ANNENBERG
BROWN UNIVERSITY

EdWorkingPaper No. 19-43

The Long-Run Impacts of Same-Race Teachers

Seth Gershenson
American University

Cassandra Hart
UC Davis

Joshua Hyman
University of Connecticut

Constance Lindsay
Urban Institute

Nicholas W. Papageorge
Johns Hopkins University

We examine the long-run impacts of having a same-race teacher. First, we leverage data from the Tennessee STAR class-size experiment to show that black students randomly assigned to a black teacher in grades K-3 are 5 percentage points (7%) more likely to graduate from high school and 4 percentage points (13%) more likely to enroll in college than their same-school, same-race peers not assigned to a black teacher. Second, we replicate these results in North Carolina using quasi-experimental methods. Finally, we formally define "role model effects" as information provision, which facilitates an exploration of possible mechanisms that drive these results.

VERSION: April 2019

Suggested citation: Gershenson, S., Hart, C., Hyman, J., Lindsay, C., & Papageorge , N.W. (2019). The Long-Run Impacts of Same-Race Teachers (EdWorkingPaper No.19-43). Retrieved from Annenberg Institute at Brown University: http://edworkingpapers.com/ai19-43

The Long-Run Impacts of Same-Race Teachers*

Seth Gershenson†
School of Public Affairs, American University and IZA

Cassandra M. D. Hart‡
School of Education, UC Davis

Joshua Hyman§
Department of Public Policy, University of Connecticut

Constance Lindsay¶
Urban Institute

Nicholas W. Papageorge‖
Department of Economics, Johns Hopkins University, IZA and NBER

ABSTRACT: We examine the long-run impacts of having a same-race teacher. First, we leverage data from the Tennessee STAR class-size experiment to show that black students randomly assigned to a black teacher in grades K-3 are 5 percentage points (7%) more likely to graduate from high school and 4 percentage points (13%) more likely to enroll in college than their same-school, same-race peers not assigned to a black teacher. Second, we replicate these results in North Carolina using quasi-experimental methods. Finally, we formally define "role model effects" as information provision, which facilitates an exploration of possible mechanisms that drive these results.

KEYWORDS: Teacher Effects, Racial Mismatch, Teacher Diversity, Educational Attainment.
JEL CLASSIFICATION: I2

*First draft: March 20, 2017. Current draft April 26, 2019. The authors thank Susan Dynarski and Diane Schanzenbach for generously sharing the Project STAR / National Student Clearinghouse linked data. We gratefully acknowledge helpful comments from Scott Carrell, Matt Chingos, Steven Durlauf, Dave Marcotte, and Robert Moffitt; conference participants at AEA, the HCEO Social Interactions and Education Conference, SOLE, SREE, AEFP, APPAM, North American Meetings of the Econometric Society, IRP Summer Workshop at the University of Wisconsin-Madison, International Workshop on Applied Economics of Education in Catanzaro, IT, and the 2nd Annual IZA Economics of Education Meeting; and seminar participants at the University of Michigan (Ford), Stanford University (CEPA), Princeton University (ERS), USC (Rossier), UC Irvine, UC Riverside, and UC Santa Barbara. Stephen B. Holt and Emma Kalish provided excellent research assistance. The usual caveats apply.

†[Corresponding author] gershens@american.edu.

‡cmdhart@ucdavis.edu.

§joshua.hyman@uconn.edu.

¶CLindsay@urban.org.

‖papageorge@jhu.edu.

1 Introduction

Racial gaps in educational attainment stubbornly persist, despite many resources being devoted to closing them. This is troubling for at least three reasons. First, schooling reduces inequality by facilitating upward socioeconomic mobility. It increases earnings, employment, and civic engagement and reduces criminal behavior, chronic illness, and dependence on social benefits.[1] Second, reducing education gaps can generate positive externalities by lowering costs associated with criminality or raising the productivity of the workforce. Third, attainment gaps might be driven by aspiration or information gaps, whereby students of color are less likely to aspire to attend college than their white peers, despite having sufficient ability to do so. If so, attainment gaps reflect sub-optimally low investments in human capital by children from low-income and historically marginalized backgrounds. Accordingly, reducing attainment gaps may not only increase equality across racial and socio-demographic groups, but could also lead to more efficient human capital investments.

We examine one factor that could reduce racial gaps in educational attainment, teacher race, which has been shown to affect short-run educational outcomes. Our focus is on long run outcomes. In particular, we study the impact of having a same-race teacher in elementary school on black students' high school graduation, college aspirations, and actual college outcomes. We identify arguably causal estimates by leveraging the random assignment of students and teachers to classrooms in the Tennessee STAR class size experiment.[2] Black students randomly assigned to a black teacher in grades K-3 are 5 percentage points (7%) more likely to graduate from high school and 4 percentage points (13%) more likely to enroll in college than their same-school, same-cohort peers who are not assigned a black teacher. The college enrollment results are consistent with similarly sized effects on the likelihood of taking a college entrance exam such as the ACT or SAT, though are primarily driven by enrollments in two-year community colleges. Rates of college degree receipt are low and possibly undercounted in the sample, so we are unable to precisely estimate impacts on college completion.[3]

Our results complement mounting evidence that same-race teachers are beneficial to minority students on a number of contemporaneous dimensions, such as test scores, attendance,

[1] See, e.g., Bailey and Dynarski (2011); Card (1999); Grossman (2006); Lochner and Moretti (2004); Moretti (2004a,b).

[2] We focus on black students because the teaching force in the U.S. is overwhelmingly white, so increasing same-race teacher assignments for white students is unlikely to improve their outcomes. We do not study Hispanic students because virtually no Hispanic students or teachers participated in the STAR experiment. However, the short- and long-run race-match effects for Hispanic students is an important, seemingly open question worthy of future inquiry.

[3] We return to this point in the Conclusion.

course grades, and beliefs in a variety of educational settings (Dee, 2004, 2005; Fairlie et al., 2014; Gershenson et al., 2016; Holt and Gershenson, 2017). Our findings are also consistent with well-established evidence that same-gender teachers and instructors affect educational outcomes, for example, encouraging women to go into STEM fields (Carrell et al., 2010). However, this literature focuses almost exclusively on short-run outcomes that are primarily of interest because they likely proxy for long-run outcomes of ultimate import, such as educational attainment.[4] Understanding whether race-match effects extend to long-run student outcomes is crucial for the design of appropriate policy interventions, including assessing the costs and benefits of increasingly urgent calls to diversify the teaching workforce. Our main contribution is to show that the benefits of same-race teachers for black students extend to long-run educational attainment and can thus contribute to the closing of stubbornly persistent attainment gaps.

More broadly, our results shed light on the well-documented importance of teachers. Indeed, teachers are among the most important school-provided inputs. Good teachers can improve students' test scores, non-cognitive skills, and long-run outcomes such as earnings and college going (Chetty et al., 2014; Jackson, 2018).[5] However, identifying effective teachers *a priori* is difficult and the channels through which teachers affect long-run outcomes remain unclear (Staiger and Rockoff, 2010). Teacher race is an interesting exception in that it is an observable characteristic that has potentially large impacts on student outcomes. Still, causal estimates of race-match effects do not pinpoint *why* same-race teachers boost the educational attainment of black students. The absence of evidence on the mechanisms through which same-race teachers improve black students' outcomes hinders policymakers' ability to effectively and efficiently respond to this information.

To explain our main empirical results, we first assess a straightforward possibility: that our estimates are not due to a race-match effect at all, but instead reflect that black teachers in schools serving black students are more effective teachers than their white counterparts. We reject this hypothesis, showing that random assignment to black teachers has no impact on white students. This is of policy relevance in its own right since it shows that white students would not be hurt if they faced a more diverse teaching force and thus fewer white teachers.[6] It also suggests that to explain our estimates, race match is key. But why?

[4] An exception in the context of gender is Lim and Meer (2017), who show that effects of gender match in the 7th grade persist through high school.

[5] More generally, our findings contribute to growing evidence that primary-school inputs over and above teachers can have impacts on long-run socio-economic outcomes, such as the number of disruptive peers (Carrell et al., 2018), class size (Dynarski et al., 2013), and general classroom quality (Chetty et al., 2011).

[6] This finding does not necessarily hold across all educational contexts. Lusher et al. (2018) find that white students obtain higher grades when matched to white college graduate student teaching assistants.

To understand what drives same-race teacher effects, we begin with the idea that schools and teachers play an important role not only in teaching and developing students' skills, but in forming students' attitudes, beliefs, and the choice sets that they optimize over. This is likely especially true for students from historically disadvantaged backgrounds and children growing up in socio-economically disadvantaged neighborhoods, as within schools teachers play an outsize role in establishing school culture and may be some of the only college-educated professionals that many students — especially those from low-income backgrounds — interact with on a daily basis.

Accordingly, we develop a simple model of human capital investments that distinguishes between two potential mechanisms, both rooted in an established literature on *culturally relevant pedagogy* (Ladson-Billings, 1995), which examines the importance of black teachers for black students. First, black teachers could be more effective at transmitting content to black students, perhaps because they possess cultural competence. They may be more likely to come from a similar background or community and can thus understand the unique challenges minority students face or provide more compelling or relevant examples to communicate concepts. An alternative mechanism is similar to what is often deemed a *role model effect*. The idea here is that too many black students in segregated neighborhoods rarely come into contact with educated, professional black individuals (Wilson, 1987). If so, teachers provide important "counterexamples" that might shift a student's own view of what is possible (Kelly, 2010).

To formalize these mechanisms, we posit an education production function with two inputs and two periods of investment, where student investments (summarized as "effort") and teacher quality determine achievement. To capture the effectiveness mechanism, we allow black teachers to raise achievement more than white teachers. To capture role model effects, we envision black students as having incorrect prior beliefs about the education production function, in particular, by underestimating the returns to effort. A black teacher is a role model in the sense that she provides a "signal" to students about the true returns to effort that leads students to update their beliefs and increase their effort (investment).

By incorporating both potential mechanisms, the model generates testable implications: if the productivity argument holds, having two black teachers should raise attainment more than having a single black teacher. Alternatively, if role model effects are key, it is possible that the first teacher leads to the crucial update in beliefs so that the marginal impact of a second is diminished. To test these hypotheses, we estimate the marginal effect of being assigned to a second black teacher, i.e., a dosage model. We show that the impact of a second black teacher is essentially zero for certain outcomes, such as taking a college entrance exam, which is consistent with the view that role model effects are a key mechanism undergirding race-match effects on some dimensions of educational attainment. In contrast, on other dimensions, such as college enrollment, we find large impacts of a second black teacher. To explain these findings, we speculate that a single black teacher might be enough to inspire students via role model effects to take actions consistent with higher education, such as sitting for a college entrance exam. However, a second black teacher could further affect costlier actions, such as enrollment, if each black teacher more effectively delivers content to black students.

We further test the model's implications using rich, longitudinal, administrative data on the population of North Carolina public school students in conjunction with the Tennessee STAR data, as the STAR data are limited in terms of power and external validity. We find similar results of race-match using a quasi-experimental strategy and the universe of elementary school students in North Carolina, bolstering the external validity of our main finding that exposure to a same-race teacher in elementary school significantly improves the long-run educational outcomes of black students. Moreover, the North Carolina data provide the power necessary to conduct two important heterogeneity analyses that provide additional insights into the mechanisms through which same-race teachers increase the college aspirations of black students. First, we show that these effects are entirely driven by the response of persistently disadvantaged students, who are less likely to interact with professional and college-educated black adults on a regular basis. Second, we document stronger effects of black male teachers on black male students, and of black female teachers on black female students. Both of these patterns are in line with the notion of role model effects playing a nontrivial role in the race-match effects on black students' college aspirations.

While multiple black teachers affect outcomes on some dimensions, exposure to one black teacher can have a large impact, which is important since black teachers are scarce — and likely will be for some time (Putman et al., 2016). Moreover, the importance of role model effects along some dimensions (e.g., in inspiring students to take initial steps towards attending college by taking entrance exams) underscores the idea that existing gaps in educational attainment might reflect a market failure, one that leads to inefficiently low investments in human capital for minority children arising from incorrect beliefs about the education production function (Cunha et al., 2013). Evidence of a market failure provides support for the productivity argument for investing in poor children (Heckman, 2006; Heckman and Masterov, 2007). Efforts to close such gaps, such as raising the likelihood that black students are exposed to a black teacher, would not only reduce racial inequality, but also lead to more efficient human capital investments.

The paper proceeds as follows. Section 2 describes the STAR data and associated analyses. Section 3 presents the main STAR results. Section 4 introduces an education production function that includes role-model effects and tests of other pathways through which race-match effects may operate. Section 5 describes the North Carolina data and associated analyses. Section 6 concludes.

2 Data and Methods

This section describes the experimental data and methods used to identify the long-run impact of random assignment to a same-race teacher on educational attainment. Section 2.1 describes the STAR experiment and section 2.2 describes the NSC data on college enrollment and completion. Section 2.3 summarizes the publicly available STAR data and the linked NSC data. Section 2.4 describes the identification strategy.

2.1 Project STAR

Tennessee's Project STAR (Student Teacher Achievement Ratio) was a seminal field experiment in education, 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 in relatively disadvantaged schools to either small- or regular-sized classrooms, with some of the regular-sized classrooms having a teacher's aide. Participation in STAR was voluntary at the school level and no one was randomly assigned to schools, so it was purely a within-school experiment. Students assigned to a particular treatment arm, say small class, were intended to receive that treatment for the duration of the experiment (through third grade). Furthermore, over the next three years, new entrants to the STAR cohort in STAR schools were added to the experiment. 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 and Whitmore, 2001) and of college enrollment and completion (Dynarski et al., 2013). These long-run effects are also larger for black students.

