model is just identified, the orthogonality of this instrumental variable cannot be tested formally. However, the evidence of randomized assignments in Table 3 suggests its reliability.

assignment to an own-race teacher is associated with a statistically significant 4 to 5 percentile-point increase in math scores. These results are quite robust to 2SLS estimation and to the introduction of the other teacher variables. The results in Table 5 document similarly robust and statistically precise effects in models of the math scores of black students, male and female. Tables 6 and 7 present the results of similar evaluations for reading scores. For white and black males and black females, assignment to an own-race teacher is associated with a statistically significant 3 to 6 percentile point increase in reading scores. For white females, these effects are positive but not statistically distinguishable from zero. The results in Tables 4 through 7 provide rather consistent and robust evidence of the link between exposure to an own-race teacher and increased student achievement. Notably, these estimated effects generally change little (within a fraction of the relevant standard errors) when controls for other teacher and peer traits are introduced and in 2SLS models. Interestingly, these models also suggest that exposure to more experienced teachers often led to statistically significant increases in achievement for white students (but at a decreasing rate). Additionally, there were statistically significant gains in the mathematics scores of black females when assigned to a teacher receiving merit pay or to a teacher with a graduate degree.

Reduced-form results

The robustness of the results in Tables 4 through 7 to 2SLS estimation suggests that treatment crossover does not confound this study's main inferences. However, the other substantive experimental violation of concern involved attrition from the experiment. If high-achieving students who were not assigned to an own-race teacher were more likely to leave the experiment, these results might be highly misleading. Table 8 presents evidence on this question by summarizing the key evaluation results from reduced-form test score models that include imputed data for students who left the experiment. More specifically, test score outcomes were crudely imputed for students who were absent or left the experiment by relying on the prior and

subsequent subject-specific test score rankings available in the data set.[25] The regression models for these data include binary indicators for age, free lunch status and the intent to assign to a small class in addition to the grade and entry school by entry wave fixed effects.[26] The key independent variable is the *intended* assignment to an own-race teacher since the actual assignment is unavailable for students who left the experiment. For comparison purposes, Table 8 also presents the results of this model when applied to only the actual test score data. These results suggest that the test score gains associated with assignment to an own-race teacher are quite robust in the full sample and across the demographic subgroups. More specifically, the right panel of Table 8 uniformly indicates that an own-race teacher increased math and reading achievement by 2 to 4 percentile points among the four subgroups. Though these estimated effects are somewhat smaller than the prior evaluation results, they are generally still statistically significant. The modest reductions in effect sizes observed here may simply reflect an attenuation bias due to some measurement error in the constructed variable representing the intended assignment.

Response heterogeneity

The results presented in Table 9 evaluate potential response heterogeneity by presenting the own-race teacher effects in several different samples of students defined by student, teacher and classroom traits. As points of reference, the first row in Table 9 reports the estimated effects of own-race teachers on math and reading scores from models that include all the available data. These estimates indicate that own-race teachers increase math scores by 3.8 percentage points and reading scores by 3.1 percentage points. The next results in Table 9 indicate that the effects of own-race teachers are concentrated only among those who were assigned to regular-sized classes. For those assigned to small classes, own-race teachers appear to have small and statistically

[25] A missing test score was first imputed by the average of scores from the prior and subsequent years. If still missing, the imputation relied on the most recent prior scores and, then, subsequent scores. Krueger (1999) adopted a similar "last observation carry forward" method.

insignificant effects of the opposite sign. These smaller and imprecise estimates do not appear to be simply an artifact of small samples or insufficient variation. In the small-class models, the hypotheses that the true effects equal the estimates based on the full sample can be rejected. The next results in Table 9 indicate that the achievement gains associated with an own-race teacher are somewhat larger among students with lower socioeconomic status (i.e. those on free lunches). These gains are also isolated among students assigned to relatively inexperienced teachers (11 or fewer years of experience) and are much smaller and statistically insignificant among those assigned to more experienced teachers. However, these effects do not appear to differ substantively among students assigned to teachers with and without graduate degrees. The effects of own-race teachers might also vary across schools with different racial compositions. However, this question cannot be fully addressed with the Project STAR data because the participating schools are highly segregated racially. The median black student attends a school that is 99.79 percent black while the median white student attends a school that is 95.91 percent white. Nonetheless, Table 10 presents separate estimates for the effects of own-race teachers for students in the most densely segregated schools and those in the remaining schools. Though several of these estimates are imprecise, they generally indicate that the achievement gains associated with assignment to an own-race teacher are concentrated in the most segregated schools, particularly for black students.

