The Long-Run Impacts of Same-Race Teachers

Seth Gershenson†
American University and IZA

Cassandra M. D. Hart
University of California, Davis

Constance A. Lindsay
American University

Nicholas W. Papageorge
Johns Hopkins University and IZA

ABSTRACT: Black primary-school students matched to a same-race teacher perform better on standardized tests and face more favorable teacher perceptions, yet little is known about the long-run, sustained impacts of student-teacher demographic match. We show that assigning a black male to a black teacher in the third, fourth, or fifth grades significantly reduces the probability that he drops out of high school, particularly among the most economically disadvantaged black males. Exposure to at least one black teacher in grades 3-5 also increases the likelihood that persistently low-income students of both sexes aspire to attend a four-year college. These findings are robust across administrative data from two states and multiple identification strategies, including an instrumental variables strategy that exploits within-school, intertemporal variation in the proportion of black teachers, family fixed-effects models that compare siblings who attended the same school, and the random assignment of students and teachers to classrooms created by the Project STAR class-size reduction experiment.

† Corresponding author: gershens@american.edu. The authors thank Robert Moffitt and conference participants at SREE and AEFP for helpful comments. Stephen B. Holt provided excellent research assistance.

1. Introduction

Educational attainment gaps between socio-demographic groups are well-documented, large, and frustratingly persistent (Bailey & Dynarski 2011). These gaps are troubling because lower rates of educational attainment are associated with a host of negative downstream outcomes, including elevated risks of unemployment or relegation to low-wage work, crime, poor health, civic disengagement, and other negative socio-economic outcomes. Attainment gaps are policy relevant in part because education generates positive externalities, but also because such gaps might reflect sub-optimal investments in human capital by historically disadvantaged groups, such as racial minorities. However, the sources of attainment gaps are poorly understood, which hinders the design and efficacy of policy responses.

Teachers are thought to be among the most important school-provided educational inputs in the K-12 years (Hanushek & Rivkin 2010). In addition to improving test scores, primary school teachers have long-run impacts on earnings and educational attainment (Chetty et al. 2014). However, with a handful of exceptions (e.g., experience (Wiswall 2013)), previous literature has not shown which observable teacher behaviors or characteristics correlate with effectiveness (Chingos & Peterson 2011; Hanushek & Rivkin 2010; Rockoff et al. 2011). This limits the policy relevance of the knowledge that teachers matter and vary widely in quality.

A notable exception is mounting evidence of teacher-student demographic match effects on short-run, immediate outcomes such as test scores, attendance, and suspensions (Dee 2004, 2007; Egalite et al. 2015; Lindsay & Hart 2017; Holt & Gershenson 2015).[1] For example, primary school students, especially those from historically disadvantaged groups such as African Americans, score higher on standardized tests when they are randomly assigned to same-race

[1] A large "teacher like me" literature documents similar effects in post-secondary educational settings as well, including community college (Fairlie et al. 2014), college (Carrell et al. 2010), and law school (Birdsall et al. 2016).

teachers (Dee 2004). This suggests that a teacher's race, which is observable at the time of hiring and school or classroom assignments, is a useful predictor of teachers' abilities to reduce demographic gaps in educational achievement. However, the policy relevance of these findings is limited in that student-teacher demographic match effects have only been shown to have short-run effects on student outcomes such as test scores, suspensions, and attendance, which might fade out over time. In other words, there is no evidence that exposure to a same-race teacher has long-run impacts on educational attainment. Our aim is to fill this gap in the literature.

In this paper, we identify the causal effect of exposure to a same-race teacher in primary school on long-run investments in human capital. Using longitudinal administrative data on all public school students in North Carolina, we show that black students who are as good as randomly assigned to a black teacher at least once in the third, fourth, or fifth grades are more likely to aspire to college and less likely to drop out of high school. The high-school dropout effect is largest among black boys who were subject to persistent poverty throughout their time in primary school, a result that is striking given that this demographic group exhibits consistently low investments in human capital. By demonstrating that poor black male students exhibit higher human capital investments due to quasi-random exposure to at least one black teacher, our findings suggest that stubbornly persistent attainment gaps are not impervious to policy changes.

The main empirical analysis leverages student-level data from all public schools in North Carolina for the cohorts who entered the third grade between 2001 and 2005, which follow students through their senior year of high school. We use an instrumental variables (IV) identification strategy that exploits transitory, conditionally random variation within schools over time in the demographic composition of teachers (e.g., Bettinger & Long 2005, 2010; Fairlie et al. 2014). Conservative estimates suggest that exposure to at least one black teacher in grades 3-5 significantly reduced the probability of dropping out of high school among low-income black males by seven percentage points, or 39%. We find no effect of having a same-race teacher on female students' high-school dropout decisions, perhaps due to females' significantly higher baseline graduation rates. Similarly, regarding postsecondary educational attainment, we find that among persistently-poor students of both sexes, exposure to at least one black teacher in grades 3-5 increased students' self-reported intent to pursue a four-year college degree (at HS graduation) by 0.06, or about 19%. Again, this effect was even larger for males (0.08, or 29%).

Because we observe student-teacher match in three grades, we also investigate whether there are dosage effects of exposure to multiple black teachers. Assignment to three black teachers has a slightly stronger impact on educational attainment than assignment to a single black teacher. However, such differences tend to be small in magnitude and statistically insignificant. Accordingly, our findings suggest that exposure to one black teacher has a meaningful effect on students' long-run outcomes, and that the marginal effect of exposure to a second black teacher is relatively small. The lack of strong dosage effects suggests an important policy implication: the number of black teachers need not be dramatically increased to close racial gaps in educational attainment. Rather, our results suggest that efforts to match black students with at least one black teacher in primary school could begin immediately, by thoughtfully matching students to current teachers.

Given the size of these effects and possible questions about IV's validity, we assess whether effects are similar elsewhere. We confirm that similar patterns exist in Tennessee, using data from the Project STAR experiment, in which students were randomly assigned to different class sizes. An unintended consequence of this randomization was exogenous variation in exposure to same-race teachers (Dee 2004). Leveraging this variation, we show that exposure to a same-race teacher results in higher graduation rates and greater likelihood of taking college entrance examinations. This finding is striking, as the data come from a different state and identification comes from a different source of exogenous variation. Using North Carolina data, treatment effects are identified off of within-school variation over time in the proportion of black teachers. In the case of the Tennessee STAR data, variation comes from the random assignment of students to classrooms. Moreover, rather than self-reported college aspirations, the Tennessee STAR data provides arguably more objective measures of college intent, sitting for an entrance exam such as the SAT or ACT, which is a costly action. In contrast, reports of intentions are nearly costless and may therefore be subject to social desirability bias.

Our findings contribute to several related literatures regarding the determinants of educational attainment, sources of racial gaps in educational attainment, and the mechanisms through which teachers affect long-run socioeconomic outcomes. Previous studies have demonstrated that student-teacher demographic match affects primary school students' short-run outcomes, including academic achievement, attendance, and behavior (Dee 2004; Egalite et al. 2015; Holt & Gershenson 2015), as well as teachers' perceptions of students (Dee 2005; Gershenson et al. 2016; Ouazad 2014). Interest in these short-run outcomes is due to their proximal relationships with long-run socio-economic outcomes such as educational attainment and success in the labor market (Currie & Thomas 1999; Jackson 2012). However, until now, there has been no evidence of long-run effects of student-teacher demographic match. We demonstrate the existence of sustained and economically meaningful long-run impacts, which means that demographic match is an observable factor explaining why some teachers are better than others at increasing racial-minority and low-income students' human capital investments.

This finding also informs our understanding of the mechanisms through which teachers affect student outcomes. In a seminal contribution, Chetty et al. (2014) show that primary school teachers have long-run impacts on earnings and educational attainment, even as their impacts on test scores fade out over time. However, the administrative data analyzed by the authors precludes precise identification of the mechanisms through which such effects operate, which limits the array of policy responses available to schools. Our findings provide a possible explanation and a policy response that could easily be implemented.

Still, mechanisms remain elusive. The current study relates to role model effects, which may reflect differences in perceptions and expectations, both of which have been shown in earlier work. For example, Gershenson et al. 2016 show that black teachers expect more from black students than do white teachers. If this channel is an important one, then there may be an information gap that justifies policy interventions. The reasoning is as follows: If demographic attainment gaps are due to differences in investment decisions made by rational, fully-informed individuals, gaps may be unfortunate, but a role for policy on efficiency grounds may be limited.[2] However, if attainment gaps can be explained in part by the simple fact that socioeconomically-disadvantaged black boys disengage from school due to lack of exposure to same-race, educated role models, the concern is that information gaps or biases in expectations about human capital investments play an outsize role and cause such students to under invest in their human capital. If so, persistent racial gaps in educational attainment reflect sub-optimally low investments in human capital made by uninformed agents. Policies that increase demographic match could thus close racial and socio-economic information gaps, improving

[2] Social costs of attainment gaps, such as increased criminality among people with less education, imply a negative externality and thus suggest an alternative justification for a policy that would narrow achievement gaps.

economic efficiency by leading students to make more informed investments in their human capital.

One possibility is that teachers affect non-cognitive skills not captured by standardized tests (Gershenson 2016; Jackson 2012). For example, some teachers are better than others in instilling or developing a child's work ethic, which may not affect timed, standardized, low-stakes (to the student) tests, but could improve labor market outcomes. However, while this is an appealing explanation of the results in Chetty et al. (2014), it is not obvious that impacts on non-cognitive skills fully explain teachers' long-run impacts on socioeconomic outcomes.

An alternative, perhaps complementary, explanation is that black teachers have higher educational expectations for black students, which in turn increase black students' educational engagement and aspirations (Papageorge et al. 2016). Indeed, the authors show that biases in teacher expectations create self-fulfilling prophecies. Thus, by increasing teachers' expectations, student-teacher racial match could shift the investments made by both teachers and students. Moreover, this type of shift in expectations over long-run investments may not affect test scores, but would affect long-run outcomes, consistent with the findings in Chetty et al. (2014).

By contrast, teachers with low expectations for a student could change how they allocate scarce resources (e.g., their time and effort) to that student. Diverting resources from students could negatively affect their academic trajectories. Moreover, students who perceive that a different-race teacher has low expectations may themselves begin to believe that educational attainment is out of reach or not worth the investment.[3] Similarly, minority students may simply lack role models if they rarely observe or interact with demographically similar individuals who

[3] On the importance of expectations, Fortin, Oreopoulos, and Phipps (2015) find that changes to gender differences in post-secondary expectations are important factors accounting for the emerging trend where girls' high school GPA has risen relative to boys'.

have high educational attainment, which could also lead them to curtail their investments. In each case, the danger is that demographic mismatch generates low expectations that become self-fulfilling prophecies.[4] However, if black students face higher expectations from black teachers, as documented in Gershenson et al. (2016), this may be one channel through which the long-run impacts documented in Chetty et al. (2014) operate.

The paper proceeds as follows: Section 2 describes the data and basic identification strategy. Section 3 presents the main results and several robustness checks. Section 4 replicates the main results using experimental data from Tennessee's Project STAR. Section 5 concludes.

2. Data and Methods

This section describes the data and methods used in the primary analysis of the long-run impact of demographic representation among primary school teachers on educational attainment. Section 2.1 describes the administrative data from North Carolina. Section 2.2 summarizes the main analytic sample. Section 2.3 describes the instrumental-variables identification strategy.

2.1 Data

The main analysis utilizes student-level longitudinal administrative data on all public school students in North Carolina who entered third grade between the 2000-2001 (2001) and 2004-2005 (2005) school-years.[5] Students' educational trajectories are recorded through their senior year of high school. These data are provided by the North Carolina Education Research Data Center (NCERDC). The NCERDC student-level records can be linked to teacher identifiers through testing records, contain information on student and teacher demographics and geocoded

[4] This is related to "stereotype threat", whereby students from disadvantaged groups buy into negative stereotypes (e.g., that they will not finish high school) and shape their behavior and investments accordingly (Steele 1997).
[5] We subsequently refer to school years by the year of the spring term (e.g., 2000-2001 school-year is 2001).

address data that identify likely family groups, and include schooling outcomes such as high-school graduation, drop-out, and self-reported college intent upon high-school graduation.

The NCERDC data enable the estimation of effects of students' exposure to demographically similar teachers, as North Carolina consistently collected data linking students in tested grades to teachers. Because the NCERDC data contain linked data for multiple cohorts, we can exploit within-school changes in the demographic composition of the teaching force over time. Finally, another advantage is that data on the entire population of students provide the statistical power necessary to implement school-specific time trend models and within-family (sibling) comparisons (Figlio et al. 2015), which we exploit in the sensitivity analysis.

Because black students have higher average drop-out rates and lower college attendance rates than their non-black counterparts both nationally and in North Carolina (Murnane 2013), the analysis focuses on black students.[6] The main analytic includes black students who entered third grade for the first time between 2001 and 2005. The sample excludes 19,872 students missing from public school data by 8th grade regardless of cause of absence (e.g., transferring to private schools, leaving the state, etc.). We further exclude an additional 4,394 students who are still observed in the public school system as of 8th grade, but who are then explicitly recorded as having exited the state's public school system for out-of-state schools, private schools, or home schools; or due to death; or who are otherwise excluded from the North Carolina cohort count used to calculate school graduation rates. We exclude 3,698 students whose elementary school teachers' races are missing in all years. Finally, for our main sample, we exclude 14,432 students for whom we lack clear indication of their graduation outcomes; that is, these students are neither recorded as having definitively graduated, nor as having definitively dropped out. We later add

[6] Hispanic students have lower graduation and college attendance rates than whites, but because there are few Hispanic teachers in North Carolina, we cannot test effects of exposure to same-race teachers for this group.

these students back into the sample in a series of robustness checks to determine the sensitivity of our results to different assumptions about the outcomes of those students.

2.2 Summary Statistics

Table 1 summarizes the analytic sample of black students overall, by sex, and for the subset of students who persistently received free- or reduced-price lunch (FRL) in each year observed in grades 3-8. We use this measure of disadvantage because Michelmore and Dynarski (2016) show that persistently-FRL students fare significantly worse on a variety of educational outcomes than their counterparts whose FRL status changes over time.

Panel A of Table 1 summarizes students' educational outcomes. The NCERDC data contain two "long run" measures associated with educational attainment, which serve as the dependent variables in the main analyses. The first is an indicator for whether students are ever observed as dropping out of high school in NCERDC records. The state counts students as dropping out of school in a particular year if they are not enrolled in North Carolina public schools by the 20th day of instruction, after having attended in the previous year and without having graduated from a North Carolina school (NCERDC, 2012). Roughly 13% of students are recorded as having dropped out of high school in the overall sample. The second is an indicator for whether the student reports plans to attend a four-year college or university. This variable is collected only for students who are recorded as graduating from a North Carolina public high school. A value of zero on this variable indicates that the student either declared no intention of attending a four-year college or did not graduate from high school. Roughly 40% of the sample graduated from high school and intended to attend a four-year school; the remaining 47% of the sample graduated from high school but did not plan to attend a four-year post-secondary institution. Consistent with patterns in national data (Bailey & Dynarski 2011), panel A shows that educational attainment in North Carolina is lower for males than females, and lower for socioeconomically disadvantaged students than for their more advantaged counterparts.

Panel B of Table 1 summarizes the primary independent variables of interest, which are likely endogenous. The baseline model uses a simple binary indicator of exposure to a black teacher, which equals one if any of the student's teachers in grades 3-5 were black, and zero otherwise.[7] There is substantial variation in this variable, as about 40 to 45% of black students in these cohorts had at least one black teacher in grades 3-5. This reflects the fact that the majority of public school teachers are white, but that black teachers are more likely to teach in majority-black schools (Hanushek et al. 2004; Jackson 2009). We also report the average count of black teachers experienced by students in grades 3-5, which we use to test for dosage effects. On average, the black students in the analytic sample had about 0.6 black teachers, though this is pulled down by the majority of students who had zero black teachers. Conditional on having at least one black teacher, the average is about 1.4, which suggests that having multiple black teachers in grades 3-5 is relatively rare.