Dee (2004) recognized that STAR's random assignment of teachers and students to classrooms created exogenous variation in students' exposure to same-race teachers. 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 of all students, and particularly large effects for black students. Penney (2017) updates this work by testing for dosage and timing effects of exposure to same-race teachers and finds some modest evidence that earlier exposure is better and that dosage effects are fairly small. Chetty et al. (2011) similarly leverage Project STAR's randomization to estimate long-run effects of teacher and peer quality during kindergarten 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.[7] We extend this prior work to estimate the effects of having a same-race teacher in early elementary school on long-run educational outcomes. We do so using publicly available Project STAR data, which includes information on high school graduation, whether students took a college-entrance exam (i.e., ACT or SAT), and a host of intermediate outcomes measured in middle school, together with data on post-secondary educational enrollment and attainment from the National Student Clearinghouse (NSC) collected by Dynarski et al. (2013).

2.2 National Student Clearinghouse Data

Data on postsecondary outcomes come from the National Student Clearinghouse (NSC). The NSC is a non-profit organization and the only nationwide source of administrative data on student-level postsecondary enrollment and degree completion. Participating colleges submit enrollment data to NSC several times each academic year, reporting whether a student is enrolled, at what school, and at what intensity (e.g., part-time or full-time). The NSC also records degree completion and the field in which the degree is earned. Dynarski et al. (2015) provide a thorough discussion of the NSC, its origins, matching process, and coverage rates.

To examine the effects of class size on postsecondary outcomes, Dynarski et al. (2013) submitted the STAR sample to the NSC in 2006 and again in 2010. The NSC then matched individuals in the STAR sample to its database using name and birth date. The STAR sample was scheduled to graduate high school in 1998, so these data capture college enrollment and degree completion for twelve years after on-time high school graduation, when the STAR sample is about 30 years old. One key advantage of the NSC data is that, because it is matched using students' identifying information collected at the time that students entered the STAR experiment, it is available even for those students who attrit from the STAR sample. While the NSC data provide valuable insights into postsecondary educational attainment, a few limitations of the data merit further discussion.

First, the NSC-STAR matching was not perfect. About twelve percent of students in the

[7] 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.

STAR sample have incomplete name and/or birth date information that reduces the chance of making a match (Dynarski et al., 2013). Because a student who attended college but did not produce a match in the NSC database is indistinguishable from a student who did not attend college, such mismatches could bias our estimates if missing name and/or birth date information is correlated with initial assignment to a black teacher. Accordingly, we add an indicator variable equal to one if a student has a missing name or date of birth, and zero otherwise, to the balance tests presented in section 2.4. Consistent with Dynarski et al. (2013), we find small, statistically insignificant differences, indicating that the probability of missing identifying information is uncorrelated with being initially assigned a black teacher.

Second, not all schools participate in NSC. Today, the NSC estimates that they capture about 97% of undergraduate enrollment nationwide. During the late 1990s, however, when the STAR subjects would have been graduating from high school, the NSC included colleges enrolling about 80% of undergraduates in Tennessee (Dynarski et al., 2015). Dynarski et al. (2013) compare the mean college enrollment rate in the STAR-NSC sample to that of a sample of Tennessee-born individuals from the 2005 American Community Survey (ACS), and show that, as expected, the enrollment rate is about 20% lower in the STAR-NSC data than in the ACS. Dynarski et al. (2013) also find that the rate of degree receipt in the STAR-NSC data is even lower than 80% of the rate found in the ACS. This is likely because degree receipt is underreported in the NSC, as not all colleges that report enrollment to the NSC report degree receipt (Dynarski et al., 2015). For this reason, and because degree completion rates of the black students in our sample are so low that we are underpowered to examine it, we focus on college enrollment, not college degree receipt, as our primary measure of educational attainment. Further, we also consider SAT/ACT exam taking, which is not subject to these concerns, as a measure of college intent and a proxy for college enrollment.

Finally, the exclusion of colleges from the NSC will cause measurement error in the dependent variable. If this error is independent of treatment (i.e., classical measurement error), then the true effect of being assigned a black teacher will be larger than our observed effect by the proportion of enrollment that is missed (approximately 20%). However, if the measurement error in college enrollment is correlated with black teacher assignment, then our effect could be either upward or downward biased. This could be the case, for example, if colleges attended by marginal students are disproportionately undercounted by NSC.

Dynarski et al. (2013) and Dynarski et al. (2015) compare the schools that participate in the NSC with those in the Integrated Postsecondary Education Data System (IPEDS), which is a federal database that includes the universe of postsecondary institutions. Those studies find that along multiple measures, such as sector, racial composition, and selectivity, the NSC colleges are similar to the universe of IPEDS colleges, with one notable exception:

the NSC tends to exclude for-profit institutions. If assignment to a black teacher causes black students who would not otherwise attend college to systematically enroll in for-profit schools, we will underestimate the effect of black teacher assignment on college attendance. Alternatively, if black teacher assignment induces students out of such schools into colleges that are in the NSC data, such as community colleges, then our estimates will be upward biased. Dynarski et al. (2013) conduct a back-of-the-envelope exercise to bound the possible upward bias attributable to this phenomenon, and find that any likely upward bias is small. Using the same procedure, we find the same result in our context: any upward bias is capped at 0.3 percentage points, accounting for only 7.5% of our estimated effect.

2.3 STAR Data

Table 1 summarizes the main analytic sample of black students who participated in Project STAR. Panel A of Table 1 summarizes students' baseline characteristics. Column 1 shows that about half of the approximately 4,000 black STAR students were male and about 80% were eligible for free- or reduced-price lunch (FRL) in at least one school year. Another baseline characteristic is an indicator for whether the student's name or date of birth (DOB) was missing. This variable is important because it is the identifier used to match the STAR and NSC data (Dynarski et al., 2013). Overall, about 12% of the sample has a missing name or DOB. Columns 2 and 3 split the sample by whether students had a black teacher in their first year in STAR, and as expected due to random assignment we see balance in sex, FRL, and the indicator for missing name or DOB. We present more formal balance tests in section 2.4, which are consistent with the results in Dee (2004). Columns 4-7 show sample means across the four cohorts of students entering the STAR sample. Cohort 1, who entered in kindergarten, was the largest cohort, comprising about half the analytic sample. The socio-demographic composition across the cohorts is about the same, though having a missing name or DOB is most common in cohort 1.

Panel B of Table 1 summarizes the classroom and school characteristics. The independent variable of interest, exposure to same-race (i.e., Black) teachers, is simply a binary indicator equal to one if the student was assigned to a black teacher in the child's first year in a STAR school. We focus on this definition because compliance with the random assignments was essentially perfect in the student's first year (Dee, 2004; Krueger, 1999) and is observed for all students, regardless of whether they later attrited from the experimental sample. Overall, 44% of the analytic sample had a same-race teacher, and the likelihood of having a black teacher increased slightly in later cohorts.

Column 1 further shows that about one quarter of students were in small classrooms.

The average class had about 21 students and was 80% black. About one third of teachers had at least a Master's degree and the average teacher had about ten years of experience. The average student was in a grade in which 43% of the teachers were black and about 67% of students were in "poor schools," which we define as schools in which at least 80% of the students were FRL, which roughly amounts to the bottom tercile of the distribution.

Columns 2 and 3 of panel B highlight some slight differences between the classrooms and schools in which students were exposed to a same-race teacher. Black teachers were less likely to be in small classes, taught in classes with more black students, were less likely to hold an advanced degree, and had more years of experience than their white counterparts. Intuitively, these differences represent the fact that black teachers taught in schools that had more black students and do not reflect a failure of random assignment of black students to black teachers within schools.

Finally, panel C summarizes several long-run educational outcomes of interest. The public-use STAR data provide indicators for whether the student took the SAT or ACT college entrance exams, which are indicators of college intent, and for high school graduation. Like the NSC data, the college entrance exam data is available regardless of whether students attrited from the experimental sample (Krueger and Whitmore, 2001). Just over a quarter of students in the sample sat for a college entrance exam. The average SAT score among test takers was about 777, though assessing the effect of same-race teachers on SAT scores requires correcting for selection into the test, which we discuss below.[8] Unfortunately, high school graduation information is only observed for 37% of the sample. The empirical analysis deals with this data problem by using both bounding and multiple imputation procedures.

The postsecondary educational outcomes, from the NSC data described in section 2.2 and Dynarski et al. (2013), show that about one third of the analytic sample enrolled in some type of college. Of those who enrolled in college, more than half first enrolled in a two-year college, one third enrolled in a four-year public college, and the balance enrolled in four-year private colleges. More than two thirds of these students persisted into year two, though less than one third completed a degree (in schools that reported degree attainment to the NSC). Column 4 shows that postsecondary outcomes of cohort 1 (kindergarten entrants) tend to be better than those of later entrants, which is unsurprising given that kindergarten was optional at this time and that cohorts 3 and 4 necessarily switched schools early in elementary school. These differences highlight the importance of conditioning on cohort of entry in addition to school (Krueger and Whitmore, 2001).

[8]ACT test takers' scores were converted to the SAT scale.

2.4 Identification Strategy

Project STAR targeted disadvantaged schools and made random *within-school* assignments of students and teachers to classrooms. Students and teachers are not randomly distributed across schools, of course, so all analyses condition on school-by-cohort fixed effects to account for systematic unobserved differences in the levels and trends of schools' performance and resources (Krueger and Whitmore, 2001). Using the limited pre-experiment data available on students, previous research has documented good balance between students assigned to small- and regular-sized classrooms and between black students assigned to same- and different-race teachers (Dee, 2004; Dynarski et al., 2013; Krueger, 1999). Chetty et al. (2011) use linked IRS earnings data for parents to provide even more convincing balance tests.

In Table 2, we perform within-school balance tests for students in our main analytic sample of black students in their first STAR year. These balance tests follow those in Dee (2004) and simply regress an indicator for assignment to a black teacher on indicators of student gender, FRL status, a small-class indicator, and a "missing name or DOB" indicator. Column 1 conducts a naive balance test that does not adjust for school-by-cohort FE for the full sample. While not statistically significant at traditional confidence levels, this test shows that ever-FRL students, students in small classes, and students whose name or DOB were missing were less likely to have a same-race teacher. However, when the school-by-cohort FE are included in column 2, the coefficients and standard errors fall by about half. This proper "within-school" balance test yields fairly precise zeros, suggesting that within-school random assignment was achieved, at least in students' first year in STAR. Columns 3-4, and 5-6, show similar patterns for black boys and black girls, respectively. This is consistent with Krueger (1999), who reports that an audit of school records for a subset of STAR schools finds that nearly all students complied with their classroom assignment in their first STAR year, as well as with the balance tests conducted by other researchers (Chetty et al., 2011; Dynarski et al., 2013; Krueger, 1999).

Thus, compliance is not an issue in students' first year in STAR. Accordingly, the simplest way to identify the causal effect of having a same-race teacher is to make within-school comparisons between the long-run outcomes of students who did have a same-race teacher in their first STAR year and those of students who did not. We make these comparisons by estimating linear models of the form

$$y_{ijk} = \theta_{jk} + \beta X_i + \delta Black_i + u_{ijk}, \tag{1}$$

where i, j, k index students, cohorts, and schools, respectively; θ is a school-by-cohort fixed effect (FE), X is a vector of observed student, first-year teacher, and first-year classroom characteristics including student's sex and FRL status, teacher's sex, experience, certification, degree attainment, and career ladder stage, and class type; *Black* is a binary indicator equal to one if the student had a black teacher in his or her first year of STAR. The parameter of interest is δ, which represents the causal effect of assignment to a same-race teacher in students' first year of STAR.[9] We estimate equation (1) by OLS, though verify below that the results are robust to specifying analogous FE-logit models instead. Other extensions of equation (1) include augmenting the model to include interaction terms that allow the effect of race match to vary by cohort and by student and teacher background, and by replacing the school-by-cohort FE with classroom FE and comparing white and black students in the same classroom. We cluster standard errors by classroom, as this is the level at which random assignments were made (Abadie et al., 2017), though this decision proves inconsequential.

3 STAR Results

This section reports the empirical analyses of the Project STAR data. Section 3.1 presents and probes the main finding that random assignment to a same-race teacher in primary school significantly increases the likelihood that black students ultimately enroll in college. Section 3.2 considers other educational outcomes such as high school graduation, taking a college entrance exam, the type of college or university that students first enroll in, persistence in higher education, and degree attainment. Finally, section 3.3 examines possible heterogeneity in the long-run effects of having a same-race teacher.

3.1 Main Results

Table 3 reports baseline estimates of equation (1) for college enrollment, the main long-run outcome of interest. These results provide the cleanest evidence of a causal relationship between exposure to a same-race teacher in elementary school and students' long-run educational attainment, as there are no concerns about compliance or attrition: classroom assignments were random in students' first STAR years and administrative college enrollment data are observed for nearly all STAR participants (Dynarski et al., 2013). Columns 1

[9]Note that this model is explicitly comparing students who did and did not have a black teacher *in their first STAR year*; it says nothing about having black teachers in subsequent years. If even a single exposure to a black teacher is what matters, these estimates will understate the effect of having a same-race teacher because the "control group" will possible have a black teacher in later years. Specifically, 38% of control group students have a black teacher in year two of STAR. Treating this as "noncompliance," we could scale the estimate as $\frac{\delta}{0.62}$ and give it a LATE interpretation. Doing so would scale up our baseline estimate of 0.040 to 0.064. Intuitively, this puts the baseline estimate closer to the estimated effect of having *at least one* black teacher, results which are discussed in section 4.3.2.

and 2 of Table 3 restrict the sample to the inaugural kindergarten cohort, and therefore estimate the long-run impact of being randomly assigned a black teacher in kindergarten. We do so because the STAR experimental randomization is cleanest for kindergartners (Krueger, 1999).[10] The estimated impact is between 4.7 and 6.2 percentage points, marginally statistically significant, and robust to controlling for observed teacher and student characteristics. The robustness to student characteristics is to be expected given the random assignment of students to classrooms. The robustness to teacher characteristics suggests that the result is not driven by within-school racial differences in observable measures of teachers' effectiveness (e.g., experience (Wiswall, 2013)). From a base enrollment rate of 34% among non-matched students, the 4.7-6.2 percentage point effect constitutes a nontrivial 14-18% increase in the likelihood of ever enrolling in college.