The results in Tables 9 and 10 suggest that other student, teacher, classroom and school traits have important consequences for the racial interactions between students and teachers. A natural and important question to consider is whether these types of response heterogeneity suggest that own-race teachers matter because of passive teacher effects (e.g. role model effects, stereotype threat) or active ones (e.g. teacher biases) The absence of these effects in small classes is arguably consistent with either hypothesis. For example, more personal interaction with a

[26] Since an actual classroom assignment was not available for the imputed data, the heteroscedasticity in these models is accommodated at the school/grade/classroom type level.

teacher in a small class may obviate the racially driven role-model effects that could occur in larger and more impersonal classes. Alternatively, the racial bias in teacher behaviors may simply be less severe in smaller classes where a teacher's finite resources are less scarce. Similarly, the concentration of these effects among more inexperienced teachers could reflect the importance of a teacher's age for race-based role-model effects (a passive effect) or the role of experience in attenuating unintended racial biases by teachers (an active effect). However, if we were willing to make the strong assumption that teachers with graduate training exhibit less racial bias in the classroom, these results would suggest the benefits of an own-race teacher are driven by passive teacher effects and not by race-specific teacher behavior. The concentration of these effects in the most segregated schools may suggest more convincingly the importance of passive teacher effects. It is not clear there could be such a dramatic shift in active teacher biases across these schools. However, the passively generated harm associated with an other-race teacher could plausibly be concentrated among those students who have had little or no experience with people of the opposite race.

The role of teacher unobservables

Overall, these results indicate that assignment to an own-race teacher was associated with large and statistically significant achievement gains for both black and white students. Furthermore, the randomized pairings of students and teachers that occurred as part of the Project STAR experiment allow us to be unusually certain that these robust associations do not merely reflect the unobserved, *student-level* determinants of educational achievement. However, these estimates do not provide entirely unambiguous evidence on the importance of racial dynamics in the classroom. It is possible that the apparent effects of teachers' race simply reflect unobserved dimensions of teacher quality that happen to vary with a teacher's race. For example, the results for black students are also consistent with the hypothesis that the predominantly black schools managed to attract and retain high-quality black teachers but only low-quality white teachers. However, three types of indirect evidence suggest that the possible biases due to unobserved

teacher quality are of limited relevance. First, the conjecture that these results merely reflect systematic racial differences in unobserved teacher quality is inconsistent with the fact that *both* black and white students appear to have gained from exposure to an own-race teacher. If black teachers were simply better than white teachers on average, we would have instead expected to find that all students gained from exposure to them. Given that both black and white students gained from exposure to own-race teachers, systematic racial differences in teacher quality would imply that the estimated effects of teacher race are overstated for students of one race but understated for students of the other race.[27]

The available data on other teacher traits possibly associated with teacher quality (experience, graduate education, merit pay status) provide a second reason to doubt the confounding influence of unobserved teacher quality. The estimated effects of assignment to an own-race teacher (Tables 4, 5, 6 and 7) were quite robust to the introduction of these teacher traits. Furthermore, these test-score results indicate that some of these controls do reflect teacher quality in that they are at least sometimes associated with statistically significant achievement gains. A related approach would be to suppose that racial differences in observed teacher quality provide an index for the amount of racial differences in unobserved teacher quality.[28] Separate auxiliary regressions for black and white students indicate that assignment to an own-race teacher was not associated with substantive changes in the other teacher traits. Specifically, for white students, assignment to an own-race teacher led to small and statistically insignificant changes in all three observed teacher traits (experience, graduate education, merit pay status). The results were similar for black students except that own-race teachers had on average roughly 4 more

[27] The high amount of racial segregation across schools implies a caveat to this observation. It may be that the white teachers that black students typically face are of relatively low quality while, simultaneously, the black teachers that white students typically face are also of relatively low quality.

[28] See Murphy and Topel (1990) and Altonjii, Elder and Taber (2000) for applications of this approach to bounding omitted variable biases.

years of teaching experience.[29] The limited amount of "selection on observables" suggests that there may not be large and systematic racial differences in unobserved teacher quality that could confound the interpretation of this study's results.

Third, the types of response heterogeneity documented in the previous section are entirely consistent with the existence of active or passive effects associated with teacher race but are much less easily reconciled with the confounding presence of unobserved teacher quality. In particular, if this study's results merely reflected racial patterns in unobserved teacher quality, it is not clear why this race-specific teacher quality should matter in larger classes but not in smaller classes. Similarly, the fact that these effects appear to be concentrated among relatively inexperienced teachers for students of *both* races cannot easily be attributed solely to the existence of race-specific but unobserved teacher quality.

