SIEPR Discussion Paper No. 08-36

Teacher Sorting and Own-Race Teacher Effects in Elementary School

By
Conrad Miller
Stanford University

May 2009

Stanford Institute for Economic Policy Research
Stanford University
Stanford, CA 94305
(650) 725-1874

Teacher Sorting and Own-Race Teacher Effects in Elementary School

Conrad Miller[§]

May 31, 2009

Abstract

I investigate "own-race teacher effects"—the extent that students benefit from having a teacher with the same racial background. Own-race teacher effects may justify recruitment of underrepresented groups in teaching and, in combination with peer effects, help to determine the optimal assignment of students to teachers. However, previous estimates of own-race teacher effects are likely confounded by the sorting of teachers across schools. To circumvent this endogenous sorting, I develop and estimate a teacher-level metric of own-race teacher effects based on teacher fixed effects using administrative data from North Carolina public schools. I find that own-race teacher effects are present for mathematics achievement, but significantly smaller than previous estimates.

Keywords: race, teacher quality, teacher sorting, achievement gap

[§]Department of Economics, Stanford University, Stanford, CA 94305 (email: ccmiller@stanford.edu). This paper was originally my undergraduate honors thesis at Stanford University. I am indebted to Caroline Hoxby for her guidance and support. I thank David Autor, Cecilia Conrad, William Darity, Jr., Giacomo DeGiorgi, Michael Lovenheim, Luigi Pistaferri, Geoffrey Rothwell, Jesse Weinstein-Gould, Shahrzad Zarafshar, and Stanford Labor and Development Reading Group participants for helpful comments and discussions. I gratefully acknowledge financial support from the Mellon Mays Undergraduate Fellowship and Stanford Undergraduate Research Assistant Program.

Introduction

Recent economic literature has established that teachers are a significant determinant of student achievement (Aaronson *et al.* 2007; Rivkin *et al.* 2005; Rockoff 2004). These studies suggest that a one standard deviation increase in a teacher's quality increases a student's achievement by 0.1 to 0.13 standard deviations. While this may come as little surprise to parents and educators, this overall effectiveness[1] does not depend on commonly used proxies for teacher quality. Easily observed teacher characteristics such as tenure, experience, certification, and education—variables schools use to hire, promote, and pay teachers—explain a small fraction of the variation in teacher quality within schools. Since most of the variation of teacher quality cannot be explained by such characteristics, many economists have now accepted the idea that only direct measures of teachers' value-added reveal the gains students can expect to experience. What determines a teacher's effectiveness is still a subject in need of study, and past conventions require further assessment.

One common and often-cited hypothesis among educators and policymakers is that minority students perform better when matched with own-race teachers. Racial minority groups have historically been underrepresented among teachers and this belief in own-race teacher effects has motivated the recruitment of minority teachers as a potential remedy for the achievement gap between white students and minority students (U.S. Department of Education 1997). It is important to understand own-race teacher effects to understand to what degree such recruitment practices can be effective in addressing the achievement gap and, more generally, in optimizing the assignment of teachers to students.

There are several sociological and psychological arguments for own-race teacher

[1] I use the phrases 'teacher quality' and 'teacher effectiveness' interchangeably. In this context, they refer to a teacher's ability to improve her student's standardized test scores. Ideally, teacher quality would be measured on several dimensions beyond standardized test scores, including graduation rates, educational attainment, earnings, and measures of social well-being. However, data shortcomings typically limit the analysis to standardized test scores.

effects, some with more empirical support than others. Ferguson (2003) provides a thorough survey of this literature. A black student may feel encouraged by the mere presence of a black teacher, perhaps because that teacher serves as a role model or contradicts a student's previously held beliefs about her own educational potential. Though intuitive and often-cited in educational policy, the existence of such role-model effects has little empirical support (Cizek 1995). Stereotype threat may also serve as a significant passive teacher effect (Steele 1997). That is, stereotypes that are emphasized when students are matched with teachers with a different racial background may have a detrimental effect on the student's achievement. Experimental evidence suggests that changes in the salience of race (e.g. the use of a pre-test demographic questionnaire) can alter a student's test performance (Steele 1997). There is empirical evidence that a teacher's allocation of class time, interaction with students, and design of class materials may be biased in favor of students who share that teacher's racial background (Ferguson 2003). Finally, racial dynamics may be important in parent-teacher interactions, which may be critical to a child's development (Ferguson 2003).

While there is a significant sociological literature investigating the importance of these racial dynamics, few studies attempt to assess whether students perform better *because* they are matched with an own-race teacher. Most notable of such studies are Dee (2004) and Ehrenberg *et al.* (1995), and their results contradict each other. Dee (2004) tests the hypothesis that students perform better when matched with own-race teachers using the data from the Tennessee Project STAR experiment, a large-scale randomized experiment intended to measure the effectiveness of small-sized classes. The author finds evidence that students perform better on a standardized test when paired with an own-race teacher. Ehrenberg *et al.* (1995) uses the National Education Longitudinal Study of 1988 to investigate how a student's subject test score gains between 8th and 10th grade relate to having an own-race 10th grade teacher in that subject. The authors demonstrate that, although teachers give more favorable subjective evaluations of own-race students, student's test gains are not significantly related to having an own-race teacher.

However, both of these studies do not to control for teacher effectiveness and its distribution across schools and racial groups.[2] Implicit in the econometric strategies of these papers is the assumption that, within a school, teachers from all racial backgrounds have the same distribution of quality. That is, the authors assume there is no differential selection of teachers *across* schools. This assumption is implicit because, although both studies include controls for observable teacher characteristics, these characteristics relate weakly to a teacher's effectiveness (Aaronson *et al.* 2007; Rivkin, *et al.* 2005; Rockoff 2004). Controlling for selection within schools could not remedy, for instance, a problem caused by more effective black teachers preferring to teach in schools with more black students, and more effective white teachers preferring to teach in schools with more white students. If there is reason to believe that, within a school, teacher quality is not identically distributed across races, this assumption is not reasonable, and results in Ehrenberg *et al.* (1995) and Dee (2004) may be misleading.