Figure 1 plots histograms that fully explore the distribution of students' exposure to black teachers in grades 3-5, separately by gender and SES. The two histograms are qualitatively similar. About half of students had no exposure to black teachers in grades 3-5, 30% had one, 15% had two, and 6% had three. Figure 2 plots similar histograms that examine the specific timing, or sequencing, of students' exposure to black teachers in grades 3-5. In the figure, zeroes

[7] The NCERDC data matches elementary-school teachers to specific classes starting in 2007, including to self-contained classes, but the long-run educational outcomes of interest in the current study have yet to be realized for these cohorts. These data are not uniformly available for the earlier cohorts who comprise the analytic sample in the current study. We therefore link students to teachers using administrative records of who administered the student's End-of-Grade exams, which are only taken in grades 3-5. While this approach imperfectly identifies students' teachers, it performs quite well: using more recent years of data we can verify that about 85% of exam proctors were in fact the actual self-contained classroom teacher (authors' calculations).

indicate non-black teachers and ones indicate black teachers. Thus, a sequence of 101 indicates that a student had black teachers in grade 3 and 5, but not in grade 4. These figures suggest that black teachers are slightly more common in grades 4 and 5 than in grade 3.

Panel C of Table 1 summarizes the instrument used in 2SLS estimation of the baseline model. The instrument is cohort and school specific. For the cohort entering a school's third grade in year *t*, the instrument measures the share of teachers who were black in grade 3 in year *t*, grade 4 in year *t*+1, and grade 5 in year *t*+2. That is, the instrument measures the fraction of black teachers in each of grades 3, 4, and 5, for each school-cohort, weighting each grade equally. Intuitively, this represents the share of black teachers a student would be expected to encounter in grades 3-5.[8] Within-school standard deviations of the instrument show that about 37% of the sample variation in the instrument occurs within, as opposed to between, schools. This suggests that there is sufficient identifying variation in the instrument to estimate models that condition on school fixed effects. The identifying variation in the instrument is also depicted in Figure 3, which plots the histogram of school-by-cohort deviations in the fraction of teachers who are black from the long-run school average. The modal deviation is zero, which is largely due to a number of schools that had no black teachers in grades 3-5 for during the sample's time period. However, more than half of school years exhibit deviations from the school's mean, again suggesting that there is sufficient identifying variation in the instrument.

Finally, panel D provides some basic background information on the students. Overall, almost half were persistently-low income (i.e., used FRL in all years observed in grades 3-8) and more than 85% used FRL at least once. About 10% were classified as exceptional learners in at

[8] The instrument can be disaggregated into grade-specific fractions, which we exploit in the timing models presented in Section 3.2.

least one year, which generally means that the student had an IEP or learning disability. Exceptionality rates are slightly higher for males and persistently-FRL students.

Though omitted from Table 1 in the interest of brevity, a number of standard school characteristics are observed and included in the baseline model. Time-varying school controls are defined for the school-year in which the student entered the third grade. They include the pupil-to-teacher ratio, (logged) total enrollment in the school, share of students using subsidized lunch, share of students black, Hispanic, or other race, geographic locale indicators (urban, town or rural, and suburban as omitted reference group), and a measure of the school's performance on the end-of-grade exams, which averages the grade-year standardized performance of students in math and reading for students across all tested grades in the school.

2.3 Identification Strategy

The goal of this paper is to estimate the impact of exposure to same-race teachers during primary school on black students' long-run educational attainment. Accordingly, interest is in obtaining consistent estimates of γ in linear models of the form

$$y_{ist} = \beta X_i + \theta_s + \delta_t + \gamma SameRace_{ist} + \varepsilon_{ist}, \tag{1}$$

where i, s, and t index students, primary schools (in third grade), and cohorts (year of entry into third grade), respectively; X is a vector of observed student characteristics; θ and δ are school and cohort fixed effects (FE), respectively; *SameRace* is a measure of students' exposure to same-race teachers in grades 3-5; and ε is an idiosyncratic error term.[9]

School and cohort FE control for sorting across schools and statewide secular trends in teacher demographics that might affect black students' educational attainment. The baseline

[9] All models include missing-variable indicators for all right-hand side variables except the teacher-student demographic match variables to preserve full information for cases with partially missing data.

model codes *SameRace* as a binary indicator that equals one if student-*i* was assigned to a black teacher in any of grades 3-5, and zero otherwise. We then test for dosage and timing effects by using alternative definitions of *SameRace*. For example, dosage effects are estimated using the number of student *i*'s classrooms that were headed by black teachers in grades 3-5, and timing effects are captured by grade-specific indicators for having had a same-race teacher in that grade.

OLS estimates of equation (1) are likely biased by unobserved student characteristics that jointly predict classroom assignments and long-run outcomes, even after conditioning on the basic socio-demographic controls in *X* and school FE (Rothstein 2010). For example, students with lower achievement (Clotfelter, Ladd & Vigdor, 2006) and greater exposure to school discipline (Lindsay & Hart, 2017) are more likely to be matched to black teachers, and these factors likely affect long-run outcomes as well. More generally, myriad unobserved factors (e.g., parental involvement) might jointly influence classroom assignments and educational attainment (Dieterle et al. 2015). Therefore, we estimate equation (1) by 2SLS using an instrumental-variables (IV) strategy similar to that pioneered by Bettinger and Long (2005, 2010).

Specifically, our preferred IV identification strategy exploits transitory (conditionally random) variation within schools over time in the demographic composition of the teaching staff. Identifying variation comes from the fact that cohorts who enter the third grade in a particular school in different years have different propensities to be assigned to same-race teachers, because teachers frequently go on leave, change schools, and even change grades within a school (Brummet et al. 2017). Conditional on school FE, transitory changes in the demographic composition of schools' teaching staffs leads to deviations from the "steady state" fraction of grades 3-5 teachers who are black, which are arguably exogenous (i.e., are excluded from equation (1)). The reason is that, net of baseline school quality and trends in school demographics, teacher entries and exits are likely driven by exogenous, idiosyncratic factors such as enrollment changes, parental leaves, and retirements. Similar IV strategies have been used to investigate the impact of faculty representation on graduation rates in community colleges (Fairlie et al. 2014).

We use the school-cohort specific racial composition of teachers in grades 3-5 (Z_{st}) to instrument for the student-specific *SameRace* variable in 2SLS estimation of equation (1). For each cohort of incoming third graders at a particular school, we calculate the share of teachers who are black among grade 3-5 teachers that a student would potentially encounter if they progressed through grades 3, 4, and 5 in the same school in the course of three academic years (i.e., if they neither change schools, repeat grades, nor skip grades). The baseline instrument Z_{st} measures the fraction of black teachers in each of grades 3, 4, and 5, for each school-cohort, weighting each grade equally. We do not require students to have actually stayed in the same school for grades 3-5, as this would induce endogenous sample selection. Nonetheless, only about 12% of students in the analytic sample changed schools in grades 3-5, and the only consequence of including these students is a potential weakening of the first-stage relationship.

3. Main Results

This section presents the main empirical results. Section 3.1 presents baseline estimates of equation (1), which document the long-run impact of having at least one black teacher in grades 3-5 on black students' educational attainment and aspirations. Section 3.2 tests for dosage and timing effects of exposure to black teachers in grades 3-5. Section 3.3 presents several sensitivity analyses that probe the robustness of the main results.

3.1 Main Results

Table 2 reports 2SLS estimates of the baseline model specified in equation (1). Column 1 estimates the model for the full analytic sample of black students who entered the third grade in North Carolina public schools between 2001 and 2005. Panel A of column 1 shows that the first stage is strong. Specifically, the first-stage coefficient estimate of 0.71 means that a ten percentage point increase in the share of a school's teachers who are black, on average, increases the probability that the student is assigned to at least one black teacher in grades 3-5 by about seven percentage points. Panel B of column 1 shows that having had at least one black teacher in grades 3-5 significantly reduced the black high school dropout rate by about four percentage points, which corresponds to an economically meaningful 31% decline. Panel C of column 1 finds that on average, having had at least one black teacher in grades 3-5 increased the probability that black high school graduates expected to attend a four-year college or university by about three percentage points (7%), though this increase is not statistically significant.

Of course, pooling male and female students in the same regression might obfuscate the actual long-run impacts of having a same-race elementary school teacher if males and females react differently to this educational input. Male students might be more responsive, as Table 1 showed that they have higher high school dropout rates than females, they are often less engaged and have lower educational expectations than females (Fortin et al. 2015; Steele 1997), and educational interventions and inputs frequently have larger effects on males than females (Carrell & Hoekstra 2014; Chetty et al. 2016; Dynarski et al. 2013; Figlio et al. 2016). Accordingly, Columns 2 and 3 of Table 2 estimate the baseline model separately for male and female students, respectively. The first stages are again strong. However, the long-run impact of having at least one same-race elementary school teacher on the probability of dropping out of high school is entirely driven by the behavior of black male students. On average, having had at least one black teacher in grades 3-5 reduces males' dropout probability by about eight percentage points, effectively halving the black male dropout rate. There is no such effect on black female students. This is consistent with role model effects being especially important for the most at-risk students.

There could also be heterogeneity by SES in students' long-run responses to exposure to same-race elementary school teachers, as students from low-income and less-educated households might disproportionately benefit from same-race role models in school due to their lower levels of exposure to, and confidence interacting with, college-educated individuals outside of the traditional school day (Dillon & Smith 2017; Lareau 1987; Lareau & Weininger 2003). Accordingly, Columns 4-6 of Table 2 re-estimate the baseline model, overall and by sex, for the subsample of black students who persistently received for free/reduced lunch (FRL) in each of grades 3-8. We focus on this particularly vulnerable subsample because Michelmore and Dynarski (2016) show that such students are substantially worse off, and lower-performing, than their counterparts who move in and out of FRL during their K-8 education.[10]

Panel A of Table 2 shows that the first-stage is strong for the persistently-FRL subsamples. Consistent with the full-sample results reported in columns 1-3, columns 4-6 of Panel B show that having had at least one black teacher in grades 3-5 significantly reduced the dropout rate of persistently-FRL black males, but has no effect on persistently-FRL black females. Consistent with the hypothesis that persistently-FRL students might be particularly responsive to role-model effects, the point estimate in column 5 is 50 percent larger than its counterpart in column 2. Specifically, among persistently-FRL black male students, the point estimate suggests that having a same-race teacher at least once in grades 3-5 reduces the likelihood of high-school dropout by about 12 percentage points. Finally, columns 4-6 of Panel

[10] In Section 3.3 we investigate the sensitivity of this analysis to alternative definitions of disadvantage.

C document a statistically significant effect of having had at least one black teacher in grades 3-5 on persistently-FRL high-school graduates' postsecondary expectations. This effect of about ten percentage points (30%) is similar in magnitude for both males and females and is in contrast to the null effect in the full sample. Again, the long-run impact of same-race elementary school teachers is substantially larger among the most socioeconomically disadvantaged black students.

The linear 2SLS estimates of equation (1) presented in Table 2 are our preferred baseline estimates and model for two reasons. First, the linear model allows for the inclusion of school fixed effects, and later, school-specific linear time trends. Second, the linear model is easily augmented to allow for nonlinear dosage and timing effects of exposure to multiple black teachers, as described in section 3.2. Despite the binary nature of both the dependent and endogenous variables, 2SLS estimates frequently provide good approximations of the underlying average partial effects (APE) of interest (Angrist & Pischke 2009; Wooldridge 2010).

Nonetheless, two limitations of the linear model suggest that a system of nonlinear probit models might provide more efficient estimates of the APE of interest. The first is that by treating the dependent variable as dichotomous, we necessarily estimate the dropout and college-aspiration models separately. The second is the usual concern over the linear probability model's constant partial effects and unconstrained predicted probabilities.

We address these concerns by specifying and estimating a two-equation system containing a probit and an ordered-probit model that have correlated errors. This system is analogous to the usual bivariate probit model used in the case of a binary dependent and endogenous variable (Wooldridge 2010).[11] The RHS of equation (1) constitutes the linear index of the ordered-probit model, where the outcome is an ordinal variable that takes one of three values: high school dropout, high-school graduate with no college intent, and high-school

[11] In Appendix Table A1 we report the "usual" bi-probit analogs to the baseline linear models estimated in Table 2.

graduate with college intent. The probit model is analogous to the first-stage regression in 2SLS estimation, which includes as a regressor the instrument that is excluded from the ordered-probit model. The two equations are jointly estimated as a mixed-process model (Roodman 2011).[12]

Estimates of the joint probit-ordered probit models, which are quite consistent with the baseline linear estimates in Table 2, are reported in Table 3. Panel A of Table 3 reports the "first stage" probit coefficients on the excluded instrument. The coefficient estimates are large, positive, and strongly statistically significant. Panel B of Table 3 reports the coefficient on the variable of interest, exposure to at least one black teacher, in the ordered-probit model. Both overall and for each subsample, the point estimate is of the expected sign and statistically significant. However, the magnitude of the ordered-probit coefficient is not directly interpretable, nor can it be directly compared to the baseline linear estimates.

Rather, APE for each of the three ordinal outcomes are reported in panel C. The APE of exposure to a same-race teacher on the probabilities of dropping out and of intending to attend college *are* comparable to the linear estimates in panels B and C of Table 2. Consistent with the baseline estimates in column 1 of Table 2, column 1 of Table 3 finds small, negative effects on the likelihood of dropping out and small, positive effects on the probability of stating an intent to attend college. However, these APE are notably more precisely estimated than the linear 2SLS estimates, as expected. Columns 2 and 3 of Table 3 show that the effects are larger for males, particularly persistently-FRL males. Again, these estimates are substantially more precisely estimated than the linear 2SLS estimates, and are slightly smaller in magnitude.[13] For example, column 3 shows that a low-income black male is seven percentage points less likely to drop out,

[12] School FE are implemented by including a full set of school dummies. The incidental parameters bias from doing so is likely minimal, as there are many (> 20) students per school (Greene 2004).
[13] The APE tend to be close to the bottom end of the corresponding linear 2SLS 95% confidence intervals.

and eight percentage points more likely to state an intent to attend college. These are smaller, but similar in magnitude, to the corresponding linear estimates of -0.12 and 0.10, respectively.

3.2 Dosage and Timing

The baseline estimates presented in Section 3.1 provide arguably causal evidence that exposure to at least one black teacher in grades 3-5 significantly increased educational attainment, particularly among males and the most socioeconomically disadvantaged students. This raises a natural, policy-relevant question: is there a marginal benefit of having a second (or third) black teacher in grades 3-5? This is an important question given the limited supply of black teachers in U.S. public schools and in the teacher-training pipeline (Goldring et al. 2013; Putman et al. 2016). Accordingly, we estimate "dosage model" specifications of equation (1) in which *SameRace* is coded as a vector of mutually exclusive indicators for the number of black teachers the student had in grades 3-5.[14]

Specifically, we estimate

$$y_{ist} = \beta X_i + \theta_s + \delta_t + \sum_{k=1}^{3} \gamma_k 1\{w_i = k\} + \varepsilon_{ist}, \tag{2}$$

where w in {0, 1, 2, 3} is the number of black teachers experienced by student i in grades 3-5 and $1\{\cdot\}$ is the indicator function. The omitted reference category is 0, which means that equation (2) contains three endogenous variables that must be instrumented (the indicators for having had 1, 2, and 3 black teachers). We obtain the identifying instruments by estimating a poisson

[14] This model can only be estimated for the subsample of students whose teachers' races were observed in each of grades 3, 4, and 5. The "dosage" analytic sample is summarized in Appendix Table A2. Along most dimensions, the dosage sample closely resembles the full analytic sample summarized in Table 1 of the main text. The largest difference is in exposure to black teachers, as students in the dosage sample, on average, are about 5 percentage points more likely to have ever had a black teacher than those in the full analytic sample. To assuage concerns that the dosage results are biased by endogenous sample selection, we re-estimate the baseline model (equation 1) on the selected dosage sample and present these estimates in Appendix Table A3. It is reassuring, then, that these point estimates are remarkably similar to those for the full sample, reported in Table 2 of the main text.

regression of w on the school-cohort specific racial composition of teachers in grades 3-5 (z_{st}) and the other covariates on the RHS of equation (2), where z was the sole instrument used in the baseline 2SLS estimation of equation (1). Then, we use the poisson estimates to predict the probabilities (p_1, p_2, p_3) that student i experienced 1, 2, and 3 black teachers, respectively.[15] Finally, equation (2) is estimated by 2SLS using p_1, p_2, and p_3 as instruments. This is a variant of the two-step IV method for dealing with endogenous, binary independent variables (Wooldridge 2010, p. 939; Angrist & Pischke 2009, p. 191).