Cumulative effects

The prior empirical models presented here have assumed a constant effect associated with a year's exposure to an own-race teacher regardless of a student's cumulative exposure. The results presented in Table 11 are based on alternative models that identify the effects of years of cumulative exposure to an own-race teacher in a relatively unrestrictive manner by using a set of four binary indicators. In general, these results suggest each year of additional exposure generates roughly equal increases of 2 to 4 percentile points in math and reading achievement. The hypotheses that the coefficients on these indicators are equal can be easily rejected. These results imply that exposure to an own-race teacher does simply confer a fixed, one-time gain but rather can have additive effects on student achievement.

[29] However, for black students, teaching experience appears to be a poor indicator of teacher quality (Tables 5 and 7).

6. CONCLUSIONS

Frequent recommendations for the aggressive recruitment of minority teachers have been motivated by the putative educational benefits for minority students. However, the available evidence that own-race teachers actually improve student achievement has at best been limited and qualified. As recent studies have recognized, it is difficult to make reliable inferences about this relationship given the pervasive specification problems associated with standard observational data on educational outcomes and the absence of compelling natural experiments. It was suggested here that the Project STAR class-size experiment presents a unique opportunity to examine the putative educational benefits of own-race teachers since it generated ostensibly random pairings of the students and teachers under study. This study presented such evaluations and found consistent evidence that there are rather large educational benefits for both black and white students from assignment to an own-race teacher in these early grades.

These results clearly provide novel support for the conventional assumption that recruiting minority teachers can generate important achievement gains among minority students. These results also suggest that one of the real and typically overlooked costs of such efforts may be a meaningful reduction in the educational achievement of non-minority students. However, there are several important caveats appropriate to considering the broader policy implications of these results. For example, these results cannot, of course, speak directly to whether these effects exist in regions outside of the Tennessee schools under study. These results also do not address the effects of own-race teachers on important long-term student outcomes such as educational attainment. But, perhaps most importantly, this study cannot provide meaningful evidence on the exact mechanisms by which own-race teachers might actually influence student achievement (ie. the varying types of passive and active teacher effects). This gap in our knowledge is particularly noteworthy since the results presented here could be narrowly construed to suggest that an increased racial segregation of teachers and students should be promoted to improve the overall levels of educational achievement. Such a recommendation could be criticized not only on

normative grounds but also because it ignores the possibility of more balanced policies informed by an improved understanding of why the racial interactions between students and teachers matter for student outcomes. A more appropriate interpretation of this study's results is that it underscores the sizable educational relevance of the racial dynamic between students and teachers as well as the need for a better understanding of what really drives this phenomenon. The results presented here offered some provocative hints that these racial interactions involve complex structural effects which interact in important ways with other student, teacher, classroom and school traits.

REFERENCES

Altonji, Joseph G., Todd E. Elder and Christopher R. Taber. "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools," National Bureau of Economic Research Working Paper 7831, Cambridge, MA, August 2000.

Ballou, D. (2001). "Pay for Performance in Public and Private Schools." Economics of Education Review **20**(1): 51-61.

Burtless, G. *Does money matter?: the effect of school resources on student achievement and adult success.* Gary Burtless, ed. Brookings Institution Press, Washington, DC: 1996

Casteel, C.A. "Teacher-Student Interactions and Race in Integrated Classrooms," Journal of Educational Research 92 (1998): 115-120.

Cizek, G.J. "On the Limited Presence of African-American Teachers: An Assessment of Research, Synthesis and Policy Implications," Review of Educational Research 65 (1995): 78-92.

Clewell, B.C. and A.M. Villegas. "Introduction" Education and Urban Society 31 (1998): 3-17.

Dee, T.S. and B. Keys. "Does Merit Pay Reward Good Teachers? Evidence From a Randomized Experiment," unpublished manuscript, Department of Economics, Swarthmore College, August 2001.

Dee, T.S., W.N. Evans and S. Murray. "Data Watch: Research Data in the Economics of Education," Journal of Economic Perspectives, 13(3), Summer 1999, pages 205-216.

Ehrenberg, R.G., D.D. Goldhaber and D.J. Brewer. "Do Teachers' Race, Gender and Ethnicity Matter? Evidence From the National Educational Longitudinal Study of 1988?" Industrial and Labor Relations Review 48 (1995): 547-561.

Ehrenberg, R.G. and D.J. Brewer. "Did Teachers' Verbal Ability and Race Matter in the 1960's? Coleman Revisted" Economics of Education Review 14 (1995): 1-21.

Ferguson, Ronald F. "Teachers' Perceptions and Expectations and the Black-White Test Score Gap," in *The Black-White Test Score Gap*, C. Jencks and M. Phillips, editors, Brookings Institution Press, Washington DC, 1998.

Graham, P.A. "Black Teachers: A Drastically Scarce Resource," Phi Delta Kappan (1987): 598-605.