In this paper I review past literature on own-race teacher effects and relevant literature on teacher sorting. Using a simple model, I show how differential teacher sorting across schools by teachers' race may bias naive estimates of own-race teacher effects. Using administrative data that includes the entire population of North Carolina public school students and teachers from 1995 to 2003, I develop a teacher-level metric for own-race teacher effects that circumvents the problem of teacher sorting, and estimate these effects. I show that when teachers are permitted to have separate effects on black and white students, there is some evidence of positive own-race effects for mathematics achievement; however, my estimates are substantially smaller than

[2] Dee (2004) takes note of this shortcoming and attempts to address it.

those in Dee (2004).

In addition to circumventing the problem of teacher sorting, my metric of own-race teacher effects has the advantage of being teacher-level, while previous studies only analyze aggregate populations. The micro nature of my metric allows for more interesting study of the properties of own-race teacher effects, including where and with which teachers and students own-race teacher effects are most pronounced.

Differential Sorting: An Illustrative Model

To see why differential sorting may bias the estimates of own-race teacher effects found in Dee (2004) and Ehrenberg *et al.* (1995), consider the following simplified model.

Suppose there are two types of students and teachers, black (B) and white (W). Teachers are either high quality (H) or low quality (L). Schools are segregated by student race. For simplicity, I assume schools are perfectly segregated (i.e. all-black or all-white), but the analysis and conclusion is similar for any degree of segregation. Suppose that in the true model for student achievement, gains depend solely on teacher quality and an idiosyncratic error. There are no own-race teacher effects. Equivalently,

$$\Delta A_{it} = \alpha_0 + \alpha_1 I_{it}^H + \epsilon_{it} \tag{1}$$

where i indexes students, t indexes years, A is the metric for student achievement and I^H is an indicator for whether the student's teacher in that period is high quality.

Further suppose that, to assign teachers to schools, districts give each teacher a list of possible schools at which he or she may choose to work, and the teacher works at her most preferred school of the available schools. Districts prefer high quality teachers, but due to wage setting constraints, cannot offer salaries based on quality. Instead, districts offer high quality teachers a longer (in general, more desirable) list of possible schools. If teachers prefer to teach students from their own race, this assignment mechanism will lead to differential sorting such that

$$\mathbf{Pr}(I^{own-race}|H) > \mathbf{Pr}(I^{own-race}|L) \tag{2}$$

where $I^{own-race}$ is an indicator for whether the student and teacher are from the same race. That is, high quality teachers are more likely to teach own-race students than low quality teachers because they are more likely to be provided the opportunity to do so. If a researcher interested in own-race teacher effects ignores teacher quality and estimates

$$\Delta A_{it} = \beta_0 + \beta_1 I_{it}^{own-race} + \epsilon_{it} \tag{3}$$

the estimate for own-race teacher effects, β_1, will be upwardly biased. More precisely,

$$Bias(\hat{\beta}_1) = \alpha_1 \frac{Cov(I^{own-race}, I^H)}{Var(I^{own-race})} > 0$$

With differential teacher sorting, the empirical strategy employed in Dee (2004) and Ehrenberg *et al.* (1995) may lead to biased estimates of own-race teacher effects. In the presence of sorting, an appropriate empirical strategy should effectively control for teacher quality. For example, under the simplified model, estimating

$$\Delta A_{it} = \beta_0 + \beta_1 I_{it}^{own-race} + \beta_2 I_{it}^H + \epsilon_{it} \tag{4}$$

would yield the correct estimate of β_1, zero.

However, the current method for measuring teacher quality utilizes student achievement, the same dependent variable we use to identify own-race teacher effects. Estimates of teacher quality already include potential own-race teacher effects, so in the presence of own-race teacher effects, we cannot simply estimate equation 4 using standard estimates of teacher quality. To see this, consider the following example.

Suppose Mrs. Jones, a black teacher, has black and white students, where $\rho_{black}^{Jones} < 1$ is the proportion of her students that are black. With white students, her fixed effect on student achievement is δ, and with black students, her fixed effect is $\delta + \theta$. Equivalently, one can interpret Mrs. Jones' fixed effect as δ, and the own-race teacher effect as θ. However, if I were to estimate Mrs. Jones' fixed effect without accounting for the heterogeneity across student groups, I would estimate her fixed effect as $\delta + \rho_{black}^{Jones}\theta$, a weighted average of her fixed effects with black and white students. This estimate includes a portion of θ, the own-race teacher effect. If I were to estimate

$$\Delta A_{ijt} = \beta_0 + \beta_1 I_{ijt}^{own-race} + T_j + \epsilon_{it} \tag{5}$$

where j indexes teachers and T_j is the standard estimate of the teacher j's quality, my overestimate of T_{Jones} would bias my estimate of β_1 (the true value is θ) downwards.

To circumvent this problem, I estimate each teacher's fixed effect with respect to black and white students separately, and use the difference to estimate θ. In the case of Mrs. Jones, I would estimate her fixed effect with white students, δ, her fixed effect with black students, $\delta + \theta$, and use the difference as an estimator for θ.