Estimates of the dosage model (equation 2) are presented in Table 4. The large (> 10) F statistics in panel A show that the first-stage relationships are strong (Stock & Yogo 2002). Panel B reports estimated long-run dosage effects on the probability of dropping out of high school. Consistent with the baseline results reported in Table 2, dosage effects tend to be larger for males, particularly persistently-FRL males. Column 2 of panel B shows that for black males, the long-run effect of having a same-race teacher is monotonically increasing in the "dose" (number) of black teachers experienced in grades 3-5. In column 5, analogous estimates for the persistently-FRL male subsample tend to be slightly larger in magnitude, though the indicator for two black teachers is imprecisely estimated. Both for all males and for persistently-FRL males, the dosage indicators are strongly jointly significant, again indicating that there are significant long-run impacts of having a same-race elementary school teacher on black males' propensity to complete high school.

For black males, the effect of having three consecutive black teachers in grades 3-5 is larger than having either one or two black teachers in this grade span. However, this difference is only statistically significant when comparing the effect of having had one to having had three

[15] Augmenting this procedure to include a cubic in z in the poisson regression yields qualitatively similar results. See Appendix Table A4. This suggests that identification is not purely due to nonlinearities in the poisson model. Estimates of the initial poisson regression are presented in Appendix Table A5.

black teachers in the full sample of black males (column 2). Nonetheless, the point estimates of the long-run impacts of having three black teachers in grades 3-5 on black males' dropout rates are large in magnitude (relative to the baseline high-school dropout rate), which raises the question of whether having three black teachers in grades 3-5 is indicative of other school features, such as schools with predominantly black teaching staffs, that may be less common among students who encounter only one or two black teachers in grades 3-5. Unfortunately, the administrative data for the cohorts under study do not identify students' K-2 teachers.

To understand whether having three consecutive black teachers in grades 3-5 strongly predicts also having three black teachers in K-2, we conduct an out-of-sample analysis of two recent cohorts from North Carolina's public schools: students entering kindergarten in 2007 and 2008. These cohorts allow us to observe all K-5 student-teacher matches, but are not usable for our main research question, as students in these cohorts have yet to graduate. The crosstabs in Appendix Table A6 show that for students in this more recent sample, the more black teachers a student encounters in grades 3-5, the more likely the student is to have also had black teachers in grades K-2. For instance, among the full sample of black students (Column 1) with zero black teachers in grades 3-5, about one quarter of students had previously been matched to a black teacher in grades K-2. By contrast, among students with three black teachers in grades 3-5, nearly two-thirds had also been matched to at least one black teacher in their K-2 years. This trend likely reflects school segregation. There are no gender differences in these patterns.

To further investigate this question within the main analytic sample, Appendix Table A7 uses personnel files that identify the self-contained teachers in grades K-5 in the years our sample cohorts attended 3rd grade, for the school at which they attended 3rd grade.[16] We compare

[16] Our data allows us to see the teachers that served in these roles during the years covered by our sample, but again does not allow us to match the classroom-level records to individual students.

the share of teachers serving in each grade who were black for students who had 0, 1, 2, or 3 black teachers in grades 3-5. As might be expected, the teaching forces in schools serving students who have a higher number of black teachers in grades 3-5 also feature higher concentrations of black teachers in lower grades as well. For instance, among students who encounter one black teacher in grades 3-5, the teaching forces for grades K-2 in the schools where they attend 3rd grade are roughly 20-25% black. By contrast, for students who encounter 3 black teachers in grades 2-3, the teaching forces in K-2 are about 50-60% black. Together, Appendix Tables A6 and A7 suggest that exposure to black teachers in grades 3-5 is positively correlated with exposure to black teachers in grades K-2. However, the lack of systematic dosage effects suggest that this data limitation does not confound the main results. Rather, this correlation coupled with the lack of strong dosage effects further reduces concerns about the mismeasurement of students' total elementary school exposure to black teachers.

A related question is whether the grade(s) in which black students are exposed to a black teacher matters. This question is policy-relevant, as it has implications for the optimal grade-level assignments of black teachers, who are in short supply. Accordingly, we estimate "timing model" variants of equation (1) in which *SameRace* is coded as a vector of indicators for whether the student experienced a black teacher in each of grades 3-5.[17] Specifically, we estimate

$$y_{ist} = \beta X_i + \theta_s + \delta_t + \sum_{g=3}^{5} \lambda_g r_{ig} + \varepsilon_{ist}, \tag{3}$$

where r_{ig} is a binary indicator equal to one if student *i* had a black teacher in grade *g*, and zero otherwise. Thus there are three grade-specific endogenous variables that need to be instrumented

[17] Like the dosage model, this timing model can only be estimated for the subsample of students whose teachers' races were observed in each of grades 3, 4, and 5.

(r_3, r_4, r_5). We use the school-cohort-grade specific racial composition of teachers in grade g to instrument r_{ig}, so the model is just-identified. Estimates of equation (3) are reported in Table 5.

Panel A of Table 5 shows that the grade-specific first-stage relationships are strong. Consistent with the main results presented in Table 2, Panel B of Table 5 shows that the long-run impact of having a same-race elementary school teacher is larger for males than females, and for low-income students than their more advantaged counterparts. However, we detect no significant differences by grade in the impact of having a same-race teacher. The estimates in Panel C, for the probability of stating college intent, are even less precisely estimated. In fact, these indicators are not jointly significant for any of the groups estimated.

The lack of strong grade-specific differences in Table 5 is perhaps unsurprising, as the analysis is underpowered, and the grade-3 education production function is arguably quite similar to that in grades 4 and 5. Another limitation of equation (3) is that it fails to incorporate potential dosage effects. Indeed, it is intuitively appealing to combine the timing and dosage elements in a single "sequencing" model in which *SameRace* is coded as a vector of mutually exclusive "sequence indicators" (e.g., same, other, same). There are eight potential sequences, with zero black teachers (i.e., other, other, other) serving as the omitted reference group.[18] Thus, the sequence model contains seven endogenous variables, so 2SLS estimation requires seven instruments. We obtain the necessary instruments by interacting the three measures of school-cohort-grade racial compositions of teachers that were used as instruments for equation (3).[19] Intuitively, this approach amounts to using the probability of observing a sequence as an instrument. For example, let z_g be the fraction of black teachers in grade g: the probability of, and instrument for, observing a sequence of same, other, other would be $z_3(1-z_4)(1-z_5)$.

[18] They are 000, 100, 010, 001, 110, 101, 011, and 111, where 1 indicates a same-race and 0 indicates other-race.
[19] They are z_3, z_4, z_5, z_3z_4, z_3z_5, z_4z_5, and $z_3z_4z_5$, where z_g are cohort-school-grade-specific shares of black teachers.

Estimates of this sequencing specification are reported in Table 6. Results are only reported for the high-school dropout outcome, as the dosage and timing models failed to precisely identify effects on college intent. Panel A provides F statistics for the joint significance of excluded instruments, and shows that the first stages are strong. Consistent with the baseline estimates reported in Table 2, Panel B of Table 6 shows that the seven sequence indicators are strongly jointly significant for males, but not for females. Like in Table 5, the point estimates in Table 6 provide modest evidence of dosage effects. However, the difference between the 100 and 001 sequences are not statistically significant. That is, conditional on having precisely one black teacher in grades 3-5, whether the exposure occurs in grade 3 or grade 5 does not matter.

3.3 Sensitivity Analyses

Having documented arguably causal long-run effects of having a same-race elementary school teacher on high-school graduation and college intent, we now probe the robustness of these results. A consistent, striking pattern thus far is that the long-run impacts of same-race teachers are concentrated among the most socioeconomically disadvantaged students. Disadvantage has been defined by persistent receipt of free or reduced price lunch (FRL) in grades 3-8, as in Michelmore and Dynarski (2016). Much previous research in education has used a student's FRL status in a given year to measure SES. But this is likely a weaker indicator of SES, as a subset of students move into, and out of, FRL eligibility each year, and by definition these students are less disadvantaged than their persistently-FRL counterparts.

Accordingly, in Table 7, we re-estimate the baseline model (equation 1) using two alternative definitions of disadvantage. Columns 1-3 identify students who were *ever* FRL-enrolled in grades 3-8, among whom persistently-FRL students comprise a subset. Consistent with the baseline estimates in columns 4-6 of Table 2, panel B of Table 7 shows that exposure to a same-race teacher significantly reduces the likelihood of ever-FRL students failing to complete high school. Once again, this effect is almost entirely concentrated among black males. The point estimates are slightly smaller, though similar in magnitude, to those for persistently-FRL students. Panel C of Table 7, however, finds no significant effect of having a same-race teacher on ever-FRL students' intent to complete a four-year college degree. Here, the ever-FRL students more closely resemble the never-FRL students than the persistently-FRL students, which highlights the importance of measuring the degree of socioeconomic disadvantage and the potential pitfalls of relying on a crude, transitory measure of SES like FRL in a given year (Michelmore & Dynarski 2016). Columns 4-6 of Table 7 define disadvantage as persistent FRL receipt in grades 3-5. This is a stronger definition of disadvantage than the ever-FRL definition used in columns 1-3, but is weaker than the persistently-FRL in grades 3-8 baseline definition. Unsurprisingly, these estimates closely resemble the baseline estimates. Taken together, the results in Table 6 confirm two general points. First, the long-run impact of same-race teachers on black students' educational attainment, particularly on the H.S. completion margin, is stronger among low-SES students regardless of how SES is measured. Second, they reinforce the importance of accounting for SES-gradients in students' responsiveness to interventions.

Table 8 presents a potpourri of sensitivity analyses. The preferred baseline estimates are reported in panel A, to serve as a benchmark. Panels B and C investigate the validity of the preferred IV strategy. The 2SLS estimates reported in panel B use a cubic in the baseline (scalar) instrument in the first-stage regression. The cubic terms are strongly jointly significant, and the resulting 2SLS estimates are not substantially different from those generated by the baseline (linear) model. The robustness of the results to the specification of the first stage is consistent with the instrument being valid (Dieterle & Snell 2016). Panel C augments the baseline model specified in equation (1) to condition on school-specific linear time trends. This model relaxes the concern that identifying variation in the instrument (within-school changes in the demographic composition of the faculty) is correlated with long-term, unobserved trends in school or student quality and composition. Specifically, the estimates in the school-trends model are identified by deviations from trend in the racial composition of the teaching faculty. It is reassuring, then, that these point estimates are quite similar to the baseline estimates reported in panel A, and estimates for the low-income samples remain statistically significant at $p<.10$, even with the reduction in statistical power. The similarity between the baseline (school FE) and school-trends estimates reinforces our choice of the former as our preferred specification, as it provides more precise estimates.

Panels D and E of Table 8 probe the robustness of the main results to the choice of analytic sample. Recall that the H.S. dropout outcome is coded as a one if there is an administrative record of the student formally dropping out, and a zero if there is an administrative record of the student graduating; students who exited the longitudinal data with no formal record of dropping out, nor of graduating, are excluded from the baseline sample. This decision raises two potential concerns. First, many of these students might have actually dropped out, but the administrative records are missing. Second, these students might have exited North Carolina's public school system, perhaps in part due to the race of their grade 3-5 teachers. This would create a sample-selection bias. To examine the practical importance of such concerns, in panel D we re-estimate the baseline model using an enlarged analytic sample that includes the students who exited the data. We do so twice, initially coding the "exiters" as graduates and again coding the "exiters" as dropouts. The former is arguably more accurate, though both specifications are plagued by potentially non-random measurement error. Both sets of estimates are qualitatively similar to those generated by the main analytic sample, which is reassuring, although the arguably more accurate coding of “exiters” as high school dropouts are slightly smaller in magnitude and are less precisely estimated.

The last panel of Table 8, panel E, estimates the baseline model on three additional samples. The first sample excludes students who move during grades 3-5; because our instrument is based on the teaching force that students would face if they stayed at the same school, we mis-measure the teaching force encountered by these students. Excluding these students makes little difference to our estimates; results are broadly similar to the main results.[20]

The second sample change expands our sample by a year in both directions. We have both graduation and drop-out data for the years 2009-2015.[21] 2009 is the expected graduation year for students who enter in 2000, and 2015 is the expected graduation year for students who enter in 2006. Restricting our sample to cohorts that started in 2001-2005 allowed us to capture accurate graduation outcomes for students who may have graduated a year ahead or behind schedule. Relaxing this restriction in favor of a larger sample of students results in a similar pattern of results, although the coefficients are somewhat smaller in magnitude.

Table 8 concludes by reporting estimates for the sample of non-black students. Importantly from a policy perspective, non-black students are not harmed by exposure to black teachers. The point estimates actually tend to be the same sign as those for black students, though they are quite small in magnitude and statistically indistinguishable from zero.

[20] The main analytic sample includes these students because omitting them would create a sample selection bias if the moves were caused by teacher race.

[21] Technically, drop-out data is available from 2004, before the graduation data becomes available, so that we can capture students who drop out prior to 2009. 2009 is the expected graduation year for the cohort of students who enter in 2000.

Finally, we augment equation (1) to control for family-school fixed effects. This robustness check exploits within-family, between-sibling variation in exposure to black teachers in grades 3-5, among siblings who attended grade 3 in the same school. This is an appealing strategy, as the identifying variation comes from comparisons of siblings who reside in the same household, and are therefore arguably exposed to the same household, parental, and neighborhood inputs (Aaronson 1998; Garces, Currie, & Thomas 2002). While family FE purge a potentially important source of bias from the error term in equation (1), the threat of within-family, student-specific variation in unobserved confounders remains (Bound & Solon 1999). Accordingly, we estimate the augmented family-FE model using the preferred IV strategy.

We identify presumptive siblings using address identifiers geocoded by the NCERDC. Because students could share an address without being related—for instance, if students reside in a multi-unit apartment building—we impose a series of restrictions in order to count two students as siblings. To be presumed as siblings, two students must: (i) be of the same race, (ii) never be observed living in different addresses from each other in the same school-year, and (iii) have either been observed in the same school-year at a shared address that housed 5 or fewer children (we assume that buildings with larger numbers of students are usually apartments), or (iv) have been matched to a mutual sibling for whom conditions (i-iii) held. Roughly 95% of students in the main analytic sample could be assigned to family groups under these conditions; however, only about 20% of students could be matched to siblings. Students may lack sibling matches either because they are only children, or because their siblings are in cohorts not covered in our main sample.[22] This highlights the biggest limitation of the family-FE approach: the lack of power, due to both the smaller sample size and the inclusion of the FE themselves.

[22] About 42% of students are matched to siblings when using data on a larger range of available cohorts (not just 2001-2005).