Columns 3 and 4 of Table 3 repeat this exercise using data for all STAR entry cohorts. The main benefit of pooling the cohorts is to increase power. Again, the point estimate is robust to adding controls and is similar in size to the kindergarten-only estimates: random assignment to a black teacher in your first STAR year increases the likelihood of enrolling in college by about 4 percentage points, or 13%, off the base of 31% among non-matched students. Of course, pooling all four cohorts in a single regression begs the question of whether the impact varies by cohort. We address this in columns 5 and 6, interacting the *Black* indicator with a set of cohort-entry indicators. The interaction terms are individually and jointly insignificant in both columns, suggesting that the race-match effect does not vary by STAR cohort. Accordingly, our preferred baseline models in subsequent analyses pool the cohorts to increase power, condition on observed student and teacher controls, but do not include the cohort-match interaction terms.

In sum, the results presented in Table 3 provide credible evidence that same-race teachers have long-run impacts on black students' educational outcomes. There are two reasons for this. First, college enrollment data are observed for nearly all students, even if they left the STAR experiment or left the Tennessee public school system, because they were obtained from the NSC. Second, there are neither compliance nor attrition concerns in students' first year in a STAR school. Because these results provide proof of concept for what follows, we now conduct a series of sensitivity analyses that verify the robustness of the results.

Overall, the main results in Table 3 prove to be quite robust. Table 4 presents a potpourri of sensitivity analyses. For reference, column 1 reproduces the baseline result, previously reported in column 4 of Table 3, which uses the full pooled sample and the full set of teacher and student controls. Column 1 also reports different plausible standard errors, clustered

[10]Moreover, Ding and Lehrer (2010) question whether later STAR entrants were randomly assigned, though we find no evidence that this is an issue for black students and their assignment to black teachers.

at different levels including school-cohort and school, which are all similar in size and yield similar statistical inferences. We prefer the classroom-clustered standard errors, as this is the level of random assignment and the level at which treatment varies (Abadie et al., 2017). In column 2, we re-estimate the model on the selected sample of students for whom name and DOB were observed, as students whose name and/or DOB were missing might have enrolled in college but been coded as non-enrolled due to a failed NSC match. The main result is robust to excluding students whose enrollment outcomes might be incorrectly measured, reducing concerns that the imperfect coverage of STAR students in the NSC data drives the results.[11]

Column 3 of Table 4 estimates the FE-Logit analog of equation (1) to account for the binary nature of the dependent variable (Chamberlain, 1980). The FE-Logit coefficient estimate remains positive and statistically significant. Its magnitude cannot be directly interpreted nor can proper average effects be computed, but an approximate scale factor can be computed using the sample average of Y.[12] Using this approximate scale factor, the implied partial effect is 0.046, which is slightly larger, though similar in size, to the baseline LPM point estimate of 0.040 reported in column 1. Column 4 re-estimates the baseline LPM using the FE-Logit's analytic sample, as the FE-Logit's MLE routine drops observations from school-cohorts that exhibited no variation in the outcome. The resultant point estimate is identical to that in the full sample, which is unsurprising given that only about 100 students were dropped.

Finally, columns 5 and 6 of Table 4 add additional classroom-level controls to the baseline model. Column 5 replaces the class-type indicators with an exact count of class size. Because class size might not be random, we follow Krueger (1999) in instrumenting for size with the randomly assigned class-type indicators. The resultant 2SLS estimate of the impact of having a black teacher on college enrollment is essentially unchanged. Similarly, column 6 adds to the baseline model a control for the share of the class that is black. Again, this is a check of random assignment, as it could be that the racial composition of a classroom jointly affects the likelihood of having a black teacher and later educational outcomes. Here too, however, we see that the baseline estimate is unchanged by adding this control. Together, the sensitivity analyses reported in Table 4 further bolster the causal interpretation of the main finding in Table 3: black primary school teachers significantly increase the chances that black students eventually enroll in college.

[11]Moreover, Appendix Table A1 reproduces the battery of robustness checks reported in Table 4 for the selected sample of students whose name and DOB are observed. Once again, we find that the main result is robust to a variety of modeling choices among this selected sample.

[12]The sample mean of enrollment is $p = 0.325$, so the approximate scalar is $p(1 - p) = 0.219375$.

3.2 Other Educational Outcomes

Table 5 estimates the baseline model (equation 1) for several long-run educational outcomes over and above college enrollment. Column 1 replaces college enrollment with an indicator for whether the student took the ACT or SAT college entrance exam. Taking a college entrance exam is interesting in its own right, as it indicates college intent during the student's junior or senior year of high school.[13] College intent is a particularly relevant outcome for economically disadvantaged students who comprise the majority of the STAR experiment's black student population, potentially distinct from actual enrollment, as the phenomenon of "summer melt" suggests that anywhere from 8 to 40% of high school graduates who intend to enroll in college at the time of graduation fail to do so (Castleman and Page, 2014). Moreover, the entrance-exam data can cross validate the NSC data by providing information on college intent obtained from an independent source (Krueger and Whitmore, 2001). The point estimate of 0.043 is essentially identical to the baseline result for enrollment, suggesting that the race-match effect on college intent or aspirations matches closely with the eventual effect on actual enrollment.[14]

The remainder of Table 5 exploits the wealth of information over and above college enrollment maintained by the NSC. Columns 2-4 disaggregate the main enrollment effect by the type of postsecondary institution in which students first enrolled. Here, while we see positive coefficients for each type of institution, the main effect was driven by community college enrollments. This is intuitive, as community colleges are the most likely landing spots for students on the margin of entering postsecondary education. Finally, columns 5 and 6 address students' persistence in postsecondary education. Column 5 shows a slightly smaller, marginally significant effect on persistence, which is a standard measure of whether first-time college goers return to campus in the fall of their second year. This is reassuring, as the main enrollment effect was on enrollment in two-year community colleges where the majority of dropouts occur during the first year and summer before the second year (Ishitani and DesJardins, 2002). Unfortunately, we are under-powered to examine postsecondary degree receipt (column 6). We find a near-zero, statistically insignificant effect on degree receipt. However, given the very low rates of degree completion among non-matched students (8.5%),

[13]Indeed, prior to obtaining the NSC data on enrollment, this was the principal outcome of interest in an earlier version of this paper (Gershenson et al., 2017).

[14]That having a same-race teacher affects the likelihood of taking a college-entrance examination raises the question of whether there are analogous effects on entrance exam *performance*. Appendix Table A2 examines this question by estimating variants of equation (1) that take ACT scores (or converted SAT scores) as the dependent variable. We account for the selection problem caused by missing exam scores for non-test-takers by using both Heckit selection corrections (Heckman, 1979) and Lee (2009) bounds. Both naive OLS and the Heckit procedure yield small, negative, statistically insignificant point estimates of about 0.2 points (1%) that are fairly precise zeros. The Lee Bounds include zero and the OLS and Heckit estimates.

we cannot rule out degree receipt effects on the order of 1 or 2 percentage points. Effects in that range would suggest the marginal matched student induced into college persisting to degree receipt at around the same rate as the inframarginal, non-matched student.[15]

Finally, Table 6 examines the effect of having a same-race teacher on high school graduation. This analysis is hindered by data availability, as high school graduation data are missing for more than half (63%) of the analytic sample, so the results should be interpreted with a healthy dose of caution.[16] Indeed, column 1 takes a sample-selection indicator as the outcome of the baseline model specified in equation (1) and shows that random assignment to a black primary school teacher significantly increases the likelihood that a student's high school graduation data was recorded in the Project STAR database. Intuitively, this positive selection into the sample is consistent with the positive impacts on enrollment documented thus far, as the presence of graduation data suggests some degree of attachment to the public school system. In this sense, the positive selection observed in column 1 provides yet another instance of random assignment to a same-race teacher positively affecting long-run educational outcomes. Nonetheless, we are interested in the long-run effects of same-race teachers on the high school graduation margin too, as a nontrivial share of economically disadvantaged black students in Tennessee in this era were closer to the high school graduation margin than to the college enrollment margin.

To show that the selected sample's education production function is not too different from that of the full analytic sample, in column 2 we estimate the baseline college enrollment model on the selected sample and find a nearly identical, albeit less precise, point estimate. This suggests that the returns to having a same-race teacher are similar for students whose high school graduation status was and was not observed.

Accordingly, we proceed in columns 3 and 4 by estimating LPM and FE-Logit models, respectively, for high school graduation on the selected sample. These are naive estimates in the sense that no correction for sample selection is made. Both find marginally significant effects of random assignment to a same-race teacher on the likelihood of high school graduation. The LPM estimate of about 0.05 suggests a 7% increase in the likelihood of graduating high school. Columns 5 and 6 show that these estimates are robust to instead using a

[15]Only 28% (=0.0868/0.314) of college enrolling black students with white teachers earn a degree. Our estimated effect on enrollment is 4.0 percentage points. Thus, if the marginal student induced into college by having a black teacher persisted to degree receipt at the same rate as their inframarginal, non-matched peer, the estimated effect on degree receipt would be 1.1 percentage points (=0.04*0.28). We cannot reject that our coefficient on degree receipt of half a percentage point is equal to 1.1 percentage points.

[16]Appendix Table A3 summarizes the basic student data by high school status. Unsurprisingly, students for whom high school records are missing are systematically worse off in terms of both baseline and long-run outcomes. This is likely why previous long-run analyses of STAR's small class size reductions on educational attainment do not investigate high school graduation (Dynarski et al., 2013; Krueger and Whitmore, 2001).

multiple imputation procedure to impute the missing high school graduation outcomes.[17]

Of course, multiple imputation does not eliminate selection bias if the dependent variable is not missing at random, so we also implement a bounding-type procedure in which we re-estimate the baseline model 11 times, each time replacing the missing graduation values with a different graduation probability between 0 and 1. The resultant point estimates and 95% confidence intervals are plotted in Figure 1. Three features of Figure 1 are consistent with a positive, arguably causal effect of having a same-race teacher on the likelihood of completing high school. First, the point estimates are universally positive and intuitively the estimates at each end of the horizontal axis provide extreme bounds of the effect. Second, at assumed (and plausible) graduation rates of 70% and below for those whose graduation records are missing, which includes the selected sample mean graduation rate of 67%, the estimated effect of race-match is statistically significantly different from zero. Finally, the magnitude of the point estimates in Figure 1 are consistent with those reported in Table 6, ranging from 0.02 to 0.05 over the range of plausible graduation rates for students whose graduation data are missing (i.e., 0.0 to 0.70). This is the plausible range because we suspect positive selection into the sample and those whose graduation records are missing are less likely to have graduated than those whose records are observed, and the graduation rate in the selected sample is 0.67. In sum, Table 6 and Figure 1, taken together with the results for college enrollment presented thus far, strongly suggest that exposure to a same-race teacher increases the probability of graduating from high school. We revisit this question in section 5 using administrative data from North Carolina that are not prone to the missing data problems that plague the STAR high school graduation records.

3.3 Heterogeneity

Table 7 reports estimates of augmented versions of equation (1) that interact the black-teacher indicator with observed student, classroom, and school characteristics. Columns 1 and 2 test whether the effect varies by sex and socioeconomic status (i.e., free lunch receipt), respectively, and find no evidence of heterogeneity along these dimensions. Column 3 tests whether the effect varies by class size, as teachers may form closer personal relationships with students in smaller classes. As in column 5 of Table 4, we follow Krueger (1999) in instrumenting for size and the match-size interaction with the randomly assigned class-type indicators and the interaction of those indicators with the race-match indicator. Here too, the 2SLS estimate suggests that the race-match effect does not vary by class size. Similarly, column 4 finds no evidence that the long-run match effect varies by racial composition of

[17] We use a probit formulation of equation (1) and 40 imputations to construct these estimates.

the classroom. Following Dee (2004) and Dynarski et al. (2013), column 5 tests whether the effect is stronger in more disadvantaged schools, where school disadvantage is proxied by a binary indicator equal to one if more than 80% of students are eligible for free lunch, and zero otherwise.[18] Once again, the interaction term is small and statistically indistinguishable from zero. Finally, column 6 tests whether the match effect is a function of the number of other black teachers in the same grade. The motivation for this is that students likely interact with, or at least see, the other same-grade teachers in the school and these teachers might similarly act as role models, mentors, and sounding boards. Here too, the interaction term is small and statistically insignificant.

In sum, the results presented in Table 7 provide no evidence that the long-run effects of having a same-race teacher vary by the observable student, classroom, or school characteristics available in the STAR data.[19] However, this could be driven by either the relatively small STAR samples and accompanying lack of power to detect small heterogeneities, the general lack of variation in student background in the STAR sample, which was purposely composed of disadvantaged schools, or by the relatively crude student-level data available in the STAR data. We address these questions in section 5 when we return to the question of heterogeneous effects using administrative data for the population of North Carolina public school students.

4 Mechanisms: Effectiveness and Role Models

This section presents a framework for defining and identifying the channels through which student-teacher race match may affect long-run educational outcomes. Specifically, section 4.1 discusses possible mechanisms, beginning with ideas from a literature on *culturally relevant pedagogy* (Ladson-Billings, 1995). Motivated by these ideas, section 4.2 presents a formal model of the education production function, which distinguishes between teacher effectiveness and role-model effects, which are modeled as information provision. Section 4.3 provides tests of the model's implications and provides evidence that role model effects help to explain our main empirical results. Finally, section 4.4 conducts some exploratory analyses of same-race primary school teachers' effects on middle-school outcomes.

[18]We use 80% as this approximately marks the bottom tercile of the distribution, which is the measure used in Dynarski et al. (2013), though the null finding is robust to using other definitions.

[19]Nor do we find any evidence that race-match effects vary by observed teacher qualifications.

4.1 Early Literature on Black Teachers

Why does assignment to a black teacher have a marked impact on black students' long-run outcomes? One straightforward explanation is that in schools serving black students, black teachers are simply better than their white counterparts. This might occur because white teachers in schools with high rates of minority children tend to have less experience than their black colleagues due to teacher sorting patterns (Hanushek et al., 2004; Jackson, 2009) and experience predicts teacher effectiveness (Wiswall, 2013); indeed, we observe this pattern in the STAR data in Table 1, where black teachers have three more years of experience, on average, than their white counterparts.