Literature Review

Own-Race Teacher Effects [3]

Using the NELS-88 dataset, Ehrenberg *et al.* (1995) test whether students experience larger test score gains when they are matched with an own-race or own-gender teacher. They also investigate whether teachers submit more positive subjective evaluations of

[3] A related literature examines own-gender teacher effects, which are potentially analogous to own-race teacher effects in causes and properties (Dee 2007; Carrell *et al.* 2009; Qian and Zafar 2009).

same-gender students conditional on the students' test scores. The authors estimate the effect of having an own-race or own-gender 10th grade teacher in a particular subject on test score gains in that subject, including controls for student, teacher, and school characteristics. Ehrenberg *et al.* (1995) find little evidence that the race and gender of teachers have a significant effect on test outcomes. In the analogous investigation of teacher subjective evaluations of students, the results indicate that white female teachers give better evaluations to white female students, and white male teachers give better evaluations to white male students.

Dee (2004) tests the hypothesis that students perform better when matched with own-race teachers using the Tennessee Project STAR experiment data. The STAR experiment was a large-scale randomized experiment intended to measure the effectiveness of small-sized classes. Kindergarten students and teachers were randomly assigned to small classes, regular-sized classes, and regular-sized classes with teacher aides within 79 participating schools. The experiment continued through the third grade for the cohort of students. The author exploits this ostensibly random assignment of students to teachers to measure the effects of racial pairings on student standardized test scores. In baseline specifications with a variety of student, teacher, and peer controls, the author finds that assignment to an own-race teacher is associated with a 3 to 6 percentile point increase in reading scores for white and black males and black females, while there is no significant effect for white females. In mathematics, the effect ranges from 3 to 5 percentile points for all groups.

Dee (2004) then addresses a potential identification issue, this time related to teacher unobservable factors. Teacher unobservables may bias the estimates of own-race teacher effects because teachers are not randomly assigned *across* schools. For example, the effect for black students could also arise because, in predominantly black schools, the black teachers are of a significantly higher quality than white teachers. An analogous explanation could be given for white students. The author attempts to address this issue by introducing classroom fixed effects. Classroom fixed effects should absorb teacher fixed effects as well as peer effects and other classroom-level factors. However, this approach does not account for potential interracial differences in school student composition beyond gender, free lunch eligibility, and year of birth. For example, the racial achievement gap may be larger (smaller) in schools that are mostly white (black), controlling for these demographic variables, for reasons other than classroom racial dynamics. Such differences would bias estimates in Dee (2004), even in the presence of classroom fixed effects. In addition, Dee (2004) does not allow the relationship between student achievement and peer characteristics to depend on student race. Just as there may be own-race teacher effects, there may be own-race *peer* effects, perhaps with ambiguous sign. Estimates of own-race teacher effects derived in this way would partially observe any differential peer-effects. The inclusion of classroom fixed effects appears to reduce the magnitude of own-race teacher effects, but does not render them entirely insignificant.

Dee (2004) also attempts to assess the issue of differential teacher sorting by analyzing different types of schools separately. He posits that if poorer and more racially homogeneous schools have difficulties recruiting high quality teachers and own-race teacher effects are concentrated in these schools, then these unobserved teacher quality issues may be important. He finds that, for black students, own-race teacher effects are concentrated among disadvantaged students and in more segregated schools. This suggests that the distribution of teacher unobservables may be a relevant confounding influence, or that own-race teacher effects are more significant in more segregated schools for some other reason. For white students, effects are similar across socioeconomic groups and schools.

Though Dee (2004) attempts to do so with the means available to him, Dee (2004) and Ehrenberg *et al.* (1995) do not control for teacher effectiveness and its distribution across schools and racial groups. Implicit in the econometric strategies of these papers is the assumption that, within a school, teachers from all racial backgrounds have the same distribution of quality. Systematic differences in a teacher's pupils' achievement are only related to the racial interaction of teacher and student, conditional on the school and the student's previous achievement. This assumption is implicit because, although both studies include control for observable teacher characteristics, these characteristics are only weakly associated with a teacher's effectiveness (Aaronson *et al.* 2007; Rivkin *et al.* 2005; and Rockoff 2004). As a robustness check, Dee (2004) includes classroom fixed effects that should absorb teacher fixed effects, but under this specification, both classroom fixed effects and own-race teacher effects are likely poorly identified.

Teacher Sorting

Research on teacher mobility and preferences suggests that the distributional assumption implicit in the previous own-race teacher effect literature is problematic. Teacher salary scales are generally rigid within states, both across schools within a district and across districts within a state. In the presence of these wage setting constraints, school and job characteristics serve as an alternative means of compensation. Districts provide highly effective teachers a choice among schools as a compensating differential, allowing those teachers to work in schools with characteristics they prefer. Thus, if black and white teachers have different preferences for schools and students, it is likely that the two groups will sort differently across schools in a manner correlated with teacher quality.

Several recent studies examine the relationship between teacher mobility and school demographics by investigating the movement of individual teachers across schools. For districts in California, Georgia, New York and Texas, researchers have found that teachers in schools with low-achieving students, particularly those with more experience—those more likely to have the opportunity to change schools—tend to move to schools with higher-achieving students, leaving many poor and minority students in schools with teaching vacancies and inexperienced, less qualified teachers (Betts, Rueben and Danenberg 2000; Lankford *et al.* 2002; Hanushek and Rivkin 2004; Scafidi *et al.* 2007). However, this pattern of movement is not uniform across teacher racial groups, and the behavior of white teachers dominates the overall pattern. Using data from Texas public schools, Hanushek and Rivkin (2004) find that white teachers systematically move to schools with fewer minority students, fewer poor students, and better prior student test scores. Conversely, black teachers move to schools with more black students and comparable prior student test scores. This differential sorting suggests that teachers prefer own-race students, schools have racial preferences for teachers, or teachers and schools have preferences over factors *associated* with student and teacher race, that are not racial *per se*.