Estimates of the family-FE specifications are reported in Table 9. Columns 1 and 2 report estimates of the baseline IV model for the full and siblings samples, respectively, and column 3 reports IV estimates adding family-school fixed effects. Consistent with all previous specifications, panel A shows that the first stages are strong, even conditional on family-school FE. Consistent with the main results, column 3 of panel B shows that the negative, statistically significant effect of having a same-race teacher on high school dropout rates is robust to conditioning on family FE. The family-FE estimate in panel C of column 3 shows that, like in the baseline model, exposure to at least one black teacher in grades 3-5 has a modest, positive, statistically insignificant effect on black students' college intent.

Columns 4-6 of Table 9 do the same for the subsample of persistently low-income black students.[23] Once again, panel A documents a strong first-stage for this subsample. Similarly, panel B shows that the impact of exposure to a black teacher on the high-school dropout decision is stronger among low-SES black students, even when making within-family comparisons of siblings. Finally, column 6 of panel C shows that the long-run effects of same-race teachers on low-SES black students' college intent is robust to conditioning on family FE: the point estimate of 0.11 for low-income students (Column 6), while imprecisely estimated, is nearly identical to the baseline estimate in column 4.

Together, the results presented in Table 9 lend further credibility to a causal interpretation of the baseline IV estimates reported in Tables 2 and 3: exposure to at least one black teacher in grades 3-5 is associated with long-run, arguably causal impacts on educational attainment. This is true even when exploiting within-family, between-sibling variation in exposure to same-race teachers due to transitory variation in the racial composition of schools' teaching staffs.

[23] We do not stratify the sample by sex due to the lack of power in the "siblings sample."

4. Project STAR Experimental Data

This section examines the robustness of the results presented in section 3 to using administrative data set from a different state and a different identification strategy that exploits the random assignment of students and teachers to classrooms. Section 4.1 provides background on field experiments, including Project STAR, in public education. Section 4.2 describes the public-use Project STAR data and methods used to estimate the long-run impacts of same-race teachers on students' educational attainment. Section 4.3 presents the results.

4.1 Field Experiments in Education

Longstanding debates about the impacts of school-provided educational inputs such as class size and teacher quality center on the correct specification of the education production function (Hanushek 1979; Rothstein 2010; Todd & Wolpin 2003). Concerns about the correct specification have been mitigated in recent years by an influx of randomized experiments that create exogenous variation in students' exposure to educational inputs and interventions (e.g., Araujo et al. 2016; Banerjee et al. 2007; Fryer 2014; Kane & Staiger 2008; Muralidharan & Sundararaman 2011). Tennessee's Project STAR (Student Teacher Achievement Ratio) was a seminal field experiment in the education context, designed to identify the impact of class size on student achievement (Krueger 1999).

Project STAR began in 1986, when it randomly assigned kindergarten students and teachers to either small- or regular-sized classrooms. Over the next three years, the experiment continued by randomly assigning students from the 1986 kindergarten cohort to small and regular classrooms in grades 1-3, and by refreshing the sample in each year. Krueger (1999) shows that small classes significantly improved student performance on standardized tests, particularly among racial-minority and low-income students. Follow-up studies document long-run effects of random assignment to a small classroom on the likelihood of taking a college entrance exam (i.e., ACT or SAT) (Krueger & Whitmore 2001) and on the likelihood of college completion (Dynarski et al. 2013). Once again, these effects are larger for black students.

While the aforementioned analyses of the Project STAR experiment have influenced debates over the efficacy of class-size reductions, Dee (2004) recognized that the random assignment of teachers and students to classrooms created exogenous variation in exposure to another educational input: having a same-race teacher. Dee (2004) leverages this variation to estimate the impact of having a same-race teacher on test scores, and finds significant effects of racial match on both math and reading scores that are largest among black students.

Chetty et al. (2011) similarly leverage Project STAR's randomization to estimate long-run effects of specific kindergarten classrooms on earnings. However, the extant literature that exploits the Project STAR randomization to estimate short- and long-run effects of class size, and to estimate the short-run effects of having a same-race teacher, has yet to leverage this variation to estimate long-run impacts of having a same-race primary school teacher on educational attainment.[24] In section 4.2 we fill this gap in the literature using publicly available Project STAR data, which includes information on high school graduation and whether students took a college-entrance exam (i.e., ACT or SAT). These outcomes closely resemble the dropout and college-intent indicators contained in North Carolina administrative data, though the latter have the advantage over self-reported intent in that they are objective measures of a costly behavior that are not subject to social desirability bias.

[24] Footnote 22 of Chetty et al. (2011) reports finding a positive but statistically insignificant effect of having a same-race teacher on earnings. The paper makes no mention of investigating the impact of having a same-race teacher on educational attainment.

4.2 Data and Methods

Table 10 summarizes the data used in this analysis. The analytic sample includes the full set of black students who entered STAR in any of the four program years (grades K-3). Panel A shows that about one third of black students did not complete high school, about 17% graduated high school but did not take a college entrance exam, and almost one half of students completed high school and took a college entrance exam. The dropout rate here is slightly larger than in North Carolina, though it is worth remembering that the STAR cohort is about ten years younger than the earliest North Carolina cohorts and that STAR targeted some of the most disadvantaged schools in the state. Panel B shows that 44% of black STAR students were assigned a black teacher in their first year of STAR, which is similar to the proportion of North Carolina students *ever* assigned a black teacher in grades 3-5. This suggests that overall, black students were slightly more likely to experience a black teacher in primary school in Tennessee than in North Carolina, though the difference is small in magnitude. Finally, panel C summarizes students' backgrounds. Almost 2/3 of students classify as "persistently poor," which is a higher rate than in North Carolina. This could be due to STAR's concentration in disadvantaged schools.

Investigating the long-run impacts of exposure to a same-race primary-school teacher on black students' educational attainment using the Project STAR data augments the North Carolina analyses in two main ways. First, by leveraging the experimental design of Project STAR, this analysis relies on an entirely different identification strategy. If the two analyses yield similar results, a Hausman-Test type argument suggests that both estimates are credible, as it is unlikely that the two approaches would be biased in similar ways. Second, the STAR analysis bolsters the study's external validity by showing that the long-run impacts of same-race teachers are not unique to North Carolina, nor to the later elementary grades. Additionally, by using college entrance exam taking, which is as an objective measure of college intent, these results assuage concerns over subjectivity and potential social desirability bias in self-reports of college intent.

We estimate the impact of having a same-race teacher in the grade in which the student entered the STAR sample on long-run educational outcomes. Because randomization occurred within schools, we condition on school-by grade of entry fixed effects (FE). Using the sample of black students, we estimate linear models of the form

$$y_{icgs} = \beta X_i + \lambda W_c + \theta_{gs} + \gamma SameRace_i + \varepsilon_{icgs}, \tag{4}$$

where *i*, *c*, *g*, and *s* index students, classrooms, grades, and schools, respectively, *X* is a vector of student characteristics including sex and persistent-FRL status, *W* is a vector of classroom characteristics that include class size indicators, the socio-demographic composition of the class, and the classroom teachers' observed characteristics (i.e., experience, educational attainment, and merit pay receipt), θ is a school-by-grade of entry FE, and *SameRace* is an indicator equal to one if the student's classroom teacher was black in the year the student entered the STAR sample. Due to the within-school random assignment of participating students and teachers to classrooms in kindergarten, then, OLS estimates of γ represent the causal effect of exposure to a same-race kindergarten teacher on long-run educational attainment (Dee 2004; Krueger 1999). We also estimate equation (4) using a fixed-effects logit estimator (Wooldridge 2010) and an ordered-probit model similar to that estimated in Table 3 to increase efficiency.

4.3 Results

Estimates of linear, FE-Logit, and ordered-probit versions of equation (4) are reported in Table 11. All reported coefficients and average partial effects (APE) are for the binary indicator equal to one if the student was assigned a black teacher in his or her first year in STAR, and zero otherwise. Each model is estimated with and without controls, and in no case does adding the controls appreciably change the results. This is consistent with past research (e.g., Dee 2004) and consistent with randomization in STAR having been achieved (Krueger 1999; Dee 2004). The first row in columns 1 and 2 reports OLS estimates of the linear model for high school dropout, and finds that having a same-race teacher reduced the probability of dropping out by about four percentage points. These estimates are similar in size to those obtained for North Carolina in section 3.1, but are imprecisely estimated. The second row of columns 1 and 2 reports OLS estimates of the linear model for taking a college entrance exam, and finds a statistically significant effect of having a same-race teacher of about four percentage points, or 10%. These estimates are quite similar in size to the APE of 0.03 uncovered for North Carolina in Table 3.

Columns 3 and 4 of Table 11 report estimates of analogous FE-logit models. Patterns in the sign and statistical significance of the FE-logit coefficients are consistent with those of the OLS estimates reported in columns 1 and 2. However, the size of the FE-logit coefficients cannot be directly compared to the corresponding OLS estimates, nor can comparable APE be computed, as the FE-logit estimator does not uncover the distribution of the FE (Wooldridge 2010). However, a scaling factor of $p(1-p)$, where p is the probability of success, can be used to generate approximate APE. The relevant scale factors for the two outcomes are 0.2244 and 0.2496, which suggest APE on high-school dropout and college entrance-exam taking of about 0.04 and 0.05, respectively. These APE are remarkably similar to OLS coefficients.

Finally, columns 5 and 6 of Table 11 use the high-school dropout and college entrance exam data to construct a single ordinal measure of educational attainment and estimate an ordered-probit model akin to that estimated in Table 3. The ordered probit model produces marginally significant coefficients and APEs that are consistent with the OLS and FE-logit results presented in columns 1-4 of Table 11, and with the North Carolina ordered-probit estimates reported in Table 3. Specifically, being assigned a black teacher in the first year of STAR reduced the likelihood of dropping out of high school by 0.05 (15%) and increased the likelihood of taking a college entrance exam by 0.05 (10%). These estimates are remarkably similar to those reported for the low-income sample in North Carolina reported in Table 3.

In sum, the results presented in Table 11 provide further evidence that there are long-run, causal effects of having a same-race teacher in elementary school on black students' educational attainment. Moreover, the results presented in this section suggest that the results presented in section 3 are not unique to North Carolina, nor to the later elementary grades (i.e., 3-5). More importantly, that we document similarly sized, arguably causal effects using an entirely different source of exogenous variation in exposure to black teachers suggests that there is, in fact, a causal relationship between black students' exposure to same-race teachers in elementary school and longer-run educational outcomes, particularly among socio-economically disadvantaged students. In other words, this finding is not a spurious artifact of the instrumental variables identification strategy described in section 2.

5. Conclusion

Exposure to a black teacher during elementary school raises long-run educational attainment for black male students, especially among those from low-income households. For the most disadvantaged black males, conservative estimates suggest that exposure to a black teacher in primary school cuts high school dropout rates 39%. It also raises college aspirations along with the probability of taking a college entrance exam. These results come from two unique analyses that utilize datasets from two states and two distinct identification strategies.

Thus, well-established evidence of demographic match effects are not limited to short-run outcomes, such as test scores or teacher perceptions or expectations. Rather, they extend to educational investments several years later. Our findings also suggest that a straightforward policy lever – assignment of black male students to black teachers – can help to close frustratingly persistent achievement gaps.

We discuss two avenues for future research. First, it would be useful to continue to understand what precise student or teacher behaviors help to explain the demographic match effect. A candidate is teacher expectations, which could affect how resources are allocated among students (e.g., teacher time and effort or advanced placement classes). Alternatively, assignment to a black teacher with higher expectations could lead a student to modify his own expectations, which could likewise affect investments and long-run educational attainment. To further uncover mechanisms underlying demographic match effects, further research could also assess other intermediate outcomes, such as risky behavior or delinquency. Finally, further work could assess whether demographic match effects extend to longer-run outcomes, including college completion, employment, course of study, occupational choice, and earnings. Evidence of increased monetary returns to match effects could be especially helpful in justifying potential costs of policies that would increase assignment of black teachers to black students and the recruitment of black teachers.

REFERENCES

Aaronson, D. (1998). Using sibling data to estimate the impact of neighborhoods on children's educational outcomes. *Journal of Human Resources*, 33(4), 915-946.

Angrist, J. D., & Pischke, J. S. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.

Araujo, M. C., Carneiro, P., Cruz-Aguayo, Y., & Schady, N. (2016). Teacher quality and learning outcomes in kindergarten. *Quarterly Journal of Economics*, *131*(3), 1415-1453.

Bailey, M. J., & Dynarski, S. M. (2011). Gains and gaps: Changing inequality in US college entry and completion. National Bureau of Economic Research Working Paper No. w17633.

Banerjee, A. V., Cole, S., Duflo, E., & Linden, L. (2007). Remedying education: Evidence from two randomized experiments in India. *Quarterly Journal of Economics*, *122*(3), 1235-1264.

Bettinger, E. P., & Long, B. T. (2005). Do faculty serve as role models? The impact of instructor gender on female students. *The American Economic Review*, *95*(2), 152-157.

Bettinger, E. P., & Long, B. T. (2010). Does cheaper mean better? The impact of using adjunct instructors on student outcomes. *The Review of Economics and Statistics*, *92*(3), 598-613.

Birdsall, C., Gershenson, S., & Zuniga, R. (2016). Stereotype threat, role models, and demographic mismatch in an elite professional school setting. IZA Discussion Paper No. 10459.

Bound, J., & Solon, G. (1999). Double trouble: on the value of twins-based estimation of the return to schooling. *Economics of Education Review*, *18*(2), 169-182.

Brummet, Q., Gershenson, S., & Hayes, M. S. (2017). Teachers' grade-level reassignments: Evidence from Michigan. *Educational Policy*, 31(2): 249-272.

Carrell, S. E., & Hoekstra, M. (2014). Are school counselors an effective education input? *Economics Letters*, *125*(1), 66-69.

Carrell, S. E., Page, M. E., & West, J. E. (2010). Sex and science: How professor gender perpetuates the gender gap. *The Quarterly Journal of Economics*, *125*(3), 1101-1144.

Chetty, R., Friedman, J. N., Hilger, N., Saez, E., Schanzenbach, D. W., & Yagan, D. (2011). How does your kindergarten classroom affect your earnings? Evidence from Project Star. *The Quarterly Journal of Economics*, *126*(4), 1593-1660.

Chetty, R., Friedman, J., & Rockoff, J. (2014). Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood. *American Economic Review*, *104*, 2633-79.

Chetty, R., Hendren, N., Lin, F., Majerovitz, J., & Scuderi, B. (2016). Gender gaps in childhood: Skills, behavior, and labor market preparedness childhood environment and gender gaps in adulthood. *The American Economic Review*, *106*(5), 282-288.

Chingos, M. M., & Peterson, P. E. (2011). It's easier to pick a good teacher than to train one: Familiar and new results on the correlates of teacher effectiveness. *Economics of Education Review*, *30*(3), 449-465.

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2006). Teacher-student matching and the assessment of teacher effectiveness. *Journal of Human Resources*, 41(4), 778-820.

Currie, J., & Thomas, D. (1999). Early test scores, socioeconomic status and future outcomes. (No. w6943). National bureau of economic research.

Dee, T. S. (2004). Teachers, race, and student achievement in a randomized experiment. *Review of Economics and Statistics, 86*(1), 195-210.

Dee, T. S. (2005). A teacher like me: Does race, ethnicity, or gender matter? *American Economic Review*, 95(2), 158–165.

Dee, T. S. (2007). Teachers and the gender gaps in student achievement. *Journal of Human Resources*, *42*(3), 528-554.

Dieterle, S. G., & Snell, A. (2016). A simple diagnostic to investigate instrument validity and heterogeneous effects when using a single instrument. *Labour Economics*, *42*, 76-86.