Hanushek, E.A. "Some Findings From an Independent Investigation of the Tennessee STAR Experiment and From Other Investigations of Class Size Effects," Educational Evaluation and Policy Analysis 21 (1999): 143-164.

Hess, F.M. and D.L. Leal. "Minority Teachers, Minority Students, and College Matriculation: A New Look at the Role-Modeling Hypothesis," Policy Studies Journal 25 (1997): 235-248.

King, S.H. "The Limited Presence of African-American Teachers," Review of Educational Research 63 (1993): 115-149.

Krueger, A.B. "Experimental Estimates of Education Production Functions," Quarterly Journal of Economics (1999): 497-532.

Krueger, A.B. and D.M. Whitmore "The Effect of Attending a Small Class in the Early Grades on College-Test Taking and Middle School Test Results: Evidence from Project STAR," Economic Journal 111 (January 2001), 1-28.

Ladson-Billings, G. *The dreamkeepers : successful teachers of African American children*. Jossey-Bass Publishers, San Francisco, CA: 1994.

Mosteller, F. "The Tennessee Study of Class Size in Early Grades," Critical Issues for Children and Youth 5 (1995): 113-127.

Murnane, R. J. and D. K. Cohen (1986). "Merit Pay and the Evaluation Problem: Why Most Merit Pay Plans Fail and a Few Survive." Harvard Educational Review **56**(1): 1-17.

Murnane, R.J., Singer, J.D., Willett, J.B., Kemple, J.J. and Olson, R.J. *Who Will Teach? Policies That Matter*. Harvard University Press, Cambridge, MA: 1991.

Murphy, Kevin M. and Robert H. Topel. "Efficiency Wages Reconsidered: Theory and Evidence," in *Advances in the Theory and Measurement of Unemployment*, Yoram Weiss and Gideon Fishelson, editors, St. Martins Press, New York, 1990, pages 204-240.

National Commission on Teaching and America's Future (NCTAF). What Matters Most: Teaching for America's Future. New York, NY: 1996.

Steele, C.M. "A threat in the air - How stereotypes shape intellectual identity and performance," American Psychologist 52: (6) 613-629 June 1997

Steele, C.M. "Stereotyping and its threats are real, " American Psychologist 53: (6) 680-681 June 1998

U.S. Department of Education, National Center for Education Statistics. *America's Teachers: Profile of a Profession, 1993-94*. NCES 97-460, by Robin R. Henke, Susan P. Choy, Xianglei Chen Sonya Geis, Martha Naomi Alt, Stephen P. Broughman, Project Officer. Washington, DC: 1997.

Zimmermann, R.S., E.L Khoury, W.A. Vega, A.G. Gil, G.J.Warheit. "Teacher and Parent Perceptions of Behavior Problems Among a Sample of African-American, Hispanic and Non-Hispanic White Students," American Journal of Community Psychology 23 (1995): 181-197.

Word, E., J. Johnston, H.P. Bain, B.D. Fulton, J.B. Zaharias, C.M. Achilles, M.N. Lintz, J.Folger, C. Breda. The State of Tennessee's Student/Teacher Achievement Ratio (STAR) Project Final Summary Report 1985-1990.

Table 1 – Student Traits by Student and Teacher Race, Pooled Project STAR K-3 Data

	White Students		Black Students	
Student Variable	Teacher of Own Race	Teacher of Other Race	Teacher of Own Race	Teacher of Other Race
Mathematics Score	56.7	50.0	40.7	37.6
	(27.8)	(28.5)	(27.6)	(26.7)
Reading Score	56.9	51.5	39.9	37.6
	(28.0)	(28.9)	(26.1)	(26.5)
Small Class	.31	.29	.30	.27
	(.46)	(.45)	(.46)	(.44)
Female Student	.48	.48	.50	.49
	(.50)	(.50)	(.50)	(.50)
Born Before 1980	.36	.38	.35	.37
	(.48)	(.49)	(.48)	(.48)
Free Lunch	.33	.33	.82	.81
	(.47)	(.47)	(.39)	(.40)
Sample Size	15,033	922	3,542	4,386

Notes - Standard deviations are reported in parentheses. Observed student characteristics are defined for the 23,883 student observations with math scores. The reading test is defined for 23,544 student observations.

Table 2 – Own-Race Teachers and Test Scores, Kindergarten Students

	Dependent Variable		
Variable	Teacher of Own Race	Mathematics Score	Reading Score
Teacher of Own Race	--	3.6†	2.9
		(1.7)	(2.1)
Small Class	-0.010	4.8‡	5.7‡
	(0.028)	(1.2)	(1.2)
Black Student	-0.614‡	-8.4‡	-6.5‡
	(0.086)	(1.8)	(2.0)
Female Student	-0.00003	4.4‡	5.5‡
	(0.008)	(0.7)	(0.7)
Born Before 1980	-0.012	5.8‡	3.7‡
	(0.007)	(0.7)	(0.7)
Free Lunch	-0.012	-12.8‡	-14.0‡
	(0.010)	(0.8)	(0.9)
Sample Size	5783	5780	5699
R^2	.5375	.2795	.3131

Notes – Robust standard errors are reported in parentheses. All models include school fixed effects. Roughly 77 percent of the kindergarten students have an own-race teacher.