This sorting also could be attributed at least partially to residential segregation. Using data from New York State, Boyd *et al.* (2005) find that teachers express preferences to work close to where they grew up and, conditional on proximity, prefer areas with characteristics similar to their hometown. Teacher spatial preferences and spatial correlations between teachers' hometowns and school demographics could generate the aforementioned relationship between teacher movement and student demographics even if teachers have no preferences for students or teachers. For example, because experienced teachers are usually given preference for new teaching positions they are more able to express their preferences for schools. If experienced teachers—in particular, white experienced teachers—tend to live in suburban neighborhoods, they may prefer to teach in those areas and move from an inner-city school to a suburban school given the opportunity.

Jackson (2008) argues that spatial preferences are unlikely to serve as a complete explanation for differential teacher sorting, however. Exploiting the reshuffling of students caused by the end of a student busing desegregation program in a school district of North Carolina, he finds that experienced, white and high value-added teachers were relatively more likely to leave schools that experienced substantial exogenous inflows of black students. Since the neighborhoods and school locations themselves did not change, this result suggests that teachers have preferences for student attributes that are correlated with race, including race itself, *independent* of school location.

Altogether, recent empirical evidence and theory on teacher sorting suggests that teachers differentially sort across schools in a manner that may bias previous estimates of own-race teacher effects.

Data

I use administrative data on students in North Carolina public schools from 1995 to 2003. The dataset is compiled, cleaned, and distributed by the North Carolina Research Data Center and contains end-of-grade test scores in math and reading for grades 3 through 8. Test scores are mapped to encrypted student identifiers so that students can be examined longitudinally. The data also identify the school staff member who administered the test, which is usually the regular teacher in elementary grades. These identifiers, too, are encrypted to ensure confidentiality. I use a linked personnel database to identify test administrators who had regular teaching assignments and improve the student-teacher match rate. For 6th grade test scores through 8th grade scores, this matching method is unreliable because middle school students typically have different teachers for each subject. Thus I restrict my attention to data from the 3rd, 4th and 5th grades. To ensure that teacher and student effects are consistently well-identified, I further limit the sample to students observed with valid teachers in all three grades. I present descriptive statistics in Table 1. I do not present the mean value of test scores because the scoring system changes across years. However, the scaling only shifts by a constant; standard deviations are consistent across years. All further analysis includes year fixed effects, which difference out these constants.

Empirical Methodology

I determine whether differences in average quality between black and white teachers within schools relate significantly to the racial composition of the school. Such a relationship would indicate that teacher sorting across schools likely biases previous estimates of own-race teacher effects.

To estimate teacher quality, I use the "value-added" approach typical of the recent teacher effects literature (see, for example, Aaronson et al. 2007; Rockoff 2004; Rivkin et al. 2005). I interpret "teacher quality" as a teacher's "fixed effect" on students' test score improvements—a student's expected test score gain when assigned to a particular teacher's classroom relative to the omitted teacher's classroom. While this measure provides no insight as to *how* a teacher affects student test scores, it provides an overall reflection of a teacher's knowledge of the course material, ability to engage students, pedagogical approach, and a number of other factors.

Reliable identification of causal teacher fixed effects requires student-teacher matched student achievement data that is longitudinal in both students and teachers. Consistent identification also requires that the information used to assign students to teachers is observable or unrelated to student achievement gains. If teachers are consistently assigned students that differ in unmeasured ways—for example, more motivated students, students with stronger unmeasured prior achievement or more engaged parents—that result in varying student achievement gains, then, rather than reflecting the talents and skills of individual teachers, estimates of teacher effects may

Table 1: Descriptive Statistics for the Student Data

	All Students (1)		Students Included in Analysis (2)	
Population Size	1,129,691		360,283	
	Mean	SD	Mean	SD
Test Score				
Mathematics, 3rd Grade		10.9		10.3
Mathematics, 4th Grade		9.96		9.60
Mathematics, 5th Grade		9.80		9.58
Reading Comprehension, 3rd Grade		9.89		9.34
Reading Comprehension, 4th Grade		9.36		9.07
Reading Comprehension, 5th Grade		8.44		8.12
Demographics				
Female	.489	.500	.503	.500
White	.615	.487	.654	.476
African American	.296	.456	.283	.451
Hispanic	.043	.204	.024	.154
Asian	.017	.130	.013	.115
Native American	.015	.120	.013	.113
Eligible for free school lunch	.362	.481	.312	.463
Eligible for reduced-price school lunch	.064	.245	.089	.285
Parental Education				
Did not finish high school	.126	.332	.105	.307
High school graduate	.438	.496	.450	.497
Trade or business school graduate	.056	.230	.056	.230
Community, technical or junior college graduate	.121	.326	.123	.328
Four-year college graduate	.161	.367	.185	.388
Graduate school degree	.086	.281	.073	.260

Notes: Standard deviations of test scores are averages of year specific test scores. I do not present the mean value of test scores because the scoring system changes across years. However, the scaling only shifts by a constant; standard deviations are consistent across years.

reflect principals' preferential treatment of certain colleagues, ability-tracking based on information not captured by prior test scores, or the advocacy of engaged parents for certain teachers (Kane and Staiger 2008). Empirically, however, such assignment based on unobservables does not appear to introduce significant bias. Using data from a random assignment experiment in the Los Angeles Unified School District, Kane and Staiger (2008) cannot reject that non-experimental estimates of teacher quality match experimental estimates.