Dieterle, S., Guarino, C. M., Reckase, M. D., & Wooldridge, J. M. (2015). How do principals assign students to teachers? Finding evidence in administrative data and the implications for value added. *Journal of Policy Analysis and Management*, *34*(1), 32-58.

Dillon, E. W., & Smith, J. A. (2017). Determinants of the match between student ability and college quality. *Journal of Labor Economics*, *35*(1), 45-66.

Dynarski, S., Hyman, J., & Schanzenbach, D. W. (2013). Experimental evidence on the effect of childhood investments on postsecondary attainment and degree completion. *Journal of Policy Analysis and Management*, *32*(4), 692-717.

Egalite, A. J., Kisida, B., & Winters, M. A. (2015). Representation in the classroom: The effect of own-race teachers on student achievement. *Economics of Education Review*, *45*, 44-52.

Fairlie, R. W., Hoffmann, F., & Oreopoulos, P. (2014). A community college instructor like me: Race and ethnicity interactions in the classroom. *The American Economic Review*, *104*(8), 2567-2591.

Figlio, D. N., Karbownik, K., & Salvanes, K. G. (2015). Education research and administrative data. National Bureau of Economic Research Working Paper No. w21592.

Figlio, D., Karbownik, K., Roth, J., & Wasserman, M. (2016). School Quality and the Gender Gap in Educational Achievement. *The American Economic Review*, *106*(5), 289-295.

Fortin, N. M., Oreopoulos, P., & Phipps, S. (2015). Leaving Boys Behind Gender Disparities in High Academic Achievement. *Journal of Human Resources*, *50*(3), 549-579.

Fryer, R. G. (2014). Injecting charter school best practices into traditional public schools: Evidence from field experiments. *Quarterly Journal of Economics*, *129*(3), 1355-1407.

Garces, E., Currie, J., & Thomas, D. (2002). Longer-Term Effects of Head Start. *American Economic Review*, *92*(4), 999-1012.

Gershenson, S. (2016). Linking teacher quality, student attendance, and student achievement. *Education Finance & Policy*, 11(2): 125-149.

Gershenson, S., Holt, S. B., & Papageorge, N. W. (2016). Who believes in me? The effect of student–teacher demographic match on teacher expectations. *Economics of Education Review*, *52*, 209-224.

Goldring, R., Gray, L., & Bitterman, A. (2013). Characteristics of Public and Private Elementary and Secondary School Teachers in the United States: Results from the 2011-12 Schools and Staffing Survey. First Look. NCES 2013-314. *National Center for Education Statistics*.

Greene, W. (2004). The behaviour of the maximum likelihood estimator of limited dependent variable models in the presence of fixed effects. *The Econometrics Journal*, *7*(1), 98-119.

Hanushek, E. A. (1979). Conceptual and empirical issues in the estimation of educational production functions. *Journal of Human Resources*, 14(3), 351-388.

Hanushek, E. A., & Rivkin, S. G. (2010). Generalizations about using value-added measures of teacher quality. *The American Economic Review*, *100*(2), 267-271.

Hanushek, E. A., Kain, J. F., & Rivkin, S. G. (2004). Why public schools lose teachers. *Journal of Human Resources*, 39(2), 326–354.

Holt, S. B., & Gershenson, S. (2015). The impact of teacher demographic representation on student attendance and suspensions. IZA Discussion Paper No. 9554.

Jackson, C. K. (2009). Student demographics, teacher sorting, and teacher quality: Evidence from the end of school desegregation. *Journal of Labor Economics*, 27(2), 213–256.

Jackson, C. K. (2012). Non-cognitive ability, test scores, and teacher quality: Evidence from 9th grade teachers in North Carolina (No. w18624). National Bureau of Economic Research.

Kane, T. J., & Staiger, D. O. (2008). Estimating teacher impacts on student achievement: An experimental evaluation. (No. w14607). National Bureau of Economic Research.

Krueger, A. B. (1999). Experimental estimates of education production functions. *Quarterly Journal of Economics*, *114*(2), 497-532.

Krueger, A. B., & Whitmore, D. M. (2001). The effect of attending a small class in the early grades on college-test taking and middle school test results: Evidence from Project STAR. *The Economic Journal*, *111*(468), 1-28.

Lareau, A. (1987). Social class differences in family-school relationships: The importance of cultural capital. *Sociology of Education*, 73-85.

Lareau, A., & Weininger, E. B. (2003). Cultural capital in educational research: A critical assessment. *Theory and Society*, *32*(5-6), 567-606.

Lindsay, C.A., & Hart, C.M.D. (2017). Teacher-student race match and student disciplinary outcomes for black students in North Carolina. In press, *Educational Evaluation and Policy Analysis.* DOI: https://doi.org/10.3102/0162373717693109

Michelmore, K., & Dynarski, S. (2016). The gap within the gap: Using longitudinal data to understand income differences in student achievement. NBER Working Paper No. 22474.

Muralidharan, K., & Sundararaman, V. (2011). Teacher performance pay: Experimental evidence from India. *Journal of Political Economy*, *119*(1), 39-77.

Murnane, R. (2013). U.S. High school graduation rates: Patterns and explanations. *Journal of Economic Literature*, 51(2), 370-422.

Ouazad, A. (2014). Assessed by a teacher like me: Race and teacher assessments. *Education Finance & Policy*, 9(3), 334-372.

Papageorge, N. W., Gershenson, S., & Kang, K. (2016). Teacher expectations matter. IZA Discussion Paper No. 10165.

Putman, H., Hansen, M., Walsh, K., & Quintero, D. (2016). High Hopes and Harsh Realities: The Real Challenges to Building a Diverse Workforce. *Brookings Institution*.

Rockoff, J. E., Jacob, B. A., Kane, T. J., & Staiger, D. O. (2011). Can you recognize an effective teacher when you recruit one? *Education Finance and Policy*, *6*(1), 43-74.

Roodman, David. (2011). Fitting fully observed recursive mixed-process models with CMP. *Stata Journal* 11(2), 159–206.

Rothstein, J. (2010). Teacher quality in educational production: Tracking, decay, and student achievement. *The Quarterly Journal of Economics*, *125*(1), 175-214.

Todd, P. E., & Wolpin, K. I. (2003). On the specification and estimation of the production function for cognitive achievement. *The Economic Journal*, *113*(485), F3-F33.

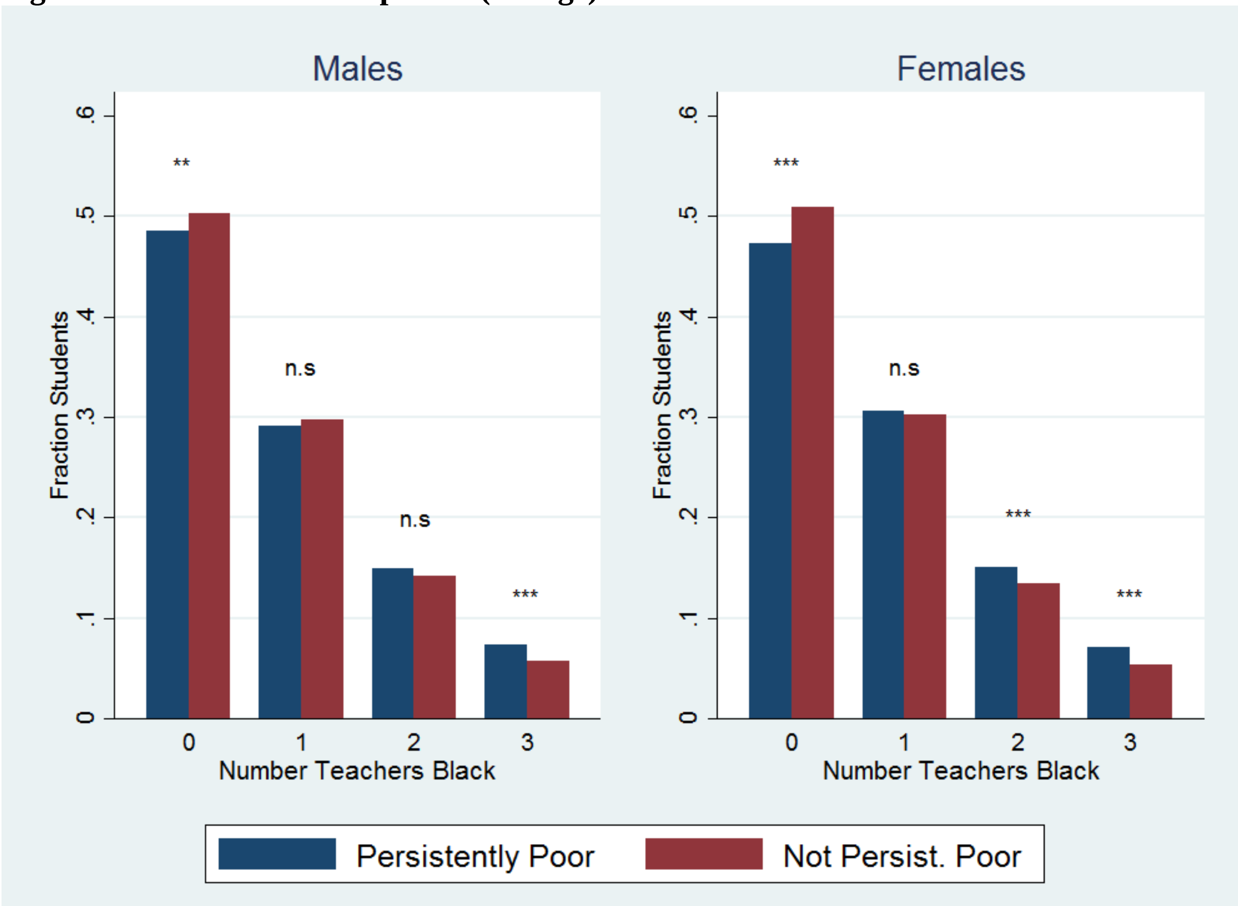
Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist*, 52(6), 613-629.

Stock, J. H., & Yogo, M. (2002). Testing for weak instruments in linear IV regression. NBER Technical Working Paper No. 284. National Bureau of Economic Research: Cambridge, MA.

Wiswall, M. (2013). The dynamics of teacher quality. *Journal of Public Economics*, *100*, 61-78.

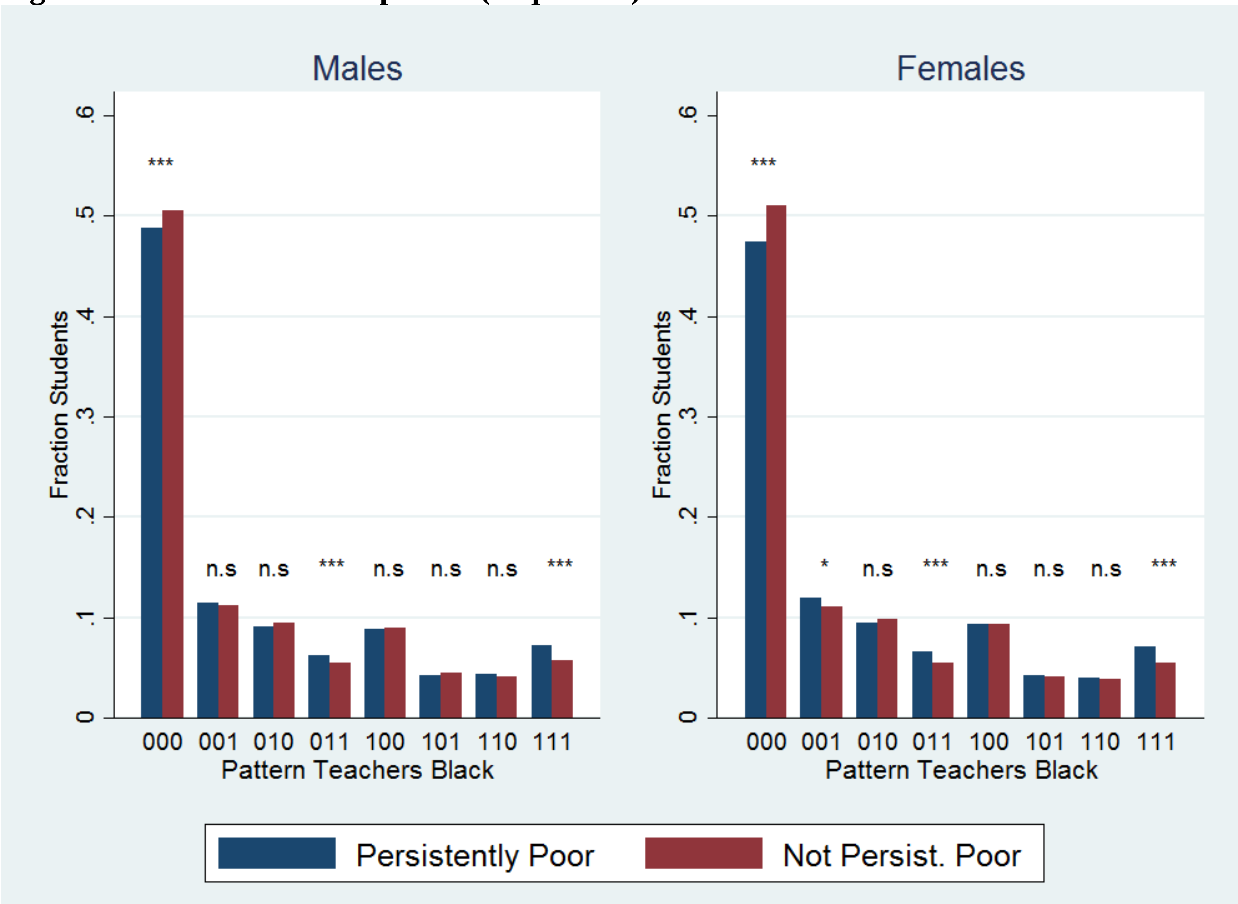
Wooldridge, J. M. (2010). Econometric analysis of cross section and panel data. MIT

Figure 1. Distribution of Exposure (Dosage) to Black Teachers in Grades 3-5



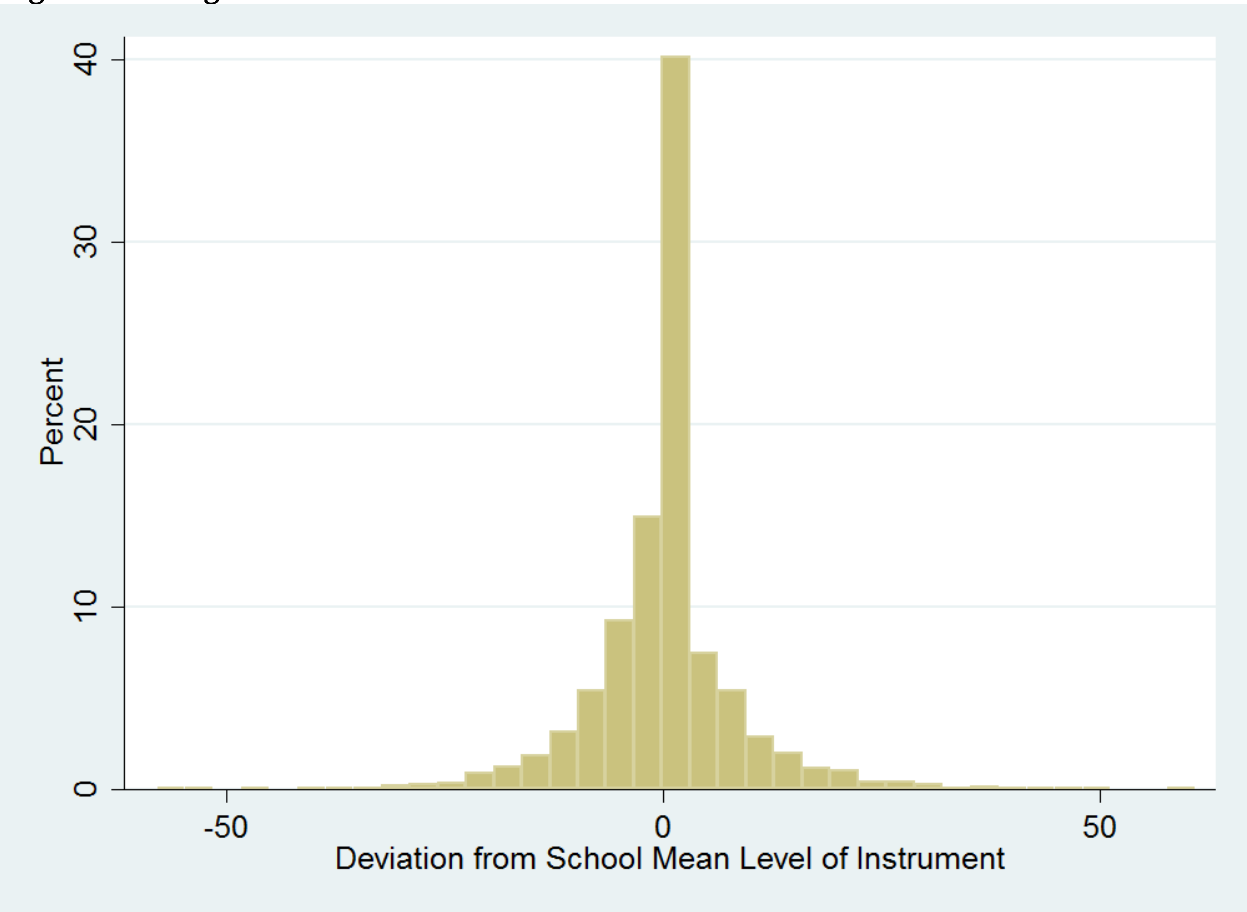
Note: Persistently poor indicates that the student was persistently eligible for free or reduced price lunch in each of grades 3-8. * p<.10, ** p<.05, ***p<.01, n.s. (not significant) p>0.10.