* Statistically significant at 10-percent level

† Statistically significant at 5-percent level

‡ Statistically significant at 1-percent level

Table 3 – Within-School Association between Student Traits and Assignment to Own-Race Teachers, By Student Race and Gender

	Sample			
Variable	White Males	Black Males	White Females	Black Females
	Dependent Variable: Own-Race Teacher			
Small Class	.013	.057	.004	.043
	(.013)	(.045)	(.013)	(.048)
Born Before 1980	-.004	.002	-.003	.001
	(.005)	(.015)	(.005)	(.017)
Free Lunch	-.005	-.028	-.001	.006
	(.004)	(.021)	(.005)	(.021)
R^2	.2491	.1952	.2428	.1910
p-value	0.23	0.37	0.94	0.84
	Dependent Variable: Intended Own-Race Teacher			
Small Class	.014	.049	.007	.034
	(.010)	(.033)	(.010)	(.033)
Born Before 1980	.002	-.004	-.002	.007
	(.004)	(.012)	(.004)	(.013)
Free Lunch	.001	-.011	-.001	.019
	(.004)	(.016)	(.004)	(.015)
R^2	.3173	.2698	.3103	.2577
p-value	0.44	0.42	0.88	0.44
Sample Size	8,328	4,024	7,665	3,939

Notes – Robust standard errors are reported in parentheses. All models include grade fixed effects and the interactions of fixed effects for the entry school and entry wave. The p-value refers to an F-test of the joint significance of the three variables.

* Statistically significant at 10-percent level

† Statistically significant at 5-percent level

‡ Statistically significant at 1-percent level

Table 4 - Estimated Effects of an Own-Race Teacher
on the Mathematics Scores of White Students by Gender

	White Males				White Females			
	OLS	2SLS	OLS	2SLS	OLS	2SLS	OLS	2SLS
Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Teacher of Own Race	4.6‡	5.0†	4.4†	4.7*	4.5†	5.1†	4.0†	4.5*
	(1.7)	(2.4)	(1.8)	(2.5)	(1.9)	(2.6)	(1.9)	(2.7)
Small Class	4.5‡	4.4‡	4.3‡	4.3‡	3.1‡	3.0‡	2.9‡	2.9‡
	(0.9)	(0.9)	(0.9)	(0.9)	(0.9)	(0.9)	(0.9)	(0.9)
Born Before 1980	-0.3	-0.4	-0.4	-0.4	-0.4	-0.4	-0.5	-0.5
	(0.7)	(0.7)	(0.7)	(0.7)	(0.7)	(0.7)	(0.7)	(0.7)
Free Lunch	-10.0‡	-10.0‡	-10.0‡	-10.0‡	-12.9‡	-12.9‡	-13.0‡	-13.0‡
	(0.7)	(0.7)	(0.7)	(0.7)	(0.7)	(0.7)	(0.7)	(0.7)
Teacher Experience	--	--	.0.32†	.0.32†	--	--	0.10	0.11
			(0.16)	(0.16)			(.0.15)	(.0.15)
Teacher Experience Squared	--	--	-.011†	-.011†	--	--	-.004	-.004
			(.005)	(.005)			(.005)	(.005)
Graduate Degree	--	--	-0.7	-0.7	--	--	1.5*	1.5*
			(0.9)	(0.9)			(0.9)	(0.9)
Merit Pay	--	--	2.5	2.5	--	--	2.2	2.2
			(1.6)	(1.6)			(1.8)	(1.8)
Percent of Classmates on Free Lunch	--	--	1.4	1.5	--	--	-5.4	-5.4
			(3.7)	(3.7)			(3.8)	(3.8)
Percent of Classmates in Kindergarten	--	--	0.9	0.9	--	--	1.1	1.1
			(3.2)	(3.3)			(3.4)	(3.4)
Percent of Classmates Black	--	--	-6.0	-6.0	--	--	-14.9†	-14.8†
			(6.6)	(6.6)			(7.1)	(7.0)
Percent of Classmates Female	--	--	0.2	0.2	--	--	-0.6	-0.7
			(4.5)	(4.6)			(4.7)	(4.7)
R^2	.1729	.1729	.1746	.1746	.1867	.1867	.1896	.1896
Sample Size	8,310	8,310	8,310	8,310	7,645	7,645	7,645	7,645

Notes – Robust standard errors are reported in parentheses. All models include grade fixed effects and the interactions of fixed effects for the entry school and entry wave.