An alternative set of hypotheses is rooted in the idea that black teachers are systematically more skilled than their white peers *at instructing black students*. This idea has received much attention outside economics, as scholars of education, sociology, and critical race theory have proposed that black teachers benefit black students by employing *culturally relevant pedagogies* (Ladson-Billings, 1995) and teaching *hidden curricula* (Foster, 1990). This literature began with ethnographic research on the role of black teachers in the segregated South, with particular attention paid to how the Brown vs. Board of Education ruling may have had a perversely negative impact on some students as it led to an exodus of black women from the teaching profession once all-black schools were legislated out of existence.[20]

Practices that constitute *culturally relevant pedagogy* can range from correctly reading student behavior and relating with appropriate cultural references to understanding how black students may perceive authority differently from non-black students. Walker (2001) emphasizes that black teachers embraced a set of ideas around teaching black students that were rooted in existing relationships with the larger black community, an idea that is echoed in Kelly (2010)'s account of black teachers visiting their students' parents at home. Foster (1990, 1997) explicitly introduced the concept of teaching a non-academic *hidden curricula*, which includes self-esteem and pride in your racial identity; cultural solidarity, affiliation, and connectedness with the larger black community; and the unique (to black students) political and social reasons for educational attainment.

Many of these ideas align with, or even motivate, the *identity economics* concepts for improving schools put forth by Akerlof and Kranton (2002). They are also adjacent to other teaching strategies and behaviors rooted in economics and psychology, including the concept of implicit bias, which might lead teachers of all backgrounds, but particularly

[20]Indeed, Kelly (2010) interviewed 44 former black teachers in North Carolina and makes the point that while segregated schools were severely under-resourced in terms of supplies and physical capital, teachers in these schools were dedicated, capable, and supported by the community.

white teachers, to *unconsciously* interact with black students in ways that harm achievement (Dee and Gershenson, 2017). For example, Tyson (2003) notes that even well-meaning white teachers might casually say and do things that harm black students' performance, such as mentioning that standardized tests are biased against black students. The idea of implicit bias is closely related to racial gaps in teachers' perceptions of black students' performance and behavior in class and their expectations for future educational success (Dee, 2005; Ferguson, 2003; Gershenson et al., 2016; Tyson, 2003). Indeed, Papageorge et al. (2018) show that biased teacher expectations do affect students' eventual educational attainment by creating self-fulfilling prophecies. Specifically, students benefit from teachers' optimism, and white teachers are systematically more optimistic about white students than black students.

There is, however, an alternative channel through which same-race teachers may matter: by serving as *role models*. Irvine (1989) details the nature in which black teachers embrace culturally-relevant pedagogical approaches that are well suited to the needs of black students. She argues that black teachers are *both* role models *and* "cultural intercessors and translators" for black students — which is directly related to increased student achievement. Similarly, Kelly (2010) finds that in addition to teaching items on the hidden curricula and deploying culturally-relevant pedagogy, black teachers were viewed as role models who represented the black middle class.[21] This distinction motivates how we formalize role model effects as separate from teacher effectiveness.

4.2 A Model of Educational Attainment and Role-Model Effects

In light of decades of theoretical and empirical research on the strategies used by black educators to effectively educate black pupils, in particular the distinction made by Irvine (1989), we develop a model of educational attainment that distinguishes between teacher effectiveness and role-model effects as two distinct channels through which black teachers improve black students' educational outcomes. Here, role models lead students to update inaccurate beliefs about the returns to human capital investments. The effectiveness story suggests that two black teachers would have a larger impact than one. In contrast, role model effects suggest that a second teacher may have a smaller effect if crucial information regarding returns to human capital has been transmitted by the first black teacher. The model formalizes this insight, develops conditions under which a second teacher has a smaller marginal effect than the first, and thus generates testable implications using dosage models.

Modeling role model effects as shifting beliefs about a parameter in a production function

[21] Related, see Dalton et al. (2015), who relate poverty to low aspirations.

is motivated by several strains of literature.[22] First, the concept of stereotype threat suggests that black students, already stereotyped as outsiders in educational spaces, are prone to "dis-identify" and disengage with educational environments (e.g., decrease effort) when they are reminded of their out-group membership (Steele, 1997). Exposure to black teachers may ameliorate students' sense of being out-group members in educational settings and reduce stereotype threat. Second, exposure to black teachers may provide students with a model of flourishing, not only in the face of challenges that apply to all students regardless of race (such as the challenge of mastering difficult abstract concepts), but also in the face of race-specific challenges that could further discourage educational attainment. These race-specific challenges may include overt racial prejudice along with more subtle micro-aggressions (commonplace intentional or unintentional racial slights that communicate prejudice or hostility (Sue et al., 2007, 2008)). Seeing role models who have flourished despite presumably facing similar race-specific challenges may make students more likely to persist in educational settings when they face those challenges themselves. Third, our conceptualization of role model effects is consistent with the notion of *growth mindsets*, which suggests that when students are taught that intelligence is a quality be developed, rather than a fixed trait, they put forth more effort and have higher achievement (Yeager and Dweck, 2012). As in our case, this points to a change in how students perceive the returns to effort. Fourth, literature on the power of teacher expectations is also relevant to our model insofar as it shows direct evidence that teachers affect educational attainment in part by shifting students' own expectations (Papageorge et al., 2018).[23] Finally, by modeling inaccurate beliefs about production function parameters, our approach to formalizing role model effects relates to Cunha et al. (2013), who show that mothers have inaccurate views of the returns to maternal time investments and thus under-invest.[24] In our case, students might make sub-optimal investments in their human capital based on incorrect assumptions about the returns to effort. Same-race teachers could lead them to update their beliefs about these returns, which leads in turn to more efficient investments and better educational outcomes.

The Production Function and Optimal Investments:

Assume two inputs, student effort E_{ig} with returns β_E and teacher effectiveness T_{ig} with

[22]Indeed, role model effects have been conceptualized in earlier literature as information provision. See, for example, Chung (2000). For empirical evidence, see Beaman et al. (2012), who relate female educational attainment in India to affirmative action policies that increase the number of women in positions of political power. Another related literature considers peer group effects. See, e.g., Lyle (2007).

[23]This is also consistent with recent research in economics suggesting that individuals make optimal choices, but might do so under inaccurate subjective beliefs (Cunha et al., 2013; Wiswall and Zafar, 2014).

[24]Indeed, experimental evidence supports a positive causal link between students' study efforts and their perceived productivity of effort (Ersoy, 2017).

returns β_T, which determine achievement (A_{ig}) for student i in grade g as follows:

$$A_{ig} = \beta_E E_{ig} + \beta_T T_{ig}. \tag{2}$$

The cost of effort is:

$$C(E_{ig}) = E_{ig}^2/2. \tag{3}$$

Maximizing A_{ig} subject to the effort constraint yields an optimal effort of $E_{ig}^* = \beta_E$. For a given teacher quality T, achievement over 2 periods is:

$$(A_{i1}^* + A_{i2}^*) \equiv \bar{A}_i^* = 2\beta_E^2 + 2\beta_T T. \tag{4}$$

Teacher Effectiveness:

Now suppose black teachers are more effective at teaching black students (e.g., due to mastery of a culturally relevant pedagogy). In this case:

$$A_{ig} = \beta_E E_{ig} + \beta_T^{SR} T_{ig}.$$

Now returns to a same-race (SR) teacher are $\beta_T^{SR} > \beta_T$. We then vary the number of black teachers so that $B \in \{0, 1, 2\}$. For a given teacher quality T, achievement over 2 periods is:

$$\begin{array}{rcl} \bar{A}_i^*(B=0) & = & 2\beta_E^2 + 2\beta_T T \\ < \bar{A}_i^*(B=1) & = & 2\beta_E^2 + (\beta_T + \beta_T^{SR})T \\ < \bar{A}_i^*(B=2) & = & 2\beta_E^2 + 2\beta_T^{SR} T \end{array} \tag{5}$$

We should expect to see gains with each black teacher. In particular, gains for having one black teacher are $(\beta_{SR} - \beta_T) \times T$ and gains for having two black teachers are $2 \times (\beta_{SR} - \beta_T) \times T$.

Role Model Effects:

A second way black teachers can raise student achievement is by acting as signals or counterexamples that lead students to update their beliefs about returns to effort. Formally, we modify the production function slightly so that:

$$A_{ig} = \tilde{\beta}_E E_{ig} + \beta_T T_{ig}. \tag{6}$$

Here, $\tilde{\beta}_E$ are beliefs about true returns to effort. Suppose students under-estimate returns to effort. Role model effects amount to information transmission whereby black teachers "signal" higher returns to effort. Student beliefs $\tilde{\beta}_E$ are that $\beta_E \in \{0, \bar{\beta}_E\}$, while $\bar{\beta}_E$ is the true value. Suppose the student's prior is $\lambda_0 \in \{0, 1\}$, which is the belief that $\beta_E = \bar{\beta}_E$. Having a black teacher B "signals" higher returns $\beta_E = \bar{\beta}_E$. In particular, $Pr[B|\beta_E = \bar{\beta}_E] = p_1$ and $Pr[B|\beta_E = 0] = p_0$. We assume that $p_1 > p_0$. In other words, black teachers provide evidence that effort pays off for black students.

In this setup, optimal effort for any beliefs λ is $E^*_{ig} = \bar{\beta}_E \lambda$. A lower prior leads to a sub-optimal decision (Cunha et al., 2013). In our case, this means lower optimal effort, which in turn leads to lower achievement. However, black students update their beliefs when matched to a black teacher. If $B = 1$, posterior beliefs are:

$$P[\beta_E = \bar{\beta}_E|B] \equiv \lambda_1 = \frac{p_1\lambda_0}{p_1\lambda_0 + p_0(1-\lambda_0)} > \lambda_0. \tag{7}$$

If $B = 2$, posterior beliefs are:

$$P[\beta_E = \bar{\beta}_E|B, B] \equiv \lambda_2 = \frac{p_1^2\lambda_0}{p_1^2\lambda_0 + p_0^2(1-\lambda_0)} > \lambda_1. \tag{8}$$

What remains to be shown are conditions under which the second update is smaller than the first, in which case there are different empirical implications than in a model in which race-specific effectiveness is key. Given optimal effort, marginal returns are $(\lambda_1 - \lambda_0)\bar{\beta}_E^2$ for the first black teacher and $(\lambda_2 - \lambda_1)\bar{\beta}_E^2$ for the second black teacher. Diminishing returns occur if $(\lambda_1 - \lambda_0) > (\lambda_2 - \lambda_1)$. This occurs when

$$\lambda_0 > \frac{p_0}{p_0 + p_1} \iff \frac{p_1}{p_0} > \frac{1-\lambda_0}{\lambda_0}. \tag{9}$$

The intuition is that diminishing returns occur when black teachers provide a "strong" signal that effort pays off, i.e., when p_1/p_0 is large. Alternatively, students' priors cannot be too small; otherwise students think a black teacher is a "fluke" and update slowly at first. A flat prior $\lambda_0 = 1/2$ means students update quickly at first.

Testable Implications:

The model generates the following testable implications. If race-specific teaching effectiveness drives the race match effect, we should not observe diminishing returns. If role model effects are the key mechanism, then there are conditions under which we would expect to observe diminishing returns. Diminishing returns occur if the first signal is more informative. This happens if prior beliefs are either flat or not "too wrong," or if black teachers are a strong signal of returns to effort. Notice, this means that diminishing returns are a sufficient, but not necessary, condition for the role model effect to be the underlying mechanism. According to the model, diminishing returns imply a role model effect.[25] If we do not observe diminishing returns, it is possible that effectiveness drives mismatch *or* that the role model effect is key, but that the inequality expressed above does not hold, i.e., prior beliefs are low or black teachers do not provide a strong enough signal. This distinction is important for the interpretation of results discussed below, where we show diminishing returns on some dimensions of educational attainment, but not on others.

4.3 Testing Model Predictions about Mechanisms

This section provides additional empirical analyses aimed at pinpointing the mechanisms underlying race-match effects. Some analyses, especially those using "dosage" models, are motivated by the model presented in the previous section.

4.3.1 White Students and Teacher Quality

One possible explanation of our results is that the black teachers in the schools and classrooms serving black students are simply better teachers than their white counterparts. Indeed, the summary statistics presented in Table 1 show that black teachers had about 3 more years of experience than white teachers, on average. The main results reported in Table 3 show that the race-match effect is robust to controlling for these observed teaching qualifications–and thus that the race-match effect is not driven by observable differences by teacher race in teacher quality. This is consistent with Krueger (1999) finding modest, at best, returns to teaching experience in the STAR sample. However, this does not mean that there are no differences by teacher race in *unobservable* teacher quality, as the bulk of the variation in teacher quality is not explained by observed qualifications (Staiger and Rockoff, 2010). We test this hypothesis below by leveraging data on white students, specifically white students in integrated classrooms, who have a black teacher.

Column 1 of Table 8 estimates the baseline model (equation 1) on the subsample of white STAR students and finds no effect of being randomly assigned a black teacher in the first year in a STAR school on the likelihood of enrolling in college.[26] Column 2 further restricts

[25]It is worth emphasizing that a different model would generate less sharp conclusions. For example, we could assume diminishing returns to teacher effectiveness, in which case diminishing returns in the data would no longer imply role model effects. The model we posit is designed to emphasize the distinction between repeated gains due to effectiveness versus the one-time provision of information that, once learned, need not be learned again, which we argue is akin to a role model effect. Accordingly, in a model of diminishing returns to teacher effectiveness, we could theoretically distinguish role model effects as leading to more drastically diminishing returns, i.e., develop conditions under which the impact of the second black teacher approaches zero, which is consistent with evidence we show below using dosage models.