Figure 2. Distribution of Exposure (Sequences) to Black Teachers in Grades 3-5



Note: Persistently poor indicates that the student was persistently eligible for free or reduced price lunch in each of grades 3-8. 0 and 1 indicate assignment to other- and same-race teachers, respectively. For example, a sequence of 101 indicates that the student had a black teacher in grade 3, a white teacher in grade 4, and a black teacher in grade 5.
* p<.10, ** p<.05, ***p<.01, n.s. (not significant) p>0.10.

Figure 3. Histogram of School-Year Deviations from School Mean % Black Teachers



Notes: The unit of analysis is the school-year. There are a total of 5,636 school-years. Because baseline models condition on school fixed effects, the identifying variation in the instrument comes from within-school variation in the fraction of teachers in grades 3-5 who are black.

Table 1. Descriptive Statistics for Black Students, Cohorts Entering Grade 3 in 2001-2005

	All Black Students			Persistently Low-Income Black Students		
	All (1)	Male (2)	Female (3)	All (4)	Male (5)	Female (6)
A. Educational Outcomes						
H.S. Dropout	13.3%	16.9%	9.9%	14.6%	18.2%	11.4%
H.S. Grad, No College Intent	46.6%	48.8%	44.4%	52.2%	53.6%	51.0%
H.S. Grad, College Intent	40.2%	34.4%	45.8%	33.2%	28.2%	37.8%
B. Own Grade 3-5 Teachers						
Exposure to ≥1 Black T	43.8%	43.7%	44.0%	45.6%	45.3%	45.8%
Number of Black Ts	0.61	0.61	0.61	0.64	0.64	0.65
	(0.80)	(0.81)	(0.80)	(0.82)	(0.83)	(0.82)
Number of Black Ts | ≥1	1.39	1.40	1.39	1.41	1.42	1.41
	(0.62)	(0.62)	(0.61)	(0.63)	(0.64)	(0.63)
C. Instrument						
% Cohort's Teachers Black	25.4	25.4	25.5	26.9	26.8	27.0
	(24.9)	(25.0)	(24.9)	(25.9)	(26.0)	(25.9)
[Within-school SD]	[9.20]	[9.18]	[9.23]	[9.47]	[9.43]	[9.51]
D. Student Characteristics						
Persistently Low-Income	45.4%	44.2%	46.6%	100%	100%	100%
Ever Low-Income	85.8%	85.6%	85.9%	100%	100%	100%
Ever LEP	0.29%	0.34%	0.24%	0.33%	0.40%	0.26%
Ever Exceptional	10.1%	13.6%	6.7%	11.0%	14.9%	7.4%
N (Students)	106,370	51,929	54,441	48,332	22,964	25,368

Notes: Standard deviations (SD) are reported in parentheses for non-binary variables. Persistently low-income is defined as being low-income in each of grades K-8. H.S. = high school. LEP = Limited English Proficient. T = teacher. H.S. dropouts are identified by the state as students who drop from enrollment in year *t* after having attended in year *t-1* and without graduating. Sample excludes students missing from public school data by 8th grade; students who exit NC school system for out-of-state schools, private schools, home schools, or death, excluded from NC cohort count; students missing own elementary teacher race composition in all years; and students missing clear indicators of either graduation or drop-out outcomes.

Table 2. Long-Run Effects of Exposure to a Black Teacher on Educational Attainment

	All Black Students			Persistently Low-Income Black Students		
	All (1)	Male (2)	Female (3)	All (4)	Male (5)	Female (6)
A. First Stage						
% Cohort Teachers Black	0.71***	0.70***	0.73***	0.70***	0.67***	0.73***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)	(0.04)
Schools	1,210	1,158	1,160	1,108	1,033	1,043
Students (N)	105,099	51,249	53,778	47,829	22,656	25,065
B. 2SLS: H.S. Dropout						
1{≥ 1 black teacher}	-0.04**	-0.07***	-0.00	-0.06**	-0.12***	0.00
	(0.02)	(0.03)	(0.02)	(0.02)	(0.04)	(0.03)
Hausman Test *(p)*	0.00	0.01	0.00	0.13	0.87	0.07
Students (N)	105,099	51,249	53,778	47,829	22,656	25,065
Outcome comparison mean	0.13	0.16	0.09	0.14	0.18	0.11
C. 2SLS: College Intent						
1{ ≥1 black teacher}	0.03	-0.00	0.04	0.11***	0.10**	0.11**
	(0.02)	(0.03)	(0.03)	(0.04)	(0.05)	(0.05)
Hausman Test *(p)*	0.04	0.31	0.07	0.44	0.75	0.18
Students (N)	103,633	50,381	53,174	47,096	22,232	24,754
Outcome comparison mean	0.41	0.35	0.46	0.33	0.28	0.37

Notes: Standard errors are clustered by the student's grade-3 school. The instrument in the 2SLS first stage is the share of Grade 3-5 teachers in the student's grade-3 school's cohort who are black. 1{.} is the indicator function. All models control for student characteristics including sex, "persistent" poverty status (in all grades 3-8), whether the student was ever classified as Limited English Proficient, and whether the student was ever classified as having a disability. All models control for school fixed effects and time-varying school characteristics (the year the student entered grade 3) including average End of Grade math and reading scores, student-pupil ratio, logged student enrollment, share of students using subsidized lunch, share of students black, Hispanic, or other race, urbanicity, and charter/magnet status. Panel A reports first-stage estimates for the H.S. drop-out regressions shown in Panel B. First-stage estimates are similarly strong for the H.S. grad and college intent estimates reported in Panel C. Hausman tests comparing the 2SLS estimates to analogous OLS estimates were conducted using control functions. "Outcome comparison mean" is the mean of the dependent variable for the relevant control group: students who experienced zero black teachers in grades 3-5 but attended schools with at least one grade 3-5 black teacher. These means are reported to contextualize the magnitudes of the point estimates. * p<.10, ** p<.05, ***p<.01.

Table 3. Probit-Ordered Probit Estimates of Long-Run Impact of Having a Black Teacher

	All	Low-Income	Low-Income Males
	(1)	(2)	(3)
A. Probit First Stage			
% Cohort Teachers Black	2.27***	2.28***	2.20***
	(0.09)	(0.11)	(0.11)
B. Ordered-Probit Cofficients			
1{≥ 1 black teacher}	0.08*	0.18***	0.26***
	(0.04)	(0.05)	(0.08)
C. Ordered Probit APE			
APE: Dropout	-0.02*	-0.04***	-0.07***
	(0.01)	(0.01)	(0.02)
APE: Graduate, No Intent	-0.01**	-0.02***	-0.02***
	(0.01)	(0.01)	(0.01)
APE: Graduate, Intend	0.03**	0.06***	0.08***
	(0.01)	(0.02)	(0.03)
Schools	1,261	1,176	1,125
Students (N)	105,150	47,897	22,748
Mean comparison dropout	0.13	0.14	0.18
Mean comparison intend	0.41	0.33	0.28

Notes: The ordered probit-probit model acknowledges that the two outcomes can be combined into an ordinal variable that takes three values: high school dropout, high school graduate with no college intent, and college graduate with college intent. The first stage remains a probit. The ordered probit-probit system is jointly estimated as a mixed-process model (Roodman 2011). APE is average partial effect. Mean comparison outcomes are the outcome means for students who experienced zero black teachers in grades 3-5, but attended schools with at least one grade 3-5 black teacher. These means are reported to contextualize the APE. * p<.10, ** p<.05, ***p<.01.

Table 4. Dosage Effects of Exposure to Black Teachers on Educational Attainment

	All Black Students			Persistently Low-Income Black Students		
	All (1)	Male (2)	Female (3)	All (4)	Male (5)	Female (6)
A. First Stages (F Statistic)						
Eq. 1: 1 Teacher Black	115.2	83.9	82.7	72.5	40.7	47.9
Eq. 2: 2 Teachers Black	86.3	46.6	84.6	50.8	22.3	47.0
Eq. 3: 3 Teachers Black	119.8	72.5	84.1	74.9	39.0	46.0
B. 2SLS: H.S. Dropout						
1{1 Teacher Black}	-0.02	-0.05	0.06	-0.08	-0.17**	0.07
	(0.04)	(0.06)	(0.04)	(0.06)	(0.08)	(0.07)
1{2 Teachers Black}	-0.03	-0.13*	0.01	-0.01	-0.07	0.02
	(0.04)	(0.08)	(0.04)	(0.05)	(0.10)	(0.06)
1{3 Teachers Black}	-0.08	-0.20***	0.07	-0.13	-0.26**	0.11
	(0.05)	(0.08)	(0.07)	(0.08)	(0.10)	(0.10)
Joint Sig., Dosage Var. (*p*)	0.15	0.00	0.43	0.29	0.01	0.63
t test, 1 = 3 (*p*)	0.12	0.03	0.88	0.48	0.35	0.64
t test, 1 = 2 (*p*)	0.84	0.52	0.44	0.41	0.49	0.59
t test, 2 = 3 (*p*)	0.50	0.61	0.53	0.29	0.24	0.48
Hausman Test *(p)*	0.04	0.36	0.03	0.65	0.89	0.21
Students (N)	46,646	22,521	24,036	20,941	9,672	11,117
C. 2SLS: College Intent						
1{1 Teacher Black}	-0.01	-0.05	0.02	0.06	0.03	0.01
	(0.06)	(0.08)	(0.07)	(0.07)	(0.10)	(0.12)
1{2 Teachers Black}	0.04	0.09	-0.01	0.13	0.21	0.08
	(0.07)	(0.11)	(0.08)	(0.09)	(0.14)	(0.11)
1{3 Teachers Black}	-0.02	-0.12	0.06	0.12	-0.06	0.13
	(0.08)	(0.10)	(0.11)	(0.10)	(0.12)	(0.15)
Joint Sig., Dosage Var. (*p*)	0.94	0.70	0.94	0.13	0.33	0.59
t test, 1 = 3 (*p*)	0.86	0.42	0.58	0.53	0.50	0.34
t test, 1 = 2 (*p*)	0.62	0.41	0.85	0.64	0.36	0.70
t test, 2 = 3 (*p*)	0.63	0.26	0.67	0.96	0.20	0.82
Hausman Test *(p)*	0.11	0.06	0.30	0.66	0.69	0.49
Students (N)	46,037	22,170	23,779	20,637	9,507	10,978

Notes: Standard errors are clustered by the student's grade-3 school. The instruments in the 2SLS first stages are the predicted probabilities that the student had one, two, and three black teachers in grades 3-5. Predicted probabilities were generated by poisson regressions of the number of black teachers actually experienced by the student on the share of Grade 3-5 teachers in the student's grade-3 school's cohort who are black. 1{.} is the indicator function. All models control for student characteristics including sex, "persistent" poverty status (in all grades 3-8), whether the student was ever classified as Limited English Proficient, and whether the student was ever classified as having a disability. All models control for school fixed effects and time-varying school characteristics (the year the student entered grade 3) including average End of Grade math and reading scores, student-pupil ratio, logged student enrollment, share of students using subsidized lunch, share of students black, Hispanic, or other race, urbanicity, and charter/magnet status. Panel A reports first-stage results for the H.S. drop-out regressions shown in Panel B. First-stage estimates are similarly strong for the H.S. grad and college intent estimates reported in Panel C. Hausman tests comparing the 2SLS estimates to analogous OLS estimates were conducted using control functions.
* p<.10, ** p<.05, ***p<.01.

Table 5. Timing Effects of Exposure to Black Teachers on Educational Attainment

	All Black Students			Persistently Low-Income Black Students		
	All	Male	Female	All	Male	Female
	(1)	(2)	(3)	(4)	(5)	(6)
A. First Stages (Joint Sig.)						
Eq. 1: G3 Teacher Black	744.9	472.7	538.1	448.4	254.8	274.3
Eq. 2: G4 Teacher Black	346.6	238.5	247.1	187.6	119.6	117.6
Eq. 3: G5 Teacher Black	324.2	229.2	170.1	180.1	84.9	101.6
B. 2SLS: H.S. Dropout						
1{G3 Teacher Black}	-0.02	-0.06**	0.02	-0.03*	-0.08**	0.01
	(0.01)	(0.02)	(0.02)	(0.02)	(0.04)	(0.02)
1{G4 Teacher Black}	-0.02	-0.06*	0.01	-0.03	-0.05	0.01
	(0.02)	(0.03)	(0.03)	(0.03)	(0.04)	(0.03)
1{G5 Teacher Black}	-0.03	-0.05	0.02	-0.00	-0.02	0.03
	(0.02)	(0.03)	(0.03)	(0.04)	(0.05)	(0.04)
Joint Sig., Timing Var. (*p*)	0.23	0.00	0.69	0.22	0.05	0.91
t test, G3 = G4 (*p*)	0.80	1.00	0.91	0.83	0.61	0.92
t test, G3 = G5 (*p*)	0.91	0.88	1.00	0.41	0.29	0.66
t test, G4 = G5 (*p*)	0.76	0.90	0.92	0.55	0.65	0.74
Students (N)	41,096	19,786	21,235	17,911	8,180	9,583
C. 2SLS: College Intent						
1{G3 Teacher Black}	0.01	-0.01	0.02	0.05	-0.02	0.08*
	(0.02)	(0.03)	(0.03)	(0.03)	(0.05)	(0.05)
1{G4 Teacher Black}	-0.00	0.03	-0.04	0.03	0.05	0.02
	(0.03)	(0.04)	(0.04)	(0.04)	(0.05)	(0.06)
1{G5 Teacher Black}	-0.00	-0.06	0.06	0.05	-0.01	0.04
	(0.03)	(0.04)	(0.04)	(0.05)	(0.06)	(0.07)
Joint Sig., Timing Var. (*p*)	0.98	0.50	0.36	0.29	0.77	0.31
t test, G3 = G4 (*p*)	0.79	0.47	0.25	0.76	0.30	0.41
t test, G3 = G5 (*p*)	0.81	0.31	0.43	0.97	0.84	0.60
t test, G4 = G5 (*p*)	1.00	0.14	0.10	0.82	0.50	0.82
Students (N)	40,560	19,469	21,017	17,662	8,044	9,471

Notes: Standard errors are clustered by the student's grade-3 school. The instruments in the 2SLS first stages are the share teachers black in grades 3-5 for each student's grade 3-school-cohort. 1{.} is the indicator function. G = grade. All models control for student characteristics including sex, "persistent" poverty status (in all grades 3-8), whether the student was ever classified as Limited English Proficient, and whether the student was ever classified as having a disability. All models control for school fixed effects and time-varying school characteristics (the year the student entered grade 3) including average End of Grade math and reading scores, student-pupil ratio, logged student enrollment, share of students using subsidized lunch, share of students black, Hispanic, or other race, urbanicity, and charter/magnet status. Panel A reports first-stage results for the H.S. drop-out regressions shown in Panel B. First-stage estimates are similarly strong for the H.S. grad and college intent estimates reported in Panel C. Hausman tests comparing the 2SLS estimates to analogous OLS estimates were conducted using control functions. * p<.10, ** p<.05, ***p<.01.