Researchers have used several variants of teacher effects specifications with different assumptions about the underlying technology of student achievement (for a full discussion, see McCaffery et al. 2003). I use the teacher fixed effect specification that Kane and Staiger (2008) find introduces the least bias. The specification has the general form

$$A_{ikt} = \gamma A_{ikt-1} + T_{it} + \mathbf{X}_i\alpha + \mathbf{C}_{ikt}\lambda + g_{it} + \tau_t + \epsilon_{ikt} \tag{6}$$

where A_{ikt} is the test score of student i at school k at time t, $\mathbf{X}_i$ is a vector of student background controls, $\mathbf{C}_{ikt}$ is a vector of classroom peer controls, and T, g, and τ are teacher, grade, and year fixed effects. The T terms are the 'quality' estimates for each teacher estimated relative to the omitted teacher. The error term ϵ_{ikt} includes time-variant influences on test scores or academic performance, such as parental divorce, test day conditions, or changes in a student's motivation. As long as students are not assigned to teachers based on these time-variant shocks, these omitted variables do not bias the teacher fixed effect estimates.

I use two variants of equation 6 to examine how robust results are to the introduction of additional controls. The base specification does not include peer controls, and the variant does. The variant is my preferred specification because it most closely matches experimental estimates (Kane and Staiger 2008). My student background

Table 2: Standard Deviation of Teacher Fixed Effects

	Mathematics		Reading	
Specification	(1)	(2)	(3)	(4)
White Teachers	1.863	1.843	.941	.925
Black Teachers	1.866	1.874	.981	.975
Combined	1.866	1.848	.947	.932
Teacher Fixed Effect				
Grade Fixed Effect	Yes	Yes	Yes	Yes
Year Fixed Effect	Yes	Yes	Yes	Yes
Classmate Characteristics	No	Yes	No	Yes

Notes: Calculations only include teacher fixed effect estimates based on at least 15 student test scores.

controls consist of indicators for student ethnicity, gender and highest level of education attained by the parent. Peer controls consist of the classroom means of the student background controls.

For statistical precision, I use empirical Bayes estimates of teacher fixed effects, effectively shrinking noisy estimates towards the mean of the value-added distribution (in this case zero). Teacher fixed effects are estimated with noise: $\hat{T}_j = T_j + \mu_j$ where μ_j is random estimation error (under the identifying assumptions) and $T_j \sim \mathbf{N}(0, Var(T))$. To adjust for this sampling error, I use the estimate $T_j^{EB} = T_j \dfrac{\sigma_T^2}{\sigma_T^2 + \sigma_{\mu_j}^2}$, where σ_T^2 is the sample estimate of the variance of the true teacher quality distribution and $\sigma_{\mu_j}^2$ is the sample estimate of the variance of teacher j's quality estimate. For a complete derivation, see Kane and Staiger (2008).

I present summary statistics of my basic teacher fixed effect estimates in Table 2.

To circumvent the confounding influence of teacher sorting, I use an own-race teacher effect metric that is teacher-level. I measure teacher quality using the same the equation 6 value-added model, but allow the estimate to depend on the racial group of the student. Due to the paucity of Latino, Asian, and Native American students and teachers in the North Carolina dataset, I limit my analysis of own-race teacher effects to white and black teachers and students. I run the fixed effect regression but allow separate fixed effects for black and white students. Thus, under the same statistical assumptions, I identify each teacher's causal effect on test score improvements for black and white students. Formally, I estimate

$$A_{ikt} = \gamma A_{ikt-1} + T_{it} + \mathbf{X}_i\alpha + ... + T_{it}^{black} + I_i^{black}(\tilde{\gamma} A_{ikt-1} + \mathbf{X}_i\tilde{\alpha} + ...) + \epsilon_{ikt} \qquad (7)$$

where $\mathbf{X}_i\alpha + ...$ are the same set of controls as in equation 6. [4] I define my own-race teacher effect metric for arbitrary teacher j as $R = T_{it}^{black}$, the coefficient of the interaction term of an indicator for teacher j and an indicator for a black student, which I interpret as the extent that teacher j improves the test scores of black students more than the test scores of white students.[5] In the absence of own-race teacher effects, this difference should not depend on the race of the teacher. If R is generally larger for black teachers, that would be evidence for the existence of own-race teacher effects.

Formally, I estimate

$$R_j = \alpha + \beta I_j^{black} + \epsilon_j \qquad (8)$$

For teacher j, where I_j^{black} is 1 if teacher j is black and 0 otherwise. [6] To interpret β, consider the following thought experiment. Suppose there are two teachers, one black and one white, with the same quality as measured with respect to white stu-

[4] The variances of both T_{it} and T_{it}^{black} are comparable to the variance of the general teacher fixed effect estimates.

[5] This metric can be used with any two groups.

[6] In equation 8, I use dependent variable, R_j, that are generated from auxiliary regressions. That is, I estimate teacher fixed effects and use those estimates to construct R_j, the dependent variables used to analyze teacher own-race teacher effects. When the dependent variable in a regression is based on estimates, the regression residual is partially comprised of sampling error, the difference between the true value of the dependent variable and the estimated value. This sampling error is heteroskedastic if the sampling variance differs across observations. Using a series of Monte Carlo experiments, Lewis and Linzer (2005) find that when the sampling error component is small relative to the total variation of the dependent variable, and the sampling error does not vary greatly across observations, Huber-White standard error estimates yield accurate results. I use only teacher fixed effect estimates based on at least 15 observations and use empirical Bayes estimates to reduce the variance of sampling error, so I am likely in this case. I use Huber-White standard error estimates throughout my analysis.

dents. Then β is the additional gain a black student would experience with the black teacher. Symmetrically, if the two teachers have the same quality with respect to black students, β is the additional gain a white student would experience with the white teacher.