[26]The estimates in Table 8 are robust to conditioning on the racial composition of classrooms.

the sample to include only white students in integrated classrooms (i.e., that contained both white and black students) and again finds a small, statistically insignificant coefficient on the black-teacher indicator.[27] This suggests that black teachers in integrated classrooms are not *universally* more effective than their white counterparts. Moreover, white students are not harmed by having a black teacher, which is unsurprising given that they are essentially guaranteed to have multiple white teachers in subsequent grades. Of course, the absence of an effect on educational outcomes does not rule out the possibility that exposure to a black teacher affects white students' racial and/or social attitudes (Carrell et al., 2015). This is an interesting and important question for future research to address.

Next, we extend the analysis of white students by estimating the baseline model on the pooled sample of white and black students. Here, we augment the model to include the interaction between student and teacher race. Columns 3 and 4 of Table 8 report these results. Consistent with the results in columns 1 and 2, the teacher-race indicators in the pooled models are insignificant, again indicating that white students are not affected by random assignment to a black teacher. The black-student coefficient is positive and statistically significant, indicating that in racially diverse STAR classrooms, black students are more likely to enroll in college than their white counterparts, regardless of who their teacher is. This is consistent with evidence that black families in mixed neighborhoods and mixed schools are penalized by being forced into neighborhoods in which they are significantly better-off socioeconomically than their white neighbors (Alba et al., 2000). Finally, the interaction terms are about 0.04, equivalent to the baseline estimates reported in column 4 of Table 3, though imprecisely estimated. These patterns are robust to replacing the school-by-cohort FE with classroom FE in columns 5 and 6 of Table 8, again suggesting that observed race-match effects for black students are not driven by differences in the overall quality of teachers or classrooms to which black students are assigned. However, it could still be that black teachers are better *at teaching black students*, which we investigate below.

4.3.2 Dosage

A key implication of the model is that if black teachers are more effective with black students, we would expect to see dosage effects of repeated exposure to black teachers. Alternatively, the role-model channel, in which the mere presence of a black teacher provides students with information regarding their potential outcomes, suggests diminishing returns to black teachers. To distinguish between these two stories, we estimate dosage models using an

[27] The similarity between columns 1 and 2 is perhaps unsurprising, as black teachers in exclusively white classrooms are rare.

instrumental variables strategy to account for noncompliance in subsequent STAR years (Dee, 2004). We also estimate the effect of *ever* having a same-race teacher, as this treatment is more comparable to a parameter we can estimate in section 5 using data from North Carolina. Here too we must instrument for the treatment because compliance with the experimentally assigned classrooms was imperfect in students' second, third, and fourth years in STAR (Dee, 2004). We estimate these models for two outcomes: college entrance exam taking and college enrollment.[28]

Because we only observe the classroom assignments of students who remain in their STAR school, we necessarily restrict the analytic sample to students who are observed in multiple STAR years.[29] To maximize power, we include all students who were in STAR for multiple years, but only code dosage for the first two years. This allows us to compare students who were in STAR for two, three, or four years on a common metric. Column 1 of Table 9 uses the baseline model specified in equation (1) to report the effect of assignment to a same-race teacher on sample selection (i.e., whether the student persisted to a second year in his or her STAR school). The estimate of about 0.03 (5%) is marginally statistically significant and indicative of positive sample selection. Like in the high-school analysis, this estimate is interesting in its own right, as persisting in the same school is arguably a positive outcome for highly-mobile, socioeconomically disadvantaged students (Schwartz et al., 2017). The results that follow are based on this selected sample and should be interpreted accordingly.[30]

Given the concerns about nonrandom attrition raised above, column 2 of Table 9 replicates the baseline analysis (equation 1) on the selected sample. Specifically, panels A and B show that the estimated effect of having a same-race teacher in the first STAR year on the likelihood of taking a college entrance exam and of enrolling in college, respectively, is about the same as in the full sample. This is reassuring in that it suggests that the education production function is similar for both attriters and non-attriters.

Column 3 presents OLS estimates of the reduced form effect of the *expected* number of black teachers over years one and two on both long-run outcomes. Following Dee (2004), the expected dose is based on what would have happened if the student complied with the experimental assignment. The expected dose is not always an integer, because in some schools there were multiple type-j classrooms taught by teachers of different races. For example, consider a school with two small first-grade classrooms, one taught by a white teacher and one taught by a black teacher. A kindergarten STAR entrant assigned to a

[28]Missing data problems in the high school graduation variable preclude a dosage analysis for this outcome.

[29]Previous research has documented significant, likely nonrandom, attrition from STAR (Ding and Lehrer, 2010; Krueger, 1999).

[30]Inverse probability weighting in the spirit of Moffit et al. (1999) yields similar results, suggesting that the results are not driven by selection on observables.

small classroom, if she complied, would then have a 50% chance of having a black teacher in the first grade. If she was randomly assigned to a black teacher in kindergarten, then her expected dose would be 1.5 black teachers over the first two years. In panel A (college exam taking) we see diminishing returns to the expected number of black teachers. However, we do not see diminishing returns in panel B for college enrollment, which will be a consistent theme of this table: there are diminishing returns to the effect of black teachers on college exam taking, but not on college enrollment.

Column 4 moves to using an actual measure of dosage, an indicator for whether the student had at least one black teacher in the first two years of STAR, and instruments for this dummy using a nonlinear function of the expected dose.[31] This definition of exposure is useful for at least two reasons. First, it alleviates concerns about the baseline model providing a lower bound for the true effect by moving students who had a white teacher in year one and a black teacher in year two from the control group to the treatment group, and comparing these students as well as those who had a black teacher in year one to students who had no black teachers. Second, this is a more interpretable, policy-relevant, and general parameter that can be compared to analogous estimates in other settings, including those presented in section 5. As expected, these estimates are slightly larger than those in column 2, consistent with the first-year model providing a lower bound due to the "noncompliance" of control-group students who have a black teacher in year two. Specifically, these results are consistent with the main finding that exposure to even one black teacher in primary school significantly increases the chances that black students aspire to, and enroll in, college.

Column 5 estimates a parametric dosage model akin to the reduced form model shown in column 3. These results are nearly identical to those in column 3, confirming that the instruments (expected dose) are strong predictors of actual dose. Here, too, we see diminishing returns in the college exam outcome but not for the college enrollment outcome. The quadratic terms are strongly jointly significant in both models, despite being individually insignificant for the enrollment outcome. This suggests that we might be underpowered to detect dosage effects. Finally, column 6 uses a nonparametric specification of dosage. Here, in panel A, we can clearly reject dosage effects on college exam taking. However, we fail to reject dosage effects on college enrollment.

[31] Table 9 uses a non-parametric specification of expected dose, though using a cubic function instead yields nearly identical point estimates.

4.3.3 Using the Model to Interpret Dosage Results

The dosage results show that there are diminishing returns to the effect of black teachers on some outcomes but not others. The model discussed in section 4.2 helps to interpret these results. According to the model, diminishing returns are a sufficient, but not necessary, condition for role model effects to occur. Thus, the fact that there are diminishing returns to black teachers' impacts on college entrance-exam taking suggests that *role-model effects* could be a key mechanism underlying teacher race-match effects on this dimension of student educational attainment. Indeed, as role models, black teachers may raise aspirations and inspire students to put forth more effort, which can be enough to nudge students into taking college entrance exams. However, we do not observe diminishing returns on the college-enrollment dimension. One possibility is that role models do matter for enrollment as well, but that the conditions under which the first update of student priors is larger than the second do not hold, i.e., the prior belief is too low or the teacher is not a strong enough signal for students to overcome their initial beliefs. Alternatively, it could be that teacher behaviors rather than role model effects are the key driver of college enrollment. We cannot distinguish between the two possibilities given the model.

Our results thus open up the possibility that race match effects are attributable to teacher effectiveness on some dimensions (e.g., college enrollment) and to role model effects on others (e.g., college test taking or college intent). In other words, the outcomes that teachers can generate as role models may be different from those that they can promote by being a more effective teacher. If so, there are dimensions, such as enrollment, where the presence of multiple black teachers may be key. For example, a single role model could affect students' own aspirations and college intent, encouraging them to sit for a college exam. However, the lack of diminishing returns on the college enrollment dimension means that multiple black teachers, potentially through behaviors based around culturally-relevant pedagogies, may consistently, year after year, provide students with knowledge and support so that prospects for completing postsecondary education rise.[32] Alternatively, multiple black teachers could communicate better with parents and thus assist families over multiple years in providing the support needed to encourage enrollment or to prevent students from falling prey to summer melt, whereby high-achieving, low-income black students plan to, but fail to, enroll in college (Castleman and Page, 2014).

In summary, black teachers positively affect black students on a variety of dimensions and the marginal impact of a second black teacher appears to vary across these dimen-

[32]This could potentially be reflected in higher test scores with higher dosage, consistent with findings in Penney (2017).

sions. Coupled with theory, these differences in dosage suggest possibly distinct channels across dimensions in why black teachers improve students' outcomes. More generally, our results suggest different production processes for different educational outcomes, which is an understudied point in education research.

4.4 Medium-Term Outcomes

Another way to probe potential mechanisms is to investigate the impacts of having a same-race teacher on intermediate outcomes. A well-documented result in the literature on teacher effectiveness is that teachers' effects on students' test scores fade out after a few years, but reappear when looking at longer-run, non-test score outcomes (Chetty et al., 2014). Jackson (2018) identifies a likely reason for this, showing that teachers who improve students' non-cognitive skills in the short-run are more likely to improve students' long-run outcomes than teachers who only improve students' test scores. Dee (2004) documents significant contemporaneous effects of race match on black STAR students' test scores and Penney (2017) extends this work to show that these effects persist one or two grades later. However, the literature has yet to investigate longer-run effects of mismatch on student test scores. We address this gap in Table 10 by estimating the medium-run effects of assignment to a same-race teacher in grades K-3 on middle school test scores.

This is an exploratory analysis that stops short of making causal inferences for two reasons. First, intermediate data on middle school test scores (and student and teacher perceptions data used below) are missing for large swathes of the main analytic sample. This is particularly true for the perceptions data. Second, hypothesis testing is plagued by the multiple comparisons problem, as there are dozens of grade-specific test scores and over 40 perception questions per student. Accordingly, we make do with the available data and provide some suggestive results.

We average each student's end-of-year test scores in grades 6-8. We do this because many students are missing one or more grade's worth of test score data for these grades, and this approach is similar to the imputation procedures used in previous STAR research (Dee, 2004; Krueger, 1999). Even so, middle school test scores are missing for a nontrivial segment of the analytic sample, though panel C of Table 10 shows that assignment to a same-race teacher did not affect the likelihood of missingness. To make the test scores comparable across grades, we standardize the tests by grade and subject to have mean 0 and SD 1.

OLS and quantile regression estimates of the impact of race match on middle school test scores, otherwise equivalent to equation (1), are reported in Table 10. Panel A reports OLS estimates, which show positive but imprecisely estimated effects ranging from 5 to 7% of a test-score SD. The lack of statistical significance here is consistent with teachers' effects on test scores fading out after several years (Chetty et al., 2014). Panel B reports quantile regression estimates for each quintile of the test-score distribution.[33] Columns 1-3 show that in math, reading, and English Language Arts (ELA) the impact of race match on the bottom half of the test-score distribution is larger than the OLS average effect. Moreover, the point estimate on either the 20th or 40th quantile is statistically significant at the 95% confidence level for each subject. This shows that exposure to same-race teachers in grades K-3 likely had some impact on medium-run (middle school) performance, specifically in the bottom to middle of the achievement distribution.

Finally, column 5 provides suggestive evidence that having a black teacher improved study skills at the bottom of the distribution, though this estimate is imprecisely estimated ($p = 0.14$). We further investigate the impact on study skills below, using data on teachers' subjective assessments. In sum, the evidence presented in Table 10 suggests that black teachers may have pulled up the bottom of the black achievement distribution. This would be consistent with our findings that having a same-race teacher (i) increased high-school graduation rates and (ii) increased college enrollment, mainly in community colleges.

Project STAR conducted two separate subjective surveys: one of fourth-grade teachers' assessments of former STAR pupils and one of former STAR pupils in the eighth grade. While over 20 questions were asked in each wave, we grouped questions thematically and report select results in Table 11. Missing data is a particularly acute problem here, as there are fewer than 500 responses for fourth grade and only 920 in eighth grade, and responses are endogenous to class-type assignments (Dee and West, 2008).[34] Accordingly, we interpret these results in a purely exploratory sense given the serious concerns about both sample selection and multiple comparisons. Nonetheless, some interesting patterns emerge.

Columns 1 and 2 report the estimated effects of the student having had a same-race teacher in his or her first STAR year (grades K-3) on fourth-grade teachers' perceptions of the student. We group a select set of responses into three categories: student's effort, curiosity, and behavior. Here, we see significant effects on the first two categories, but not on behavior. Specifically, regarding effort, teachers were about ten percentage points more likely to report that students who previously had a same-race teacher always or almost always were "persistent," "made an effort," and "tried to finish difficult work." Together with evidence of contemporaneous race-match effects on student attendance (Holt and Gershenson, 2017;

[33]We use the Quantile-FE estimator proposed by Powell (2016).

[34]The reason that more responses are missing in fourth grade is that the STAR experiment purposely limited the survey to ≤ 10 students per classroom, to lessen the burden on teachers (Finn et al., 2007). See Finn et al. (2007) for a thorough description of the medium-run testing and survey instruments.

Tran and Gershenson, 2018), and the central result in Jackson (2018), this suggests that non-cognitive skills such as effort and engagement are one mechanism through which race-match effects on long-run outcomes operate. This is also consistent with our interpretation of role-model effects providing information that causes students to update priors about the returns to effort.