Table 6. Sequencing Effects of Exposure to Black Teachers on Educational Attainment

	All Black Students			Persistently Low-Income Black Students		
	All	Male	Female	All	Male	Female
	(1)	(2)	(3)	(4)	(5)	(6)
A. First Stages (Joint Sig.)						
Eq. 1: 100	128.8	89.2	68.5	68.9	46.1	31.4
Eq. 2: 010	109.0	67.8	76.4	61.7	31.4	38.6
Eq. 3: 001	100.8	62.8	59.5	52.9	29.7	38.0
Eq. 4: 110	54.0	34.9	33.6	37.6	24.9	18.7
Eq. 5: 101	58.4	29.9	38.2	25.3	9.5	18.0
Eq. 6: 001	69.2	53.1	37.9	39.0	17.7	28.8
Eq. 7: 111	41.7	27.2	31.7	30.9	17.5	19.4
B. 2SLS: H.S. Dropout						
Sequence 100	0.04	-0.03	0.10**	-0.04	-0.17**	0.12
	(0.03)	(0.05)	(0.04)	(0.04)	(0.08)	(0.08)
Sequence 010	0.01	-0.03	0.06	-0.02	-0.10	0.11
	(0.04)	(0.06)	(0.05)	(0.05)	(0.09)	(0.08)
Sequence 001	-0.02	-0.14**	0.07	-0.07	-0.20**	0.09
	(0.04)	(0.07)	(0.04)	(0.06)	(0.10)	(0.07)
Sequence 110	-0.13**	-0.22***	-0.04	-0.17***	-0.24**	-0.11
	(0.05)	(0.08)	(0.07)	(0.06)	(0.11)	(0.12)
Sequence 101	-0.05	0.01	-0.05	-0.02	0.09	-0.05
	(0.05)	(0.09)	(0.06)	(0.08)	(0.16)	(0.12)
Sequence 011	-0.01	-0.05	0.03	-0.03	-0.05	0.02
	(0.04)	(0.06)	(0.05)	(0.06)	(0.12)	(0.07)
Sequence 111	-0.05	-0.22***	0.08	-0.04	-0.15	0.13
	(0.05)	(0.08)	(0.08)	(0.08)	(0.11)	(0.10)
p-values, post-hoc tests						
Joint Sig., Sequence Var.	0.09	0.00	0.43	0.11	0.02	0.84
t test, 001 = 100	0.12	0.15	0.43	0.58	0.74	0.71
Hausman Test	0.03	0.01	0.21	0.54	0.62	0.76
Students (N)	40,573	19,498	20,999	17,658	8,055	9,456

Notes: Standard errors are clustered by the student's grade-3 school. The instruments in the 2SLS first stages are the share teachers black in each of grades 3-5 for each student's grade 3-school-cohort, along with all possible interactions. Sequence 000 is the omitted reference group. G = grade. All models control for student characteristics including sex, "persistent" poverty status (in all grades 3-8), whether the student was ever classified as Limited English Proficient, and whether the student was ever classified as having a disability. All models control for school fixed effects and time-varying school characteristics (the year the student entered grade 3) including average End of Grade math and reading scores, student-pupil ratio, logged student enrollment, share of students using subsidized lunch, share of students black, Hispanic, or other race, urbanicity, and charter/magnet status. Hausman tests comparing the 2SLS estimates to analogous OLS estimates were conducted using control functions.
* p<.10, ** p<.05, ***p<.01.

Table 7. Effects of Exposure to a Black Teacher Using Alternate Definitions of "Low-Income"

	Ever FRL, Grades 3-8			Always FRL, Grades 3-5		
	All	Male	Female	All	Male	Female
	(1)	(2)	(3)	(4)	(5)	(6)
A. First Stage						
% Cohort Teachers Black	0.70***	0.68***	0.72***	0.71***	0.68***	0.74***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)	(0.04)
Schools	1,189	1,134	1,128	1,135	1,074	1,076
Students (N)	90,182	43,882	46,216	65,244	31,524	33,626
B. 2SLS: H.S. Dropout						
1{≥ 1 black teacher}	-0.05**	-0.09***	-0.01	-0.05**	-0.10***	-0.01
	(0.02)	(0.03)	(0.02)	(0.02)	(0.04)	(0.03)
Hausman Test (*p*)	0.00	0.01	0.00	0.00	0.16	0.00
Students (N)	90,182	43,882	46,216	65,244	31,524	33,626
C. 2SLS: College Intent						
1{ ≥1 black teacher}	0.04	0.02	0.06	0.10***	0.08*	0.11***
	(0.03)	(0.04)	(0.04)	(0.03)	(0.04)	(0.04)
Hausman Test (*p*)	0.16	0.54	0.26	0.23	0.80	0.23
Students (N)	88,868	43,116	45,665	64,202	30,918	33,193

Notes: Standard errors are clustered by the student's grade-3 school. The instrument in the 2SLS first stage is the share of Grade 3-5 teachers in the student's grade-3 school's cohort who are black. 1{.} is the indicator function. All models control for student characteristics including sex, "persistent" poverty status (in all grades 3-8), whether the student was ever classified as Limited English Proficient, and whether the student was ever classified as having a disability. All models control for school fixed effects and time-varying school characteristics (the year the student entered grade 3) including average End of Grade math and reading scores, student-pupil ratio, logged student enrollment, share of students using subsidized lunch, share of students black, Hispanic, or other race, urbanicity, and charter/magnet status. Panel A reports first-stage estimates for the H.S. drop-out regressions shown in Panel B. First-stage estimates are similarly strong for the H.S. grad and college intent estimates reported in Panel C. Hausman tests comparing the 2SLS estimates to analogous OLS estimates were conducted using control functions. * p<.10, ** p<.05, ***p<.01.

Table 8. Sensitivity of 2SLS Estimates of Impact of Exposure to a Black Teacher on Attainment

	H.S. Dropout			Intends College		
	All	Low-Income	Low-Income Males	All	Low-Income	Low-Income Males
	(1)	(2)	(3)	(4)	(5)	(6)
A. Baseline Estimates						
1{≥1 black teacher}	-0.04**	-0.06**	-0.12***	0.03	0.11***	0.10**
	(0.02)	(0.02)	(0.04)	(0.02)	(0.04)	(0.05)
Students (N)	105,099	47,829	22,656	103,633	47,096	22,232
B. Cubic First Stage						
Joint Sig. of IV (F)	251.0	174.2	112.3	255.2	180.4	114.9
1{≥1 black teacher}	-0.03*	-0.06***	-0.12***	0.01	0.09***	0.06
	(0.02)	(0.02)	(0.04)	(0.02)	(0.03)	(0.04)
Students (N)	105,099	47,829	22,656	103,633	47,096	22,232
C. School Time Trends						
1{≥1 black teacher}	-0.03	-0.06*	-0.07	0.02	0.13***	0.08
	(0.02)	(0.03)	(0.05)	(0.03)	(0.05)	(0.07)
Students (N)	105,099	47,829	22,656	103,633	47,096	22,232
D. Missing H.S. Grad.						
Code as H.S. Grad	-0.03**	-0.05**	-0.11***	N/A	N/A	N/A
	(0.02)	(0.02)	(0.03)			
Students (N)	119,269	54,140	26,302			
Code as Dropout	-0.02	-0.04	-0.08*	N/A	N/A	N/A
	(0.02)	(0.03)	(0.04)			
Students (N)	119,269	54,140	26,302			
E. Analytic Sample						
Non-Movers	-0.04**	-0.06**	-0.11***	0.03	0.11***	0.12***
	(0.02)	(0.03)	(0.04)	(0.02)	(0.04)	(0.05)
Students (N)	92,050	41,932	19,854	90,767	41,282	19,471
2000-06 G3 Cohorts	-0.02	-0.03*	-0.07**	0.03	0.07***	0.08**
	(0.01)	(0.02)	(0.03)	(0.02)	(0.03)	(0.03)
Students (N)	143,277	66,074	31,126	141,262	65,025	30,530
Non-Black Students	-0.01	0.01	-0.01	0.02	0.01	0.00
	(0.01)	(0.03)	(0.05)	(0.02)	(0.03)	(0.04)
Students (N)	253,648	44,084	22,210	250,871	43,518	21,894

Notes: Standard errors clustered by student's grade-3 school. Instrumental variable (IV) in 2SLS first stage is share of Grade 3-5 teachers in student's grade-3 school's cohort who are black. 1{.} is the indicator function. All models control for student characteristics including sex, "persistent" poverty status (in all grades 3-8), whether the student was ever classified as Limited English Proficient, and whether the student was ever classified as having a disability. All models control for school fixed effects and time-varying school characteristics (in year student entered grade 3) including average End-of-Grade math and reading scores, student-pupil ratio, logged student enrollment, share of students using subsidized lunch, share of students black, Hispanic, or other race, urbanicity, and charter/magnet status. Panel A replicates the baseline estimates from Table 2. Panel B uses a cubic of the IV in the first stage. Panel C adds school-specific linear time trends to the baseline model. Panel D adds students whose high school graduation status is missing to the analytic sample, and codes them as dropouts and graduates, respectively. Panel E estimates the baseline model on alternative samples: Non-school changers, students in an expanded range of 3rd grade entry cohorts (2000-2006), and nonwhite students. * p<.10, ** p<.05, ***p<.01.

Table 9. Sibling-School FE IV Estimates of Impact of Exposure to a Black Teacher on Attainment

	All			Low-Income		
	(1)	(2)	(3)	(4)	(5)	(6)
A. First Stage						
% Cohort Teachers Black	0.71***	0.74***	0.73***	0.70***	0.70***	0.65***
	(0.03)	(0.05)	(0.05)	(0.03)	(0.06)	(0.08)
Students (N)	105,099	15,298	15,298	47,829	8,116	8,116
B. 2SLS: H.S. Dropout						
1{≥ 1 black teacher}	-0.04**	-0.01	-0.08**	-0.06**	-0.06	-0.17**
	(0.02)	(0.04)	(0.04)	(0.03)	(0.05)	(0.07)
Hausman Test	0.00	0.09	0.24	0.13	0.67	0.28
Students (N)	105,099	15,298	15,298	47,829	8,116	8,116
C. 2SLS: College Intent						
1{ ≥1 black teacher}	0.03	0.04	0.04	0.11***	0.07	0.11
	(0.02)	(0.07)	(0.07)	(0.04)	(0.10)	(0.11)
Hausman Test	0.04	0.15	0.01	0.44	0.19	0.16
Students (N)	103,633	15,096	15,096	47,096	8,004	8,004
Family Fixed Effects	No	No	Yes	No	No	Yes
Sample	Baseline	Sibling	Sibling	Baseline	Sibling	Sibling

Coefficient (cluster robust standard error). Cluster defined by Grade 3 school. Significance: * p<.10, ** p<.05, ***p<.01. Sample includes black students entering 3rd grade in NC Public Schools from 2001 to 2005. Sample excludes 19872 students missing from public school data by 8th grade; 4394 students who exit NC school system for out-of-state schools, private schools, home schools, or death, excluded from NC cohort count; 3698 students missing own elementary teacher race composition in all years; and 14432 students missing clear indicators of either graduation or drop-out outcomes. Instrument is share of Grade 3-5 teachers in student's grade 3 school-cohort who share the student's demographic trait of interest. Student controls in all models include student sex and race controls, controls for whether students ever use subsidized lunch, are ever classified as Limited English Proficient, or ever are classified as having a disability. School controls are based on values for the student's 3rd grade year, and include average End of Grade score for the school, student-pupil ratio, logged student enrollment, share of students using subsidized lunch, share of students black, Hispanic, or other race, urbanicity, and charter/magnet status. Missing dummy variables included for all variables, except race match. FE=Fixed Effects.

Table 10. Descriptive Statistics for Black Students in Public-Use Project STAR Data

	All (1)
A. Educational Outcomes	
High School Dropout	0.34
High School Graduate	0.17
Took college entry exam (SAT/ACT)	0.48
B. Own Grade 3-5 Teachers	
1{Black T in initial STAR year}	0.44
C. Student Characteristics	
Persistently Low-Income	0.64
Older student	0.36
Male	0.45
Small class at entry	0.24
Regular class at entry	0.38
Regular class with aid at entry	0.38
N (Students)	1,549

Notes: Persistently low-income is defined as being eligible for free or reduced price lunch in each grade from K-3 that the student was observed in the data.

Table 11. Experimental Estimates of Long-Run Mismatch Effects from Project STAR

	OLS		FE Logit		Ordered Probit	
	(1)	(2)	(3)	(4)	(5)	(6)
Coefficient (HS Dropout)	-0.04	-0.03	-0.19	-0.16		
	(0.03)	(0.03)	(0.14)	(0.15)		
Coefficient (College Exam)	0.04	0.04	0.23	0.21		
	(0.02)**	(0.01)**	(0.09)**	(0.09)**		
Coefficient (O-Probit)					0.16	0.15
					(0.09)*	(0.08)*
APE (Dropout)	-0.04	-0.03			-0.05	-0.05
	(0.03)	(0.03)			(0.03)*	(0.03)*
APE (HS Diploma)					-0.00	-0.00
					(0.00)*	(0.00)*
APE (College Exam)	0.04	0.04			0.06	0.05
	(0.02)**	(0.01)**			(0.03)*	(0.03)*
Controls	No	Yes	No	Yes	No	Yes
N (Students)*	1,549	1,549	1,433	1,433	1,549	1,549

Notes: *The N for the college-exam linear and FE logit models are 4,166 and 4,042, respectively. Standard errors are clustered by school. The FE-logit samples are slightly smaller than the others because the FE-logit estimator drops observations from units (school-grades) that exhibited no variation in the outcome (Wooldridge 2010). The analytic samples are restricted to black students who participated in the Project STAR experiment. Random assignments were made in students' first year in STAR, and were intended to persist through grade 3. The independent variable of interest in these models is a binary indicator equal to one if the student was assigned to a black (same-race) teacher in his or her first year of STAR participation. Controls include indicators for the student's initial class type (small, regular, or regular + aide), age, sex, and poverty status. Poverty status is defined by an indicator for being persistently eligible for FRL in each year the student is observed in the STAR data. All Models control for school-by-grade of entry fixed effects (FE). * p<.10, ** p<.05, ***p<.01

Appendix Table A1. Bi-Probit IV Estimates of Long-Run Impact of Having a Black Teacher

	Drop-Out			Graduates and Intends 4-Year		
	All	Low-Income	Low-Income Males	All	Low-Income	Low-Income Males
	(1)	(2)	(3)	(4)	(5)	(6)
A. Bi-Probit First Stage						
% Cohort Teachers Black	2.27***	2.27***	2.21***	2.27***	2.27***	2.20***
	(0.09)	(0.11)	(0.14)	(0.09)	(0.11)	(0.14)
B. Bi-Probit Cofficients						
1{≥ 1 black teacher}	-0.13*	-0.16	-0.28**	0.10	0.30***	0.36**
	(0.07)	(0.10)	(0.13)	(0.06)	(0.10)	(0.17)
C. Bi-Probit APE						
1{≥ 1 black teacher}	-0.03*	-0.03	-0.07**	0.04	0.10***	0.11**
	(0.01)	(0.02)	(0.03)	(0.02)	(0.03)	(0.05)
Schools	1,261	1,176	1,125	1,261	1,172	1,125
Students (N)	105,150	47,897	22,748	103,688	47,161	22,748

Notes: Bi-probit estimates acknowledge the binary nature of the outcomes (high school dropout, college intent) and endogenous variable (exposure to at least one black teacher). Here, equation (1) and the underlying first stage are estimated jointly, as a system of two probit models with correlated error terms (Wooldridge 2010). * p<.10, ** p<.05, ***p<.01.