Note that this symmetry is not imposed by the model, but is instead a consequence of interpreting β from a student perspective. To determine a student's expected gain from having an own-race teacher based on quality estimates for individual teachers, I must compare the student's expected gains with a black and white teacher of similar quality, defining 'similar' using either the estimate based on white students' or black students' performances. For example, if I want to determine own-race teacher effects for black students, I would compare the teacher quality estimates with respect to black students for black teachers and white teachers, *controlling* for teacher quality estimates with respect to white students. Otherwise, I would misattribute general quality differences (with respect to black *and* white students) as own-race teacher effects. Once I control for teacher quality in this way, I am effectively comparing R for black and white teachers. If I were to perform the same exercise for white students, I would do the same but using $-R$, the extent that a teacher improves the test scores of white students more than the test scores of black students. Thus, the symmetry is inherent to the outlined method when interpreting own-race teacher effects from a student perspective.

Alternatively, my metric of own-race teacher effects can be interpreted as a racial group of teachers' *comparative advantage* in teaching own-race students. While a particular black teacher may have an *absolute advantage* with both black and white students (e.g. generating larger test score gains for both groups of students) compared to a particular white teacher, in the presence of own-race teacher effects (and constraints on class size), the white teacher will have a comparative advantage with white students. This relationship is symmetric in the sense that the 'extent' a white teacher has a comparative advantage with white students is equivalent to the 'extent' the black teacher has a comparative advantage with black students. In this case, the 'extent' is exactly my metric of own-race teacher effects.

However, this symmetry is not necessary when interpreting own-race teacher effects from a teacher perspective (e.g. it is not necessarily the case that a black teacher's quality with respect to white students is equal to a white teacher's quality with respect to black students). In this case, we are comparing the size absolute advantages, which need not be symmetric.

Results

I estimate equation 8 using teachers with fixed effect estimates based on at least 15 black student test scores and 15 white student test scores.[7] This effectively limits the sample to teachers in sufficiently racially integrated schools. A significantly positive estimate for β would imply the existence of own-race teacher effects—relative to white teachers, black teachers perform better with black students than with white students. I estimate equation 8 with respect to mathematics and reading achievement scores separately. Under both teacher fixed effect specifications, I estimate equation 8 using no controls, school-level controls, and school fixed effects to control for school type. I use black enrollment share, free lunch share, and reduced lunch share as school-level controls, all from the 2000 Public School Universe Survey. The results are in Table 4 and Table 5.

For mathematics, the estimates suggest that own-race teacher effects exist and are statistically significant. In my preferred specification, (5), point estimates for the coefficient on an indicator for whether a teacher is black is .155—roughly equivalent to 1/10 standard deviations in teacher quality—and statistically significant at the 1%

[7] However, results are generally robust to using a less restrictive requirement of 10 black student test scores and 10 white student test scores.

level. The degree that teachers perform 'better' with students from their racial group is approximately equivalent to 1/10th a standard deviation in teacher quality, .016 standard deviations in student achievement, or 2% of the black-white achievement gap in the data. However, the point estimate is at least 3.5 to 7 times smaller in magnitude than estimates presented in Dee (2004).[8] Estimates do not change significantly when I weight teachers by number of observations. When I substitute school fixed effects for school-level controls, the point estimate is reduced to .129 and is statistically significant at the 5% level, but the estimates are not statistically distinguishable. School fixed effects are perhaps overly restrictive. For example, if some schools have only white teachers with valid fixed effect estimates, a school fixed effect may eliminate their contribution to the estimation. If school-level contributions to within-teacher differences in effectiveness are sufficiently controlled for with the variables I have, then a specification without school fixed effects is preferable because it allows comparison across similar schools.

For reading, all of the estimates are smaller and statistically insignificant. Own-race teacher effects do not appear to be a significant determinant of reading achievement. This is in contrast to the substantial effects found for black students and white male students in Dee (2004), which were similar to the results for mathematics.

There are a few potential reasons that my estimates are substantially smaller than those in Dee (2004). The populations analyzed are different on several dimensions. I examine 3rd, 4th, and 5th graders in North Carolina from 1995 to 2003; Dee (2004) examines one cohort of kindergarteners through 3rd grade in Tennessee beginning in 1985. Own-race teacher effects may be pronounced in earlier graders, in the state of Tennessee, or in an earlier time period. In addition, my sample is limited to teach-

[8]Dee (2004) states estimates in terms of percentile rather than standard deviations. He finds an own-race teacher effect of 2-4 percentile points. To make them comparable, I convert the Dee (2004) estimates to standard deviations by assuming a normal distribution and evaluating at the mean. This conversion is likely an underestimate because percentiles translate to larger standard deviations on every other portion of a normal distribution.

Table 3: Own-Race Teacher Effects, Mathematics

Specification	(1)	(2)	(3)	(4)	(5)	(6)
Black Teacher	.191‡	.165‡	.164‡	.145‡	.155‡	.129†
	(.048)	(.054)	(.054)	(.047)	(.049)	(.054)
Constant	.143‡	.150‡		-.044†	-.051	
	(.020)	(.020)		(.020)	(.056)	
School Fixed Effect?	No	No	Yes	No	No	Yes
School-Level Controls?	No	Yes	No	No	Yes	No
Teacher Fixed Effect						
Grade Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes
Classmate Characteristics	No	No	No	Yes	Yes	Yes
# of Teachers	4394	4246	4394	4394	4246	4394
# of Schools	889	847	889	889	847	889
Adjusted R^2	.004	.060	.057	.002	.003	.043

Notes: Huber-White robust standard errors are reported in parentheses.
* Statistically significant at 10% level.
†Statistically significant at 5% level.
‡Statistically significant at 1% level.

ers who teach at least 15 black students *and* 15 white students, teachers in at least somewhat integrated schools. It may be the case that own-race teacher effects are less relevant in such schools. However, these are also the schools in which teacher sorting would introduce the most bias to traditional estimates. It is possible that my causal teacher fixed effects are poorly identified; however, Kane and Staiger (2008) show experimental evidence that similar estimates of causal teacher fixed effects are accurate.