* Statistically significant at 10-percent level

† Statistically significant at 5-percent level

‡ Statistically significant at 1-percent level

Table 5 - Estimated Effects of an Own-Race Teacher
on the Mathematics Scores of Black Students by Gender

	Black Males				Black Females			
	OLS	2SLS	OLS	2SLS	OLS	2SLS	OLS	2SLS
Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Teacher of Own Race	3.2†	4.0†	3.2†	3.9*	3.2†	5.3†	3.7†	5.3†
	(1.5)	(2.0)	(1.6)	(2.1)	(1.5)	(2.2)	(1.5)	(2.3)
Small Class	7.2‡	7.1‡	4.9‡	4.9‡	6.3‡	6.2‡	3.6†	3.6‡
	(1.5)	(1.5)	(1.7)	(1.7)	(1.5)	(1.5)	(1.6)	(1.7)
Born Before 1980	-.11	-.11	-.09	-.08	0.8	0.8	0.9	0.9
	(0.9)	(0.9)	(0.9)	(0.9)	(1.0)	(1.0)	(1.0)	(1.0)
Free Lunch	-8.4‡	-8.3‡	-8.4‡	-8.3‡	-7.6‡	-7.6‡	-7.3‡	-7.3‡
	(1.2)	(1.2)	(1.2)	(1.2)	(1.1)	(1.1)	(1.1)	(1.1)
Teacher Experience	--	--	-.17	-.16	--	--	.16	.15
			(.26)	(.26)			(.26)	(.26)
Teacher Experience Squared	--	--	.006	.006	--	--	-.004	-.004
			(.008)	(.008)			(.007)	(.007)
Graduate Degree	--	--	1.3	1.4	--	--	3.1†	3.4†
			(1.6)	(1.6)			(1.6)	(1.6)
Merit Pay	--	--	2.1	2.1	--	--	4.1†	4.2†
			(2.1)	(2.1)			(2.1)	(2.1)
Percent of Classmates on Free Lunch	--	--	7.6	7.7	--	--	3.3	3.6
			(7.0)	(7.0)			(6.8)	(6.9)
Percent of Classmates in Kindergarten	--	--	16.2‡	16.3‡	--	--	19.8‡	19.7‡
			(4.7)	(4.7)			(4.8)	(4.9)
Percent of Classmates Black	--	--	-3.3	-3.3	--	--	-7.5	-8.0
			(8.8)	(8.8)			(8.3)	(8.3)
Percent of Classmates Female	--	--	3.9	4.0	--	--	8.7	9.3
			(8.2)	(8.2)			(7.3)	(7.4)
R^2	.1945	.1943	.2011	.2010	.2238	.2227	.2376	.2369
Sample Size	4,005	4,005	4,005	4,005	3,923	3,923	3,923	3,923

Notes – Robust standard errors are reported in parentheses. All models include grade fixed effects and the interactions of fixed effects for the entry school and entry wave.

* Statistically significant at 10-percent level

† Statistically significant at 5-percent level

‡ Statistically significant at 1-percent level

Table 6 - Estimated Effects of an Own-Race Teacher
on the Reading Scores of White Students by Gender

	White Males				White Females			
	OLS	2SLS	OLS	2SLS	OLS	2SLS	OLS	2SLS
Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Teacher of Own Race	4.1†	4.6†	3.9†	4.2*	1.6	1.1	1.2	0.6
	(1.6)	(2.2)	(1.6)	(2.2)	(1.7)	(2.2)	(1.8)	(2.2)
Small Class	3.6‡	3.6‡	3.4‡	3.4‡	3.3‡	3.3‡	2.9‡	2.9‡
	(0.8)	(0.8)	(0.9)	(0.9)	(0.8)	(0.8)	(0.9)	(0.9)
Born Before 1980	-2.7‡	-2.7‡	-2.7‡	-2.7‡	-2.4‡	-2.4‡	-2.4‡	-2.5‡
	(0.7)	(0.7)	(0.7)	(0.7)	(0.7)	(0.7)	(0.7)	(0.7)
Free Lunch	-11.6‡	-11.6‡	-11.6‡	-11.6‡	-13.3‡	-13.3‡	-13.4‡	-13.4‡
	(0.7)	(0.7)	(0.7)	(0.7)	(0.7)	(0.7)	(0.7)	(0.7)
Teacher Experience	--	--	.49‡	.49‡	--	--	.39‡	.39‡
			(.15)	(.15)			(.13)	(.13)
Teacher Experience Squared	--	--	-.014‡	-.014‡	--	--	-.010‡	-.010‡
			(.004)	(.004)			(.004)	(.004)
Graduate Degree	--	--	-1.0	-1.0	--	--	0.7	0.7
			(0.8)	(0.8)			(0.8)	(0.8)
Merit Pay	--	--	2.1	2.1	--	--	3.0†	3.0†
			(1.6)	(1.6)			(1.5)	(1.5)
Percent of Classmates on Free Lunch	--	--	-1.9	-1.8	--	--	-10.8†	-10.8†
			(3.7)	(3.7)			(3.8)	(3.8)
Percent of Classmates in Kindergarten	--	--	1.5	1.5	--	--	2.7	2.6
			(3.2)	(3.2)			(3.3)	(3.3)
Percent of Classmates Black	--	--	-12.9†	-12.8†	--	--	-9.2	-9.4
			(6.2)	(6.1)			(6.9)	(6.8)
Percent of Classmates Female	--	--	2.0	2.0	--	--	3.0	3.1
			(4.5)	(4.5)			(4.3)	(4.3)
R^2	.1887	.1887	.1918	.1918	.2086	.2088	.2139	.2139
Sample Size	8,154	8,154	8,154	8,154	7,518	7,518	7,518	7,518