Similarly, those same teachers were marginally more likely to report that previously race-matched students always or almost always "asked questions" and "discussed subject matter outside of class." These questions suggest higher levels of intellectual curiosity among the students who were randomly assigned to a black teacher in their initial STAR year. Together with the null finding for behavior, this suggests non-cognitive skills such as work ethic and academic engagement are possible channels through which black teachers affect black students' long-run outcomes. Indeed, the null effect on behavior suggests that this is not a channel through which long-run effects operate, and is consistent with Tyson (2003)'s view that both white and black teachers focus on black students' behavior.[35]

Columns 3 and 4 report the results for students' eighth grade self reports. There are fewer significant results here, though one result stands out: students who had a black teacher in K-3 were significantly more likely to agree or strongly agree with the belief that "I can talk to teachers about problems." This is the closest question there is to assessing the advising, mentoring, or nurturing role that same-race teachers might play, but it is suggestive that the singular experience of having a same-race teacher early in elementary school can permanently shift the way students view and interact with teachers. Put simply, exposure to a black teacher can build students' trust in teachers more generally (Foster, 1990). We further pursue the likely channels through which the long-run effects of student-teacher race match might operate in section 5, using rich administrative data on the population of public school students in North Carolina.

5 Replication Exercise

We now replicate the main findings of the Project STAR analysis–that exposure to same-race teachers in primary school affects long-run educational outcomes such as high school graduation and college aspirations–using administrative data on the population of North Carolina public school students.[36] We do so by exploiting within-school transitory variation

[35]This suggests that contemporaneous effects of race match on behavior (Holt and Gershenson, 2017; Lindsay and Hart, 2017) are due to perceptual and reporting differences and not actual behavioral changes.

[36]An earlier version of this paper placed greater emphasis on these results (Gershenson et al., 2017). We now emphasize the Project STAR results, as the identification of causal effects is cleaner and more

in the racial composition of schools' teaching staffs, in the spirit of Bettinger and Long (2005, 2010). Section 5.1 describes the North Carolina data and the quasi-experimental identification strategy. Section 5.2 presents the main replication results.

5.1 Data and Methods

We replicate the STAR results using student-level longitudinal administrative data on all black public school students in North Carolina who entered third grade between the 2000-2001 (2001) and 2004-2005 (2005) school-years. Students' educational trajectories are recorded through their senior year of high school. These data are publicly available to qualified researchers via the North Carolina Education Research Data Center (NCERDC) and are commonly used in the economics of education literature (Figlio et al., 2016; Jackson, 2018; Rothstein, 2010; Wiswall, 2013). The NCERDC student-level records can be linked to teacher identifiers through testing records, contain information on student and teacher demographics, and include schooling outcomes such as high-school graduation, drop-out, and self-reported college intent upon high-school graduation. The use of testing records to link students to teachers means that our analysis is restricted to tested grades (grades 3-5).[37]

The NCERDC data complement and improve upon the STAR data in several ways. First, they follow multiple cohorts, so we can exploit within-school changes in the demographic composition of the teaching force over time. Second, they cover the entire state population of public school students, which provides the statistical power and variation in student background necessary to identify heterogeneous treatment effects. Third, they provide better coverage of high school graduation than does the STAR data. Finally, by coming from a different state and decade, the North Carolina data provide a useful check of the external validity of the STAR results.

The trade-off, of course, is that there was no explicit policy of random assignment of students to classrooms in North Carolina, so we must account for potential sorting into same-race classroom pairings (Rothstein, 2010). Because we are interested in one-off long-run outcomes such as high school graduation rather than repeated measures such as end-of-grade test scores, student fixed effect (FE) and value-added strategies are not identified. Instead, we use panel data methods that exploit transitory, within-school variation in the racial composition of schools' teaching staffs. This strategy is motivated by the work of Bettinger and Long (2005, 2010), who leverage within-unit variation in the racial and faculty-rank

straightforward due to the experimentally-induced random assignment of students and teachers to classrooms.

[37]More recent waves of these data include administrative class roster data that link students to teachers in all primary school grades. Unfortunately for the purposes of this exercise, those cohorts have not yet reached high school.

composition of university departments as instrumental variables (IV) for assignment to a demographically-matched or adjunct instructor.

However, we focus on the reduced form effect of the would-be instrument, the school's share of teachers who are black, rather than the IV estimate because the exclusion restriction is suspect in the primary school context: black teachers might serve as mentors, advocates, and role models for black students in the same grade even when they are not the student's classroom teacher. The intuitive identification argument, then, is that within-school transitory fluctuations in the racial composition of a school's faculty are conditionally random. Identifying variation comes from the fact that students who enter the third grade in a particular school in different years (i.e., different cohorts) have different propensities to be assigned to, and interact with, same-race teachers, because teachers frequently go on leave, retire, change schools, and even change grades within a school (Brummet et al., 2017; Ost and Schiman, 2015). Of course, schools that experience high levels of teacher turnover and teacher grade switching are likely different on other dimensions as well, so we condition on school fixed effects (FE) and in some cases school-specific linear time trends. Conditional on school FE and time trends, then, transitory changes in the demographic composition of schools' teaching staffs are deviations from schools' "steady state" demographic composition, which are arguably exogenous. The reason is that, net of baseline school quality and trends in school quality and student composition, grade-specific teacher entries and exits are likely driven by exogenous, idiosyncratic factors such as enrollment changes, parental leaves, and retirements. We provide a balance test of this assumption in Appendix Table A4 and find that with the exception of the share of black students, which we directly control for in equation (10), changes in observed school characteristics do not predict the share of black teachers in the school. This bodes well for the exogeneity of the potential instrument, and thus the validity of the reduced form estimates we focus on.

Specifically, we estimate linear models of the form

$$y_{ist} = \beta_1 X_i + \beta_2 W_{st} + \delta Share_{st} + \theta_s + \gamma_t + u_{ist}, \tag{10}$$

which can be augmented to include school-specific time trends ($t \times \theta_s$), for student i who enters school s in third-grade cohort t. The vectors X and W include observed student and time-varying school characteristics while θ and γ are school and third-grade cohort FE, respectively. *Share* is the independent variable of interest, which in its simplest form measures the black share of self-contained third- through fifth-grade classroom teachers the student would potentially encounter if they remain in school s through fifth grade and follow an "on schedule" progression from grade 3 to 5 in the course of three academic years (i.e., if they neither change schools, repeat grades, nor skip grades).[38] Coding *Share* in this way eliminates concerns about endogenous grade repetition and school transfers. The parameter of interest is δ, which captures the partial effect of changing a school's share of black teachers from 0 to 1. This is an out-of-sample prediction, of course, so we scale the point estimates by 0.1, to get a more useful estimate that corresponds to the effect of increasing the share of black teachers by ten percentage points.

Table 12 summarizes the analytic sample, which contains five cohorts of black students in North Carolina who entered third grade for the first time between 2001 and 2005. These means are reported both overall and separately by the intersection of sex and "persistently disadvantaged." Following Michelmore and Dynarski (2016), the latter is defined as being eligible for free or reduced price lunch in each of grades 3-8, as these are the years that FRL eligibility is observed for these cohorts of students. The disadvantaged sample is arguably more comparable to the STAR sample, which intentionally recruited schools serving disadvantaged communities.

Panel A of table 12 summarizes students' educational outcomes. The NCERDC data contain two "long run" measures associated with educational attainment, which serve as the dependent variables in estimates of equation (10). The first is an indicator for whether students are ever observed as dropping out of high school.[39] Roughly 13% of students are recorded as having dropped out of high school in the overall sample, though this masks a sizable seven percentage point gender gap seen in columns 2 and 3.[40] Columns 4-6 show that dropout rates are nominally higher in the persistently-poor subsample.

The second outcome is an indicator for whether the student self-reported plans to attend a four-year college or university after graduation. 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 postsecondary institution. This self-reported college intent is arguably comparable to the indicator for taking a college entrance exam observed in the STAR data, as both are recorded in high school and are

[38]Specification tests suggest the effect is approximately linear, as cubic terms are individually insignificant and plots of the predicted probabilities are approximately linear. See Appendix Figure A1.

[39]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. Students who leave the state's public school system to enroll in either a private school or in another state are thus incorrectly counted as having dropped out.

[40]These implied graduation rates are slightly higher, though similar in magnitude, to corresponding national averages of 72% for males and 84% for females (Murnane, 2013).

binary proxies for a student's postsecondary educational plans. Consistent with national trends in college enrollment and completion (Bailey and Dynarski, 2011), college intent is higher among females than males and lower among persistently disadvantaged students.

Panel B of table 12 summarizes students' exposure to black teachers. About 44% of black students have at least one black classroom teacher in grades 3-5. The modal number of black teachers in these grades is zero. The majority of students who do have a black teacher have exactly one (about 30%). Only about 14% of black students have multiple black teachers in grades 3-5. These figures are fairly similar across socio-demographic groups. These variables are endogenous, of course, so we instead focus on the next variable, share of the cohort's teachers who are black, as the key independent variable in equation (10). That said, there is a strong, mechanical first stage between the share of black teachers in a grade and likelihood of being assigned a black teacher, and assignment to a black teacher is a primary channel through which the share of black teachers might affect long-run outcomes. The average cohort's teacher pool was about 25% black, with an in-school standard deviation of about ten percentage points.[41] Again, this is fairly constant across socio-demographic groups.

Panel C of table 12 summarizes the students themselves. About 45% of students were persistently eligible for FRL and 85% were eligible at least once between grades 3-8. These numbers are constant across genders, which is expected given that boys and girls live in the same households and neighborhoods. However, gender differences in learning disabilities, proxied by the presence of an IEP (individualized education plan), are sizable. About 12% of students had a parent with a college degree, and again there is a stark difference in parents' education between the full and disadvantaged samples.

5.2 Replication Results

Table 13 presents estimates of equation (10), which identify the reduced-form effect of the racial composition of schools' teaching staffs on students' long-run educational outcomes. It is reduced form in the sense that there are several channels through which this effect could operate. The primary channel is that the greater the share of black teachers, the greater the likelihood that students are assigned to a black classroom teacher. However, the presence of other black teachers in the student's grade can affect outcomes as well, by acting as mentors and advocates for black students and by mentoring white teachers.[42]

[41] About 40% of schools have zero variation in this variable, which tend to be small rural schools with zero black teachers, and serve a small share of the black student population. The main results are robust to dropping these schools from the analytic sample.

[42] Jackson and Bruegmann (2009) document the importance of teacher peer effects generally but do not investigate the possible racial dimension. We leave to future work the question of whether white teachers

Panel A of Table 13 estimates equation (10) for the full sample of black students in North Carolina. Column 1 shows a negative, statistically significant effect on the probability of dropping out of high school. Columns 2 and 3 repeat this exercise separately by sex, and find that the dropout effect is entirely driven by the response of male students. Columns 4-6 show a modest, but statistically insignificant effect on students' self-reported college intent. However, recall that the STAR experiment targeted disadvantaged schools. To replicate the STAR findings, we now turn to the subset of persistently disadvantaged black students.

Panel B of Table 13 restricts the sample to students who were eligible for FRL in each of grades 3-8. In columns 1-3 we see a larger effect on high school dropout than in the full sample, and that once again the effect on dropout is entirely driven by male students. Columns 4-6 show significant effects on college intent among the disadvantaged sample that are approximately equal for both male and female students. To put these effect sizes in perspective, a ten percentage point ($\approx$ one within-school SD) increase in the share of black teachers reduces the male dropout rate and increases self-reported college intent by almost one percentage point (4.6% and 2.1%, respectively). Because the primary goal here is to replicate the STAR results, and because the effects seem to be concentrated among disadvantaged students, all subsequent analyses are restricted to the disadvantaged sample.

Panels C-E probe the robustness of these main results. First, panel C restricts the sample to schools that exhibited variation in the share of black teachers. This excludes about 6,000 students from the sample. The point estimates are robust, which is to be expected given that the baseline model conditions on school fixed effects. Second, panel D introduces linear school-specific time trends to the model. This is an important sensitivity check because it addresses the concern that unobserved school trends are jointly determining student outcomes and the racial make-up of the teaching force. Here, too, the point estimates are quite robust, suggesting that unobserved factors are not driving the results. Finally, panel E reports the FE-logit version of equation (10) that accounts for the binary nature of the outcomes. Once again, the main results are robust to this modeling choice.

Finally, panels F and G of Table 13 probe some possible mechanisms. Panel F estimates equation (10) for the sample of white students in North Carolina. As in Table 8, the idea is to rule out the possibility that the black teachers that black students are exposed to are simply better teachers than their white counterparts. Like in Tennessee, the white students in North Carolina are not affected by the share of black teachers in their school. In fact, the point estimates are fairly tight zeros.

Panel G estimates an augmented version of equation (10) that distinguishes the share of

learn to more effectively educate black students from black teachers in their grade or school.

black male teachers from the share of black female teachers. This was not possible in the STAR data because nearly all teachers were female. As in the baseline model, columns 1-3 of panel G show that the effect of black teachers on high-school dropout is entirely concentrated among male students. Specifically, column 2 shows that the effect on black males' dropout decisions of the share of black male teachers is about three percentage points larger than that of black female teachers, but the two point estimates are not significantly different from one another, are both individually significant, and bound the baseline estimate from panel B. This suggests that on the high school dropout margin, black teachers of either sex significantly benefit black boys. Columns 4-6 conduct the same exercise for college intent. In the pooled sample (column 4), we see approximately equal effects of the shares of black male and black female teachers that are in line with baseline estimate reported in panel B. However, unlike the high-school dropout results, columns 5 and 6 show stark differences by student sex in how students' college intent response to the shares of black-male and black-female teachers. The most striking result is that the effect of the share of black-male teachers is three times larger than that of the share of black-female teachers on male students' college intent, a difference that is marginally significant. Similarly, column 6 shows that the female students' college intent is entirely affected by the share of black-female teachers.

The gender differences observed in columns 4-6 of panel G reinforce a causal interpretation of the baseline estimates of equation 10. These differences, especially for male students, suggest at least some role for the role-model phenomenon presented in section 4. The reason is that the ability to teach a culturally-relevant pedagogy or hidden curriculum is not exclusively sex-specific, though there is likely to be a sex-match dimension to the role-model effect mechanism, as the signal provided by a same-race *and* same-sex teacher is likely stronger. Moreover, that the sex of black teachers matters for college intent but not for high school graduation in Table 13 is consistent with the lack of dosage effects on college entrance exam taking but not college enrollment in Table 9, as both discrepancies suggest that the production functions underlying different types of educational attainment are themselves different.