Previous teaching sorting results suggest that a substantial portion of the difference may be attributed to the manner in which teachers differentially sort across schools. This endogenous sorting would significantly bias estimates such as those found in Dee (2004) upwards. Even when Dee (2004) includes classroom fixed effects, he does not control for students' past achievement to account for differential student sorting across schools or allow for differential peer effects.

Table 4: Own-Race Teacher Effects, Reading

Specification	(1)	(2)	(3)	(4)	(5)	(6)
Black Teacher	-.030	-.031	.010	-.013	-.018	.015
	(.037)	(.038)	(.043)	(.036)	(.037)	(.041)
Constant	.150‡	.028‡		.032†	-.079*	
	(.016)	(.044)		(.016)	(.043)	
School Fixed Effect?	No	No	Yes	No	No	Yes
School-Level Controls?	No	Yes	No	No	Yes	No
Teacher Fixed Effect						
Grade Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes
Classmate Characteristics	No	No	No	Yes	Yes	Yes
# of Teachers	4372	4372	4372	4372	4224	4372
# of Schools	889	847	889	889	847	889
Adjusted R^2	.000	.007	.074	.000	.005	.066

Notes: Huber-White robust standard errors are reported in parentheses.
* Statistically significant at 10% level.
†Statistically significant at 5% level.
‡Statistically significant at 1% level.

Teacher Race and Other Student Categorizations

One potential worry about the metric of own-race teacher effects estimated in equation 8 is that it is unrelated to racial dynamics, but instead acting as a proxy for other comparative advantages black and white teachers have with particular groups of students. For example, it may be the case that black teachers are more effective with low-achieving students and white teachers are more effective with high-achieving students, across racial groups. Since black students are more likely to be low-achieving in the North Carolina data, equation 8 may simply be picking up this teacher race to student achievement relationship. Alternatively, there may be an analogous relationship between teacher race and student affluence.

To assess the degree of this potential confounding influence, I estimate equation 8 again, but instead of using $R = T_{it}^{black}$ as my dependent variable, I use an analogous measure for a teacher's additional effectiveness with low-achieving students relative to high-achieving students, and free and reduced lunch eligible students relative to ineligible students. I define low-achieving students as those whose third grade test score places them in the bottom third of their year's overall distribution, and high-achieving students as those who place in the top third. In this context, a significant and positive β would indicate that black teachers generally have a comparative advantage with low-achieving or poor students.

The results are presented in Tables 6 and 7. I use the intermediate teacher fixed effect specification, my preferred specification that includes classroom peer controls, for all cases. When comparing free and reduced lunch eligible students to ineligible students, the estimated coefficient on an indicator for whether a teacher is black is statistically insignificant at the 10% level. Black and white teachers appear to have no systematic comparative advantages with poorer students and affluent students that might confound my own-race teacher effect estimates. However, black teachers do have a comparative advantage with low-achieving students. Point estimates for the coefficient on an indicator for whether a teacher is black is .239 or .207 for mathematics scores, depending on the controls used for school type, and statistically significant at the 1% level. Analogous estimates for reading scores are small and statistically insignificant. For mathematics, these estimates are somewhat larger than my estimates for own-race teacher effects. When I further subdivide the categories of low-achieving and high-achieving students by student race, the point estimates are similar but the standard errors increase enough to render them statistically insignificant. The increase in standard error is possibly due to the reduction in sample size. Thus, it is possible that my estimates of own-race teacher effects are driven by black teachers' comparative advantage with low-achieving students. It may also be that both comparative advantages independently exist. I am unable to distinguish between these cases. For the 3177 teachers with 15 observations with white, black, low-achieving, and high-achieving students, the correlation between the achievement

Table 5: Teacher Race, Student Achievement, and Student Socioeconomic Status, Mathematics

	Student Achievement		Student SES	
Specification	(1)	(2)	(3)	(4)
Black Teacher	.239‡	.207†	.023	.028
	(.074)	(.082)	(.041)	(.047)
Constant	.040		.085‡	
	(.053)		(.020)	
School Fixed Effect?	No	Yes	No	Yes
School-Level Controls?	Yes	No	Yes	No
# of Teachers	5138	5290	5479	5658
# of Schools	1084	1131	1070	1125
Adjusted R^2	.003	.059	.003	.021

Notes: Huber-White robust standard errors are reported in parentheses. All specifications use the preferred teacher fixed effect specification, which includes mean classroom peer controls.
* Statistically significant at 10% level.
†Statistically significant at 5% level.
‡Statistically significant at 1% level.

differentials and race differentials is .451.

Conclusion

The results presented here suggest that previous estimates of own-race teacher effects may be confounded by the endogenous sorting of teachers across schools. Using a teacher-level metric of own-race teacher effects that circumvents this sorting problem, I find that own-race teacher effects are present for mathematics achievement, but substantially smaller than estimates in Dee (2004). I find that the degree that teachers perform 'better' with students from their racial group is approximately equivalent to 1/10th a standard deviation in teacher quality, or .016 standard deviations in student achievement, or 2% of the black-white achievement gap—estimates at least 3.5 to 7 times smaller than those in Dee (2004). I also find a similar comparative advantage

Table 6: Teacher Race, Student Achievement, and Student Socioeconomic Status, Reading

	Student Achievement		Student SES	
Specification	(1)	(2)	(3)	(4)
Black Teacher	.065	.026	-.018	.005
	(.047)	(.051)	(.034)	(.037)
Constant	.042		-.003	
	(.032)		(.029)	
School Fixed Effect?	No	Yes	No	Yes
School-Level Controls?	Yes	No	Yes	No
# of Teachers	5362	5516	5444	5624
# of Schools	1094	1141	1071	1126
Adjusted R^2	.004	.090	.002	.023

Notes: Huber-White robust standard errors are reported in parentheses. All specifications use the preferred teacher fixed effect specification, which includes mean classroom peer controls.
* Statistically significant at 10% level.
†Statistically significant at 5% level.
‡Statistically significant at 1% level.

for black teachers with low-achieving students, which is potentially driving own-race teacher effect estimates.