Notes – Robust standard errors are reported in parentheses. All models include grade fixed effects and the interactions of fixed effects for the entry school and entry wave.

* Statistically significant at 10-percent level

† Statistically significant at 5-percent level

‡ Statistically significant at 1-percent level

Table 7 - Estimated Effects of an Own-Race Teacher
on the Reading Scores of Black Students by Gender

	Black Males				Black Females			
	OLS	2SLS	OLS	2SLS	OLS	2SLS	OLS	2SLS
Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Teacher of Own Race	3.3†	4.7†	3.0†	4.5†	3.7‡	6.0‡	3.7†	5.7†
	(1.4)	(2.0)	(1.5)	(2.1)	(1.4)	(2.2)	(1.4)	(2.2)
Small Class	6.6‡	6.5‡	5.6‡	5.6‡	6.8‡	6.7‡	5.0‡	4.9‡
	(1.4)	(1.5)	(1.6)	(1.6)	(1.4)	(1.5)	(1.6)	(1.6)
Born Before 1980	-3.2‡	-3.3‡	-3.3‡	-3.3‡	-1.3	-1.3	-1.3	-1.3
	(0.9)	(0.9)	(0.9)	(0.9)	(0.9)	(0.9)	(0.9)	(0.9)
Free Lunch	-6.8‡	-6.7‡	-6.7‡	-6.7‡	-10.7‡	-10.7‡	-10.3‡	-10.3‡
	(1.2)	(1.2)	(1.1)	(1.2)	(1.3)	(1.3)	(1.3)	(1.3)
Teacher Experience	--	--	0.13	0.13	--	--	0.44*	0.44*
			(.26)	(.26)			(.24)	(.24)
Teacher Experience Squared	--	--	-.0003	-.001	--	--	-.009	-.010
			(.007)	(.007)			(.007)	(.007)
Graduate Degree	--	--	1.0	1.2	--	--	1.8	2.1
			(1.5)	(1.5)			(1.5)	(1.5)
Merit Pay	--	--	0.5	0.6	--	--	1.1	1.3
			(2.0)	(2.0)			(2.1)	(2.2)
Percent of Classmates on Free Lunch	--	--	3.9	4.1	--	--	2.6	3.0
			(6.6)	(6.6)			(6.7)	(6.7)
Percent of Classmates in Kindergarten	--	--	6.2	6.3	--	--	11.9‡	11.8†
			(4.4)	(4.4)			(4.7)	(4.8)
Percent of Classmates Black	--	--	-7.5	-7.7	--	--	-12.8*	-13.5*
			(8.3)	(8.4)			(7.4)	(7.4)
Percent of Classmates Female	--	--	8.1	8.4	--	--	20.5‡	21.3‡
			(7.6)	(7.6)			(7.2)	(7.1)
R^2	.1871	.1865	.1911	.1905	.2151	.2136	.2291	.2280
Sample Size	3,972	3,972	3,972	3,972	3,900	3,900	3,900	3,900

Notes – Robust standard errors are reported in parentheses. All models include grade fixed effects and the interactions of fixed effects for the entry school and entry wave.