Appendix Table A2. Descriptive Statistics for Black Students (Dosage Sample)

	All Black Students			Persistently Low-Income Black Students		
	All	Male	Female	All	Male	Female
	(1)	(2)	(3)	(4)	(5)	(6)
A. Educational Outcomes						
H.S. Dropout	12.81	16.36	9.47	14.31	18.00	11.09
H.S. Grad, No College Intent	45.57	47.76	43.52	51.51	52.78	50.40
H.S. Grad, College Intent	41.74	35.96	47.14	34.33	29.36	38.65
Ever Suspended in H.S.	41.53	47.59	35.86	47.59	53.89	42.10
12th Grade GPA	-0.30	-0.51	-0.12	-0.42	-0.62	-0.25
	(0.97)	(0.94)	(0.95)	(0.94)	(0.91)	(0.93)
B. Own Grade 3-5 Teachers						
Exposure to ≥1 Black T	50.62	50.47	50.75	52.09	51.40	52.69
Number of Black Ts	0.78	0.78	0.77	0.81	0.81	0.82
	(0.92)	(0.92)	(0.91)	(0.94)	(0.95)	(0.93)
Number of Black Ts | ≥1	1.53	1.54	1.52	1.56	1.58	1.55
	(0.71)	(0.71)	(0.70)	(0.72)	(0.73)	(0.72)
C. Instrument						
% Cohort's Teachers Black	24.75	24.77	24.73	26.03	26.07	25.99
	(24.35)	(24.45)	(24.27)	(25.39)	(25.56)	(25.23)
[Within-school SD]	[8.49]	[8.47]	[8.51]	[8.68]	[8.63]	[8.72]
D. Student Characteristics						
Persistently Low-Income	45.04	43.36	46.61	100.00	100.00	100.00
Ever Low-Income	85.18	84.93	85.42	100.00	100.00	100.00
Ever LEP	0.23	0.27	0.19	0.28	0.37	0.20
Ever Exceptional	9.48	13.01	6.17	10.35	14.40	6.82
N (Students)	46,732	22,616	24,116	21,047	9,807	11,240

Notes: Standard deviations (SD) are reported in parentheses for non-binary variables. Persistently low-income is defined as being low-income in each of grades K-8. H.S. = high school. LEP = Limited English Proficient. T = teacher. H.S. dropouts are identified by the state as students who drop from enrollment in year *t* after having attended in year *t-1* and without graduating. Sample excludes students missing from public school data by 8th grade; students who exit NC school system for out-of-state schools, private schools, home schools, or death, excluded from NC cohort count; students missing own elementary teacher race composition in all years; and students missing clear indicators of either graduation or drop-out outcomes.

Appendix Table A3. Baseline Estimates of Exposure to a Black Teacher (Dosage Sample)

	All Black Students			Persistently Low-Income Black Students		
	All (1)	Male (2)	Female (3)	All (4)	Male (5)	Female (6)
A. First Stage						
% Cohort Teachers Black	0.90***	0.91***	0.89***	0.86***	0.87***	0.84***
	(0.04)	(0.05)	(0.05)	(0.05)	(0.06)	(0.06)
Schools	1,121	1,058	1,059	993	884	904
Students (N)	46,646	22,521	24,036	20,941	9,672	11,117
B. 2SLS: H.S. Dropout						
1{≥ 1 black teacher}	-0.05**	-0.14***	0.04	-0.06*	-0.16***	0.05
	(0.02)	(0.03)	(0.03)	(0.03)	(0.05)	(0.05)
Hausman Test *(p)*	0.01	0.24	0.00	0.86	0.50	0.16
Students (N)	46,646	22,521	24,036	20,941	9,672	11,117
C. 2SLS: College Intent						
1{ ≥1 black teacher}	0.01	-0.02	0.02	0.12**	0.08	0.11
	(0.03)	(0.04)	(0.05)	(0.05)	(0.07)	(0.08)
Hausman Test *(p)*	0.02	0.11	0.10	0.37	0.88	0.34
Students (N)	46,037	22,170	23,779	20,637	9,507	10,978

Notes: The analytic sample is restricted to students for whom the races of teachers in grades 3-5 are observed. Standard errors are clustered by the student's grade-3 school. The instrument in the 2SLS first stage is the share of Grade 3-5 teachers in the student's grade-3 school's cohort who are black. 1{.} is the indicator function. All models control for student characteristics including sex, "persistent" poverty status (in all grades 3-8), whether the student was ever classified as Limited English Proficient, and whether the student was ever classified as having a disability. All models control for school fixed effects and time-varying school characteristics (the year the student entered grade 3) including average End of Grade math and reading scores, student-pupil ratio, logged student enrollment, share of students using subsidized lunch, share of students black, Hispanic, or other race, urbanicity, and charter/magnet status. Panel A reports first-stage estimates for the H.S. drop-out regressions shown in Panel B. First-stage estimates are similarly strong for the H.S. grad and college intent estimates reported in Panel C. Hausman tests comparing the 2SLS estimates to analogous OLS estimates were conducted using control functions. * $p<.10$, ** $p<.05$, *** $p<.01$.

Appendix Table A4. Dosage Effects (Cubic in IV)

	All Black Students			Persistently Low-Income Black Students		
	All	Male	Female	All	Male	Female
	(1)	(2)	(3)	(4)	(5)	(6)
A. First Stages (F Statistic)						
Eq. 1: 1 Teacher Black	154.9	100.9	101.8	90.4	52.3	54.0
Eq. 2: 2 Teachers Black	56.2	34.3	56.2	30.1	15.0	31.6
Eq. 3: 3 Teachers Black	78.2	51.9	64.1	48.8	29.3	38.2
B. 2SLS: H.S. Dropout						
1{1 Teacher Black}	-0.03	-0.07	0.04	-0.09*	-0.16**	0.06
	(0.04)	(0.06)	(0.04)	(0.05)	(0.08)	(0.06)
1{2 Teachers Black}	-0.02	-0.13	0.02	0.02	-0.07	0.01
	(0.05)	(0.09)	(0.05)	(0.06)	(0.11)	(0.07)
1{3 Teachers Black}	-0.09*	-0.19**	0.05	-0.16*	-0.24**	0.11
	(0.06)	(0.08)	(0.07)	(0.08)	(0.10)	(0.11)
Joint Sig., Dosage Var. (*p*)	0.14	0.00	0.55	0.17	0.01	0.63
t test, 1 = 3 (*p*)	0.19	0.11	0.93	0.34	0.42	0.56
t test, 1 = 2 (*p*)	0.88	0.66	0.68	0.22	0.58	0.66
t test, 2 = 3 (*p*)	0.42	0.68	0.74	0.15	0.34	0.51
Hausman Test *(p)*	0.03	0.37	0.03	0.64	0.88	0.22
Students (N)	46,646	22,521	24,036	20,941	9,672	11,117
C. 2SLS: College Intent						
1{1 Teacher Black}	-0.013	-0.036	-0.008	0.062	0.007	0.017
	(0.050)	(0.072)	(0.065)	(0.071)	(0.095)	(0.104)
1{2 Teachers Black}	0.081	0.108	0.045	0.168	0.318*	0.100
	(0.079)	(0.123)	(0.094)	(0.103)	(0.170)	(0.123)
1{3 Teachers Black}	-0.056	-0.100	-0.020	0.075	-0.137	0.110
	(0.085)	(0.105)	(0.122)	(0.116)	(0.146)	(0.167)
Joint Sig., Dosage Var. (*p*)	0.77	0.76	0.97	0.10	0.18	0.62
t test, 1 = 3 (*p*)	0.55	0.52	0.91	0.90	0.34	0.50
t test, 1 = 2 (*p*)	0.39	0.41	0.69	0.47	0.17	0.65
t test, 2 = 3 (*p*)	0.35	0.29	0.74	0.61	0.07	0.97
Hausman Test *(p)*	0.10	0.05	0.30	0.66	0.69	0.49
Students (N)	46,037	22,170	23,779	20,637	9,507	10,978

Notes: This table is comparable to Table 3 of the main text. Standard errors are clustered by the student's grade-3 school. The instruments in the 2SLS first stages are the predicted probabilities that the student had one, two, and three black teachers in grades 3-5. Predicted probabilities were generated by poisson regressions of the number of black teachers actually experienced by the student on a cubic in share of Grade 3-5 teachers in the student's grade-3 school's cohort who are black. 1{.} is the indicator function. All models control for student characteristics including sex, "persistent" poverty status (in all grades 3-8), whether the student was ever classified as Limited English Proficient, and whether the student was ever classified as having a disability. All models control for school fixed effects and time-varying school characteristics (the year the student entered grade 3) including average End of Grade math and reading scores, student-pupil ratio, logged student enrollment, share of students using subsidized lunch, share of students black, Hispanic, or other race, urbanicity, and charter/magnet status. Panel A reports first-stage results for the H.S. drop-out regressions shown in Panel B. First-stage estimates are similarly strong for the H.S. grad and college intent estimates reported in Panel C. Hausman tests comparing the 2SLS estimates to analogous OLS estimates were conducted using control functions. * p<.10, ** p<.05, ***p<.01.

Appendix Table A5. Poisson Regressions of Counts of Students' Black Teachers on Fraction of Black Teachers in Students' School-Cohort

	All Black Students			Persistently Low-Income Black Students		
	All	Male	Female	All	Male	Female
	(1)	(2)	(3)	(4)	(5)	(6)
Fraction Teach. Black	1.892***	1.827***	1.952***	1.814***	1.701***	1.881***
	(0.056)	(0.081)	(0.080)	(0.083)	(0.124)	(0.115)
Mean Predictions (IV)						
1 Teacher Black	0.26	0.26	0.26	0.26	0.25	0.26
2 Teachers Black	0.12	0.12	0.12	0.12	0.12	0.12
3 Teachers Black	0.05	0.05	0.05	0.05	0.05	0.05
N (Students)	46,701	22,597	24,104	21,032	9,799	11,233

Notes: These are the First-Step poisson regressions used to make student-specific predictions that serve as instrumental variables (IV) in the estimation of the dosage models in equation (2), Table 3, of the main text. Standard errors are clustered by Grade-3 school. Student controls in all models include student sex and race controls, controls for whether students are persistently poor (use subsidized lunch in all years observed from grades 3-8), are ever classified as Limited English Proficient, or ever are classified as having a disability. School controls are based on values for the student's 3rd grade year, and include average End of Grade score for the school, student-pupil ratio, logged student enrollment, share of students using subsidized lunch, share of students black, Hispanic, or other race, urbanicity, and charter/magnet status. All models include fixed effects for 3rd grade school.
* p<.10, ** p<.05, ***p<.01.

Appendix Table A6. Extrapolating to Predict Exposure to Black Teachers in K-2

	All Black Students			Persistently Low-Income Black Students		
	All	Male	Female	All	Male	Female
	(1)	(2)	(3)	(4)	(5)	(6)
A. 0 Black Ts in Grades 3-5						
0 Black Ts in K-2	71.61%	71.16%	72.07%	80.56%	80.45%	80.67%
1 Black T in K-2	22.61%	22.36%	21.96%	16.06%	16.09%	16.03%
2 Black Ts in K-2	5.29%	5.50%	5.06	2.84%	2.87%	2.81%
3 Black T in K-2	.94%	.98%	.91%	.54%	.59%	.49%
	(30,204)	(15,225)	(14,979)	(12,502)	(6,252)	(6,250)
B. 1 Black T in Grades 3-5						
0 Black Ts in K-2	53.28%	52.46%	54.14%	71.04%	70.14%	71.98%
1 Black T in K-2	31.55%	32.06%	31.03%	22.53%	22.90%	21.74%
2 Black Ts in K-2	12.35%	12.32%	12.38%	5.77%	5.98%	5.55%
3 Black T in K-2	2.82%	3.16%	2.45%	.89%	.98%	.73%
	(14,828)	(7,537)	(7,291)	(7,279)	(3,695)	(3,584)
C. 2 Black Ts in Grades 3-5						
0 Black Ts in K-2	42.63%	43.10%	42.15%	62.63%	62.89%	62.36%
1 Black T in K-2	34.28%	33.85%	34.73%	26.45%	25.84%	27.09%
2 Black Ts in K-2	17.64%	17.59%	17.68%	8.99%	9.18%	8.79%
3 Black T in K-2	5.44%	5.47%	5.44%	1.94%	1.38%	1.76%
	(6,394)	(3,286)	(3,108)	(3,864)	(1,981)	(1,883)
D. 3 Black Ts in Grades 3-5						
0 Black Ts in K-2	34.93%	35.58%	34.11%	56.70%	57.50%	55.77%
1 Black T in K-2	36.62%	35.45%	32.08%	28.74%	28.17%	29.41%
2 Black Ts in K-2	22.57%	23.02%	22.02%	12.29%	12.09%	12.52%
3 Black T in K-2	5.88%	5.98%	5.79%	2.27%	2.24%	2.30%
	(1808)	(973)	(835)	(1,421)	(762)	(695)

Notes: Authors' calculations using NCERDC data for students observed in Kindergarten in school years 2007 & 2008. Persistently low-income is defined as being low-income in each of grades K-8. Sample sizes are different than those in Figures 1, 2, and Table 1 because this is an out-of-sample analysis to look at patterns of black teachers from grades K-5, however the results are similar.

Appendix Table A7. Characteristics of School in Year Student Entered 3rd Grade, by Number of Black Teachers Student Matched to in Grades 3-5

	Number Teachers Black Grades 3-5			
	0	1	2	3
Teacher Composition (% Black)				
Kindergarten	13.73	22.18	34.26	48.80
	(18.14)	(23.38)	(27.86)	(30.36)
Grade 1	14.98	23.42	35.44	52.46
	(19.00)	(22.77)	(26.77)	(29.59)
Grade 2	10.85	24.46	40.08	62.52
	(15.85)	(22.97)	(27.21)	(27.06)
Grade 3	13.33	26.63	43.58	64.18
	(19.18)	(25.14)	(30.50)	(30.29)
Grade 4	14.95	28.36	45.77	65.03
	(20.43)	(26.67)	(31.21)	(29.67)
Grade 5	14.27	21.78	33.42	48.33
	(18.40)	(22.65)	(27.13)	(29.99)
Student Body % Black	40.48	53.21	67.39	81.16
	(20.00)	(20.90)	(21.06)	(16.91)
School Average EOG	-0.06	-0.20	-0.35	-0.43
	(0.34)	(0.35)	(0.34)	(0.32)
Unique students	23,078	14,002	6,698	2,954

Authors' calculations from NCERDC Data. Variables captured represent composition in student's school during the year of 3rd grade entry.