These estimates of own-race teacher effects are small; however, it is possible that these effects are more substantial under certain circumstances. There is substantial heterogeneity in own-race teacher effects across teachers. The micro nature of my teacher-level metric allows for future study of the properties of own-race teacher effects, including where and with which teachers and students they are most pronounced. Such exercises may help determine why own-race teacher effects exist and in what context they are most relevant for policy.

References

Aaronson, Daniel, Lisa Barrow, and William Sander, "Teachers and Student Achievement in the Chicago Public High Schools," *Journal of Labor Economics*, 2007, *25* (1), 95–135.

America's Teachers: Profile of a Profession, 1993-94, Washington D.C.: U.S. Department of Education National Center for Education Statistics, 1997.

Betts, Julian R., Kim S. Rueben, and Anne Danenberg, "Equal Resources, Equal Outcomes? The Distribution of School Resources and Student Achievement in California," Technical Report, Public Policy Institute of California, San Francisco 2000.

Boyd, Donald, Hamilton Lankford, Susanna Loeb, and James Wyckoff, "The Draw of Home: How Teachers' Preferences for Proximity Disadvantage Urban Schools," *Journal of Policy Analysis and Management*, 2003, *24* (1), 113–132.

Carrell, Scott B., Marianne E. Page, and James E. West, "Sex and Science: How Professor Gender Perpetuates the Gender Gap," February 2009. Working paper.

Clotfelter, Charles T., Helen F. Ladd, and Jacob L. Vigdor, "Who Teaches Whom? Race and the Distribution of Novice Teachers," *Economics of Education Review*, August 2005, *24* (4), 377–392.

– **, – , and** – , "Teacher-Student Matching and the Assessment of Teacher Effectiveness," *Journal of Human Resources*, Fall 2006, *41* (4), 778–820.

Dee, Thomas S, "Teachers, Race and Student Achievement in a Randomized Experiment," *Review of Economics and Statistics*, February 2004, *86* (1), 195–210.

– , "A Teacher Like Me: Does Race, Ethnicity or Gender Matter?," *American Economic Review*, May 2005, *95* (2), 158–165.

– , "Teachers and the Gender Gaps in Student Achievement," *Journal of Human Resources*, Summer 2007, *44* (42), 528–554.

Downey, Douglas B. and Shana Pribesh, "When Race Matters: Teachers Evaluations of Students Classroom Behavior," *Sociology of Education*, October

2004, *77* (4), 267–282.

Ehrenberg, Ronald G., Dan D. Goldhaber, and Dominic J. Brewer, "Do Teachers Race, Gender and Ethnicity Matter? Evidence from the National Educational Longitudinal Study of 1988," *Industrial and Labor Relations Review*, April 1995, *48* (3), 547–561.

Ferguson, Ronald F., "Teachers Perceptions and Expectations and the Black-White Test Score Gap," *Urban Education*, 2003, *38* (4), 460–507.

Hanushek, Eric A., John F. Kain, and Steven G. Rivkin, "Why Public Schools Lose Teachers," *Journal of Human Resources*, Spring 2004, *39* (2), 326–354.

Jackson, C. Kirabo, "Student Demographics, Teacher Sorting, and Teacher Quality: Evidence from the End of School Desegregation," *Journal of Labor Economics*, 2009, *forthcoming.*

Kane, Thomas J. and Douglas O. Staiger, "Are Teacher-Level Value-Added Estimates Biased? An Experimental Validation of Non-Experimental Estimates," December 2008. NBER Working Paper No. 14607.

Lankford, Hamiliton, Susanna Loeb, and James Wyckoff, "Teacher Sorting and the Plight of Urban Schools: A Descriptive Analysis," *Educational Evaluation and Policy Analysis*, 2002, *24* (1), 37–62.

Lewis, Jeffrey B. and Drew A. Linzer, "Estimating Regression Models in Which the Dependent Variable is Based on Estimates," *Political Analysis*, Autumn 2005, *13* (4), 345–364.

McCaffrey, Daniel, J.R. Lockwood, Daniel Koretz, and Laura Hamilton, *Evaluating Value-Added Models for Teacher Accountability*, Santa Monica, CA: RAND Corporation, 2003.

Qian, Yi and Basit Zafar, "Do Female Faculty Influence Female Students' Choice of College Major and Why?," March 2009. Working paper.

Rivkin, Steven G., Erick A. Hanushek, and John F. Kain, "Teachers, Schools, and Academic Achievement," *Econometrica*, March 2005, *73* (2), 417–458.

Rockoff, Jonah, "The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data," *American Economic Review*, May 2004, *94* (2), 247–252.

Scafidi, Benjamin, David L. Sjoquist, and Todd R. Stinebrickner, "Race, Poverty, and Teacher Mobility," *Economics of Education Review*, 2007, *26* (2), 145–159.

Steele, Claude M, "A Threat in the Air: How Stereotypes Shape Intellectual Identity and Performance," *American Psychologist*, June 1997, *52* (6), 613–629.