* Statistically significant at 10-percent level

† Statistically significant at 5-percent level

‡ Statistically significant at 1-percent level

Table 8 - Estimated Effects of an Intended Own-Race Teacher, Actual and Imputed Test Scores

	Actual Test Scores				Actual and Imputed Test Scores			
Sample	Math Score	Sample Size	Reading Score	Sample Size	Math Score	Sample Size	Reading Score	Sample Size
Full Sample	4.2‡ (1.2)	23,883	3.7‡ (1.2)	23,544	3.0‡ (0.9)	34,317	2.4‡ (0.8)	33,978
White Male	4.4† (2.2)	8,310	4.0* (2.1)	8,154	2.8* (1.5)	11,679	2.8† (1.3)	11,535
White Female	4.4† (2.2)	7,645	1.0 (1.9)	7,518	4.0‡ (1.5)	10,506	1.8 (1.3)	10,379
Black Male	3.6* (1.9)	4,005	4.2† (2.0)	3,972	2.7† (1.3)	6,270	3.1† (1.3)	6,219
Black Female	4.4† (1.9)	3,923	4.9‡ (2.0)	3,900	2.4* (1.3)	5,862	2.3 (1.4)	5,845

Notes – Robust standard errors are reported in parentheses. All models include binary indicators for race, gender, age, free lunch status, intended small class assignment, grade fixed effects and the interactions of fixed effects for the entry school and entry wave.

* Statistically significant at 10-percent level

† Statistically significant at 5-percent level

‡ Statistically significant at 1-percent level

Table 9 – Estimated Effects of an Own-Race Teacher
On Mathematics and Reading Scores, By Sample

Sample	Math Score	Sample Size	Reading Score	Sample Size
Full Sample	3.8‡ (1.0)	23,883	3.1‡ (0.9)	23,544
Regular-Sized Class	3.9‡ (1.1)	16,699	3.6‡ (1.0)	16,437
Small Class	-0.2 (1.6)	7,184	-1.6 (1.6)	7,107
No Free Lunch	2.8† (1.2)	12,214	2.1* (1.2)	12,074
Free Lunch	4.4‡ (1.2)	11,669	3.6‡ (1.1)	11,470
Inexperienced Teachers	6.9‡ (1.3)	12,363	5.4‡ (1.1)	12,227
Experienced Teachers	0.9 (1.4)	11,520	1.0 (1.4)	11,317
Graduate Degree	3.4† (1.7)	8,991	3.3† (1.6)	8,855
No Graduate Degree	3.8‡ (1.2)	14,892	2.6† (1.1)	14,689

Notes – Robust standard errors are reported in parentheses. All models include grade fixed effects and the interactions of fixed effects for the entry school and entry wave. Each model also includes the controls for student's race, age, gender, class assignment, teacher experience and its square, graduate degree, merit pay status and classroom peers when variation in the variable exists for the given sample. "Inexperienced" teachers are defined as those with 11 or fewer years of experience; experienced teachers have more than 11 years of experience.

* Statistically significant at 10-percent level

† Statistically significant at 5-percent level

‡ Statistically significant at 1-percent level

Table 10 – Estimated Effects of an Own-Race Teacher On Mathematics and Reading Scores, By Student Race and School Racial Composition

School Trait	Math Score	Sample Size	Reading Score	Sample Size
White Students				
Percent White Greater than 95.91 Percent	6.7 (4.1)	8,192	5.0* (2.7)	7,953
Percent White Less than 95.91 Percent	3.2* (1.7)	7,763	1.4 (1.5)	7,719
Black Students				
Percent Black Greater than 99.79 Percent	6.4‡ (2.2)	3,967	6.1‡ (2.0)	3,953
Percent Black Less than 99.79 Percent	0.03 (1.7)	3,961	0.6 (1.5)	3,919

Notes – Robust standard errors are reported in parentheses. All models include grade fixed effects and the interactions of fixed effects for the entry school and entry wave. Each model also includes the controls for student's age, gender, class assignment, teacher experience and its square, graduate degree, merit pay status and classroom peers when variation in the variable exists for the given sample.

* Statistically significant at 10-percent level

† Statistically significant at 5-percent level

‡ Statistically significant at 1-percent level

Table 11 – Estimated Effects of Cumulative Years with an Own-Race Teacher on Mathematics and Reading Scores

Variable	Math Score	Reading Score
Cumulative Years with an Own-Race Teacher		
One	3.3‡	2.3†
	(1.0)	(1.0)
Two	6.7‡	5.0‡
	(1.2)	(1.2)
Three	9.9‡	8.6‡
	(1.7)	(1.6)
Four	13.9‡	11.6‡
	(2.1)	(2.0)
R^2	.2241	.2464
Sample Size	23,883	23,544

Notes – Robust standard errors are reported in parentheses. All models include grade fixed effects and the interactions of fixed effects for the entry school and entry wave. Each model also includes the controls for student's age, free lunch status, small class assignment, teacher experience, graduate degree, merit pay status and classroom peers.

* Statistically significant at 10-percent level

† Statistically significant at 5-percent level

‡ Statistically significant at 1-percent level