For the sake of comparison with the Project STAR results, we now use the share of black teachers to instrument for whether the student had at least one black teacher in grades 3-5 in the spirit of Bettinger and Long (2005). The first-stage estimates are reported in panel A of Table 14. As expected given the mechanical relationship, they are quite strong. The IV estimates are reported in panel B. These results are consistent with the reduced form results presented in Table 13, as exposure to at least one black teacher only affects high school dropout rates of male students and significantly increases the college intent of all students. While there are theoretical reasons to question whether the exclusion restriction strictly holds, it is possible that these estimates still provide good approximations to the true causal effect of interest. Indeed, these estimates are similar in magnitude to the baseline STAR results. Specifically, we focus on the college intent outcome, which is quite similar to the college-exam and college-enrollment outcomes in the STAR data. The IV estimate in column 4 is 0.07, slightly larger than the effects on SAT/ACT taking and college enrollment reported in Tables 5 and 3, respectively. Of course, those STAR estimates are the effect of having a black teacher in a specific year, and not having at least one black teacher over a three-year span. Accordingly, an arguably more accurate comparison comes from the IV estimates reported in column 4 of Table 9, which show that having had at least one black teacher in the first two years of STAR raised the probability of taking an entrance exam and enrolling in college by about 10 and 8 percentage points, respectively. These two sets of STAR estimates bound the IV estimates from North Carolina. Taken together, these North Carolina results corroborate the basic result: exposure to even one black teacher in primary school significantly increases the odds that socioeconomically disadvantaged black students aspire to, and enroll in, college.[43]

How credible are the IV estimates? Black et al. (2017) describe an intuitive test, which amounts to estimating the reduced form (equation 10) separately by treatment status, which in this case refers to whether or not the student was ever assigned a black teacher in grades 3-5. These estimates are reported in panels C and D. Intuitively, if the instrument is valid, the share of black teachers should not significantly affect the outcome among individuals who are not treated.[44] It does, which suggests that either the exclusion restriction fails, there is selection, or both. Because we have theoretical reasons to mistrust the exclusion restriction and we show balance on the "instrument" in Appendix Table A4, we view the results of the Black et al. (2017) test as evidence against the exclusion restriction, and thus against the consistency of the IV estimates reported in panel B. That said, this does not invalidate the reduced-form estimates presented in Table 13 and even so, the IV estimates might not be too far off the mark. Indeed, the similarity with the STAR estimates suggests as much.

[43] We also probe the robustness of these linear 2SLS estimates to using a nonlinear model that accommodates both a binary outcome and a binary endogenous variable. We do so by jointly estimating a probit-ordered probit mixed-process model (Roodman, 2011), where the ordinal outcome takes one of three values: dropout, high school, high school plus college intent. This system is analogous to the usual bivariate-probit model used in the case of a binary dependent and endogenous variable (Wooldridge, 2010). These estimates, including average partial effects comparable to those reported in panel B of Table 14, are reported in Appendix Table A5. The results are qualitatively similar, suggesting that the IV results are not driven by a linear functional form.

[44] In the context discussed by Black et al. (2017) where there is a binary IV and treatment effect, the IV should should have no effect on the treated group either. In our case, it is possible to detect a relationship between the IV and the outcome among the treated since treatment is not binary, i.e., there is variation in how much treatment students receive. A positive coefficient might capture students who had multiple black teachers, for example.

6 Conclusion

We provide causal evidence that black students randomly assigned to a black teacher are 5 percentage points (7%) more likely to graduate high school and 4 percentage points (13%) more likely to enroll in college than their peers who are not assigned to a black teacher. Our main analyses leverage the Tennessee STAR experiment, which randomly assigned students to classrooms and teachers. These results are robust. We generate similar results using administrative data from North Carolina, a later time period, and a different identification strategy. Specifically, we exploit transitory shifts in the racial composition of teachers by grade, school, and year to isolate exogenous variation in black students' exposure to same-race teachers. Thus, our key findings are not limited to a specific state, time and experimental setting, but replicate across contexts. The magnitudes are large and, in the case of North Carolina, vary across specifications. For example, depending on the baseline, exposure to a single black teacher can reduce black male students' high school dropout rates by a third.

We also provide suggestive evidence that teachers as role models constitute an important mechanism through which race-match effects operate on some dimensions, e.g., taking a college entrance exam. Here, the marginal return to a second black teacher is essentially zero. A simple model suggests that lack of dosage provides evidence consistent with role model effects. In particular, for students facing aspirations gaps due to limited information on what is possible, a black teacher provides an effective counterexample that causes them to update their priors and change their beliefs, effort, and trajectories. However, on other dimensions, such as college enrollment, we do not find diminishing returns, suggesting that multiple black teachers have compounding effects on black students' outcomes, perhaps through more effective delivery of content.

Findings from this research provide some reason to be optimistic as they provide a path to reducing stubbornly persistent racial attainment gaps. However, they also raise a number of questions, some of which could be addressed in future research, surrounding efforts to diversify the teaching workforce. For example, while our study provides strong support for the idea that diversifying the teaching work force could *ceteris paribus* have a strong and positive effect on historically disadvantaged students, a pipeline that could achieve massive growth in the number of black teachers is non-existent (Putman et al., 2016). Hiring practices that attempt to diversify while maintaining high teacher quality would thus necessitate, for example, re-allocating college educated blacks from other lucrative fields to teaching, a relatively low-paid occupation. Doing so might lead to unintended consequences, such as exacerbating existing racial wage gaps, at least in the short run.

To put this issue into perspective, consider the following back-of-the-envelope calculation.

Of the roughly 3.8 million K-12 teachers in the U.S., approximately 256,000, or 6.7%, are black (NCES, 2017). Comparing this fraction to the 15.4% of K-12 students who are black suggests that doubling the number of black teachers would begin to get us close to aligning the work force with the student body they are supposed to teach. Doing so would necessitate steering 256,000 additional black college graduates from other occupations into teaching. Using the 2018 March Current Population Survey (CPS), and focusing on females with a Bachelor's or Master's degree, the group that comprises most teachers, we find that median earnings for blacks who are not teachers is roughly $49,000 while median earnings for blacks who are teachers is $45,000 (Ruggles et al., 2018).[45] Supposing non-teachers who became teachers were previously earning the median non-teacher income and now earn the median teacher income, efforts to diversify the teaching workforce imply a $4,000 pay cut for 256,000 black workers, thus reducing total income for blacks by more than one billion dollars.[46].

How to address the fact that the burden of increasing diversity would likely be borne by people of color in the form of pay cuts is not clear. Explicitly paying black teachers more than white teachers is likely a nonstarter for both practical and legal reasons. A more feasible policy response may be to make better use of incentives and bonuses for teaching in "hard-to-staff" schools, which include both low-achieving and high-poverty schools, and are the sorts of schools in which both black teachers and black students are over-represented (Hanushek et al., 2004). Indeed, such incentive schemes have been shown to work in North Carolina, where a $1,800 bonus reduced the teacher turnover rates by 14% (Clotfelter et al., 2008).

In light of these difficulties, it is likely that, for the foreseeable future, most teachers of black students will be white females. An important question is thus whether the lessons we learn here help us to harness the power of black teachers given the teaching workforce we have. Evidence on role model effects and, broadly, the idea that a single black teacher can have a large impact on a student's educational pathways, is helpful. It suggests that scarce black teachers could be creatively allocated to face more black students. It also opens up the possibility that black teachers and other black professionals can serve as role models without teaching students for a full year, but could work through more limited exposure: for example, a recent experiment finds that one-off, one-hour visits from female scientists in high-school science classes increases the likelihood that female students apply to selective

[45] This gap is at the low end of other comparisons of teacher and observationally-similar non-teacher salaries and ignores the fact that such gaps are larger among individuals with STEM degrees (Goldhaber, 2010).

[46] $\$4,000 \times 256,000$ = $1,024,000,000, which is a conservative figure. A summary of these calculations, including more details about the data sources and sample restrictions, and a range of alternative estimates, is provided in Appendix B. Estimated pay cuts depend on how we construct alternative earnings, and range from $4,000 to $14,759. This is in line with the range of gaps observed in the Baccalaureate and Beyond Longitudinal Study, which allows for more detailed comparisons (Goldhaber, 2010)

science majors in college (Breda et al., 2018). Still, this solution is only partial given the lack of diminishing returns on key dimensions of student educational attainment, suggesting the impact of black teachers goes far beyond acting as role models.

Given shortages of black teachers, another possibility is to reduce implicit bias and otherwise train white teachers to embrace a culturally relevant pedagogy, growth mindsets, and a culture of high expectations for all students. In considering these alternatives, the aim is not to undermine the goal of diversity, but to recognize the demographic realities of the current and potential teaching workforce. Indeed, literature on culturally relevant pedagogy continues to grapple with identifying what makes black teachers unique in their approaches, and how this might be used to train non-black teachers. One particularly promising type of intervention here is "light touch" empathy workshops, which have been shown to change teachers' purviews, reduce behavioral infractions, and improve student achievement (Okonofua et al., 2016).

Finally, while our findings on high school completion and enrollment in postsecondary education are indeed successes, the two-year programs and community colleges where these enrollments are happening have relatively low returns. Moreover, many individuals who enroll in two-year degree programs do not complete them, which means students make a costly educational investment but do not see the returns. Whether black teachers lead to an increase in two-year or four-year degree completion remains an open question and is not something we can identify with the STAR data. Whether we expect them to, whether they do, and whether we are adequately supporting these students in postsecondary schooling are important questions that we leave to future work.

References

Abadie, Alberto, Susan Athey, Guido W. Imbens, and Jeffrey Wooldridge, "When should you adjust standard errors for clustering?," *National Bureau of Economic Research Working Paper No. 24003*, 2017.

Akerlof, George A. and Rachel E. Kranton, "Identity and schooling: Some lessons for the economics of education," *Journal of Economic Literature*, 2002, *40* (4), 1167–1201.

Alba, Richard D., John R. Logan, and Brian J. Stults, "How segregated are middle-class African Americans?," *Social Problems*, 2000, *47* (4), 543–558.

Bailey, Martha J. and Susan M. Dynarski, "Gains and gaps: Changing inequality in US college entry and completion," *National Bureau of Economic Research Working Paper No. 17633*, 2011.

Beaman, Lori, Esther Duflo, Rohini Pande, and Petia Topalova, "Female Leadership Raises Aspirations and Educational Attainment for Girls: A Policy Experiment in India," *Science*, 2012, pp. 1–8.

Bettinger, Eric P. and Bridget T. Long, "Do faculty serve as role models? The impact of instructor gender on female students," *The American Economic Review*, 2005, *95* (2), 152–157.

_ **and** _ , "Does cheaper mean better? The impact of using adjunct instructors on student outcomes," *The Review of Economics and Statistics*, 2010, *92* (3), 598–613.

Black, Dan A., Joonhwi Joo, Robert LaLonde, Jeffrey A. Smith, and Evan J. Taylor, "Simple tests for selection: Learning more from instrumental variables," *Center for Economic Studies and Ifo Institute Working Paper No. 6392*, 2017.

Breda, Thomas, Julien Grenet, Marion Monnet, and Clémentine Van Effenterre, "Can female role models reduce the gender gap in science? Evidence from classroom interventions in French high schools," Mimeo, Paris School of Economics 2018.

Brummet, Quentin, Seth Gershenson, and Michael S. Hayes, "Teachers grade-level reassignments: Evidence from Michigan," *Educational Policy*, 2017, *31* (2), 249–271.

Card, David, "The Causal Effect of Education on Earnings," *Handbook of Labor Economics*, 1999, *3*, 1801–1863.

Carrell, Scott E., Marianne E. Page, and James E. West, "Sex and science: How professor gender perpetuates the gender gap," *The Quarterly Journal of Economics*, 2010, *125* (3), 1101–1144.

Carrell, Scott E, Mark Hoekstra, and James E West, "The Impact of Intergroup Contact on Racial Attitudes and Revealed Preferences," NBER Working Paper 2015.

Carrell, Scott, Elira Kuka, and Mark Hoekstra, "The Long-run Effects of Disruptive Peers," Forthcoming, *American Economic Review* 2018.

Castleman, Benjamin L. and Lindsay C. Page, "A trickle or a torrent? Understanding the extent of summer melt among college-intending high school graduates," *Social Science Quarterly*, 2014, *95* (1), 202–220.

Chamberlain, Gary, "Analysis of covariance with qualitative data," *The Review of Economic Studies*, 1980, *47* (1), 225–238.

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff, "Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood," *American Economic Review*, 2014, *104* (9), 2633–2679.

_ , _ , Nathaniel Hilger, Emmanuel Saez, Diane W. Schanzenbach, and Danny Yagan, "How does your kindergarten classroom affect your earnings? Evidence from Project STAR," *The Quarterly Journal of Economics*, 2011, *126* (4), 1593–1660.

Chung, Kim-Sau, "Role Models and Arguments for Affirmative Action," *American Economic Review*, 2000, *90* (3), 640–648.

Clotfelter, Charles, Elizabeth Glennie, Helen Ladd, and Jacob Vigdor, "Would Higher Salaries Keep Teachers in high-poverty schools? Evidence from a Policy Intervention in North Carolina," *Journal of Public Economics*, 2008, *92* (5-6), 1352–1370.

Cunha, Flávio, Irma Elo, and Jennifer Culhane, "Eliciting maternal expectations about the technology of cognitive skill formation," *National Bureau of Economic Research Working Paper No. 19144*, 2013.

Dalton, Patricio S, Sayantan Ghosal, and Anandi Mani, "Poverty and Aspirations Failure," *The Economic Journal*, 2015, *126* (590), 165–188.

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

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

_ and Martin West, "The non-cognitive returns to class size," *NBER Working Paper No. 13994*, 2008.

_ and Seth Gershenson, *Unconscious Bias in the Classroom: Evidence and Opportunities*, Mountain View, CA: Google Inc., 2017.

Ding, Weili and Steven F. Lehrer, "Estimating treatment effects from contaminated multiperiod education experiments: The dynamic impacts of class size reductions," *The Review of Economics and Statistics*, 2010, *92* (1), 31–42.

Dynarski, Susan M., Joshua Hyman, and Diane W. Schanzenbach, "Experimental evidence on the effect of childhood investments on postsecondary attainment and degree completion," *Journal of Policy Analysis and Management*, 2013, *32* (4), 692–717.

_ , **Steven W. Hemelt, and Joshua M. Hyman**, "The Missing Manual: Using National Student Clearinghouse Data to Track Postsecondary Outcomes," *Educational Evaluation and Policy Analysis*, 2015, *37* (1S), 53S–79S.

Ersoy, Fulya, "Effects of Perceived Productivity on Study Effort: Evidence from a Field Experiment," Mimeo, Loyola Marymount University 2017.

Fairlie, Robert W., Florian Hoffmann, and Philip Oreopoulos, "A community college instructor like me: Race and ethnicity interactions in the classroom," *The American Economic Review*, 2014, *104* (8), 2567–2591.

Ferguson, Ronald F., "Teachers' perceptions and expectations and the Black-White test score gap," *Urban Education*, 2003, *38* (4), 460–507.

Figlio, David, Krzysztof Karbownik, and Kjell G. Salvanes, "Education research and administrative data," in "Handbook of the Economics of Education," Vol. 5, Amsterdam, NL: Elsevier, 2016, pp. 75–138.

Finn, J., J. Boyd-Zaharias, R. Fish, and S. Gerber, "Project STAR and beyond: Database user's guide. Lebanon, TN: HEROS, Inc.," 2007.

Foster, Michele, "The politics of race: Through the eyes of African-American teachers," *Journal of Education*, 1990, *172* (3), 123–141.

_ , *Black teachers on teaching*, New York, NY: New Press, 1997.

Gershenson, Seth, Cassandra Hart, Constance Lindsay, and Nicholas W. Papageorge, "The long-run impacts of same-race teachers," *IZA Discussion Paper No. 10630*, 2017.

_ , **Stephen B. Holt, and Nicholas W. Papageorge**, "Who believes in me? The effect of student–teacher demographic match on teacher expectations," *Economics of Education Review*, 2016, *52*, 209–224.

Goldhaber, Dan, "Teacher Pay Reforms: The Political Implications of Recent Research," CEDR Working Paper 2010.

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

Grossman, Michael, "Education and Nonmarket Outcomes," *Handbook of the Economics of Education*, 2006, *1*, 577–633.

Hanushek, Eric A., John F. Kain, and Steven G. Rivkin, "Why public schools lose teachers," *Journal of Human Resources*, 2004, *39* (2), 326–354.

Heckman, James J., "Sample selection bias as a specification error," *Econometrica*, 1979, *47* (1), 153–161.

_ , "Skill formation and the economics of investing in disadvantaged children," *Science*, 2006, *312* (5782), 1900–1902.

_ **and Dimitriy V. Masterov**, "The productivity argument for investing in young children," *Applied Economic Perspectives and Policy*, 2007, *29* (3), 446–493.

Holt, Stephen B. and Seth Gershenson, "The impact of demographic representation on absences and suspensions," *Policy Studies Journal*, 2017.

Irvine, Jacqueline J., "Beyond role models: An examination of cultural influences on the pedagogical perspectives of Black teachers," *Peabody Journal of Education*, 1989, *66* (4), 51–63.

Ishitani, Terry T. and Stephen L. DesJardins, "A longitudinal investigation of dropout from college in the United States," *Journal of College Student Retention: Research, Theory & Practice*, 2002, *4* (2), 173–201.

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

_ , "What do test scores miss? The importance of teacher effects on non–test score outcomes," *Journal of Political Economy*, 2018, *126* (5), 2072–2107.

_ **and Elias Bruegmann**, "Teaching students and teaching each other: The importance of peer learning for teachers," *American Economic Journal: Applied Economics*, 2009, *1* (4), 85–108.

Kelly, Hilton, *Race, remembering, and Jim Crows teachers*, Abingdon, UK: Routledge, 2010.

Krueger, Alan B., "Experimental estimates of education production functions," *The Quarterly Journal of Economics*, 1999, *114* (2), 497–532.

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

Ladson-Billings, Gloria, "Toward a theory of culturally relevant pedagogy," *American Educational Research Journal*, 1995, *32* (3), 465–491.

Lee, David S., "Training, wages, and sample selection: Estimating sharp bounds on treatment effects," *The Review of Economic Studies*, 2009, *76* (3), 1071–1102.

Lim, Jaegeum and Jonathan Meer, "Persistent Effects of Teacher-Student Gender Matches," Forthcoming *Journal of Human Resources* 2017.

Lindsay, Constance A. and Cassandra M.D. Hart, "Exposure to same-race teachers and student disciplinary outcomes for black students in North Carolina," *Educational Evaluation and Policy Analysis*, 2017, *39* (3), 485–510.

Lochner, Lance and Enrico Moretti, "The Effect of Education on Crime: Evidence from Prison Inmates, Arrests, and Self-Reports," *American Economic Review*, 2004, *94* (1), 155–189.

Lusher, Lester, Doug Campbell, and Scott Carrell, "TAs Like Me: Racial Interactions between Graduate Teaching Assistants and Undergraduates," *Journal of Public Economics*, 2018, *159*, 203–224.

Lyle, David S, "Estimating and Interpreting Peer and Role Model Effects from Randomly Assigned Social Groups at West Point," *The Review of Economics and Statistics*, 2007, *89* (2), 289–299.

Michelmore, Katherine and Susan Dynarski, "The gap within the gap: Using longitudinal data to understand income differences in student achievement," *National Bureau of Economic Research Working Paper No. 22474*, 2016.

Moffit, Robert, John Fitzgerald, and Peter Gottschalk, "Sample attrition in panel data: The role of selection on observables," *Annales d'Economie et de Statistique*, 1999, *55* (56), 129–152.

Moretti, Enrico, "Estimating the social return to higher education: Evidence from longitudinal and repeated cross-sectional data," *Journal of Econometrics*, 2004, *121* (1-2), 175–212.

_ , "Workers' education, spillovers, and productivity: Evidence from plant-level production functions," *American Economic Review*, 2004, *94* (3), 656–690.

Murnane, Richard J., "US high school graduation rates: Patterns and explanations," *Journal of Economic Literature*, 2013, *51* (2), 370–422.

NCES, "Digest of Education Statistics," National Center for Education Statistics, Washington, D.C.: U.S. Department of Education 2017.

Okonofua, Jason A., David Paunesku, and Gregory M. Walton, "Brief intervention to encourage empathic discipline cuts suspension rates in half among adolescents," *Proceedings of the National Academy of Sciences*, 2016, *113* (19), 5221–5226.

Ost, Ben and Jeffrey C. Schiman, "Grade-specific experience, grade reassignments, and teacher turnover," *Economics of Education Review*, 2015, *46*, 112–126.

Papageorge, Nicholas W., Seth Gershenson, and Kyungmin Kang, "Teacher Expectations Matter," *NBER Working Paper No. 25255*, 2018.

Penney, Jeffrey, "Racial interaction effects and student achievement," *Education Finance and Policy*, 2017, *12* (4), 447–467.

Powell, David, "Quantile regression with nonadditive fixed effects," *Quantile Treatment Effects*, 2016.

Putman, Hannah, Michael Hansen, Kate Walsh, and Diana Quintero, "High Hopes and Harsh Realities: The Real Challenges to Building a Diverse Workforce.," *Brookings Institution*, 2016.

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

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

Ruggles, Steve, Sarah Flood, Ronald Goeken, Josiah Grover, Erin Meyer, Jose Pacas, and Matthew Sobek, "IPUMS USA: Version 8.0," Minneapolis, MN: IPUMS, 2018. https://doi.org/10.18128/D010.V8.0 2018.

Schwartz, Amy E., Leanna Stiefel, and Sarah A. Cordes, "Moving matters: The causal effect of moving schools on student performance," *Education Finance and Policy*, 2017, *12* (4), 419–446.

Staiger, Douglas O. and Jonah E. Rockoff, "Searching for effective teachers with imperfect information," *Journal of Economic Perspectives*, 2010, *24* (3), 97–118.

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

Sue, Derald Wing, Jennifer Bucceri, Annie I Lin, Kevin L Nadal, and Gina C Torino, "Racial Microaggressions and the Asian American Experience.," *Cultural Diversity and Ethnic Minority Psychology*, 2007, *13* (1), 72.

_ **, Kevin L Nadal, Christina M Capodilupo, Annie I Lin, Gina C Torino, and David P Rivera**, "Racial Microaggressions against Black Americans: Implications for Counseling," *Journal of Counseling & Development*, 2008, *86* (3), 330–338.

Tran, Long and Seth Gershenson, "Experimental Estimates of the Student-Attendance Production Function," IZA Discussion Paper No. 11911 2018.

Tyson, Karolyn, "Notes from the back of the room: Problems and paradoxes in the schooling of young Black students," *Sociology of Education*, 2003, *76* (4), 326–343.

Walker, Vanessa S., "African American teaching in the South: 1940–1960," *American Educational Research Journal*, 2001, *38* (4), 751–779.

Wilson, William Julius, "The truly disadvantaged: The inner city, the underclass, and public policy," *Chicago: University of Chicago*, 1987.

Wiswall, Matthew, "The dynamics of teacher quality," *Journal of Public Economics*, 2013, *100*, 61–78.

_ **and Basit Zafar**, "Determinants of college major choice: Identification using an information experiment," *The Review of Economic Studies*, 2014, *82* (2), 791–824.

Wooldridge, Jeffrey M., *Econometric analysis of cross section and panel data*, Cambridge, MA: MIT Press, 2010.

Yeager, David Scott and Carol S Dweck, "Mindsets that Promote Resilience: When Students Believe that Personal Characteristics Can Be Developed," *Educational psychologist*, 2012, *47* (4), 302–314.

Figures and Tables

Table 1: Analytic Sample Means

	All (1)	Black T (2)	White T (3)	Cohort 1 (4)	Cohort 2 (5)	Cohort 3 (6)	Cohort 4 (7)
A. Student Characteristics							
Male	0.52	0.52	0.53	0.50	0.56	0.55	0.54
Free Lunch (FL)	0.81	0.80	0.82	0.80	0.83	0.81	0.80
FL Missing	0.02	0.02	0.02	0.00	0.02	0.04	0.05
Missing name or DOB	0.12	0.11	0.12	0.23	0.00	0.01	0.02
K Cohort	0.50	0.46	0.53	1.00	0.00	0.00	0.00
G1 Cohort	0.21	0.21	0.21	0.00	1.00	0.00	0.00
G2 Cohort	0.17	0.19	0.16	0.00	0.00	1.00	0.00
G3 Cohort	0.12	0.14	0.10	0.00	0.00	0.00	1.00
B. Classroom & School Characteristics							
Black Teacher (T)	0.44	1.00	0.00	0.40	0.43	0.47	0.53
Small Class	0.24	0.22	0.26	0.29	0.16	0.20	0.24
Regular Class	0.38	0.41	0.36	0.34	0.50	0.36	0.37
Regular w/ Aide	0.38	0.38	0.38	0.37	0.34	0.44	0.39
Class Size	21.4	21.6	21.3	20.8	21.6	22.2	22.3
Class %Black	0.81	0.90	0.75	0.83	0.75	0.84	0.82
T Masters+	0.31	0.23	0.37	0.26	0.32	0.36	0.42
T Experience	10.6	12.4	9.2	8.4	11.5	12.8	15.2
Share Black Ts	0.43	0.60	0.30	0.40	0.43	0.46	0.52
Poor School	0.67	0.69	0.66	0.70	0.67	0.61	0.65
C. Long-Run Outcomes							
HS Grad Observed	0.37	0.38	0.37	0.40	0.35	0.38	0.32
HS Grad	0.68	0.67	0.68	0.74	0.61	0.60	0.61
Took SAT/ACT	0.27	0.29	0.25	0.31	0.22	0.26	0.22
SAT Score	777.4	775.2	779.4	790.4	772.5	745.2	765.8
College Enrollment	0.33	0.34	0.31	0.36	0.29	0.29	0.30
Enrolled 2-Yr	0.18	0.18	0.17	0.19	0.16	0.16	0.17
Enrolled 4-Yr Public	0.11	0.12	0.11	0.13	0.10	0.10	0.09
Enrolled 4-Yr Private	0.04	0.04	0.04	0.04	0.04	0.03	0.04
Persisted	0.24	0.25	0.24	0.27	0.21	0.21	0.23
Earned Degree	0.09	0.10	0.09	0.11	0.08	0.07	0.07
N	4142	1806	2336	2057	876	717	492

Notes: Sample size refers to full analytic sample; means for high school (HS) graduation and SAT score only reported for those whose HS records are observed and who took the SAT/ACT, respectively. The reported SAT score is the converted ACT score for those who took the ACT. Share black teachers refers to the share of black teachers in the student's entry grade. Poor school is defined as more than 80% of students being eligible for free lunch.

Table 2: Balance Test

	All		Male		Female	
	(1)	(2)	(3)	(4)	(5)	(6)
Male	-0.007	0.001				
	(0.016)	(0.012)				
FRL Eligible	-0.022	-0.014	-0.040	-0.040	-0.005	0.016
	(0.038)	(0.021)	(0.043)	(0.027)	(0.045)	(0.031)
Small Class	-0.055	-0.022	-0.038	-0.018	-0.071	-0.026
	(0.058)	(0.058)	(0.058)	(0.059)	(0.062)	(0.067)
Missing Name or DOB	-0.034	-0.024	-0.033	-0.035	-0.036	-0.018
	(0.040)	(0.021)	(0.048)	(0.029)	(0.046)	(0.030)
Observations	4,131	4,131	2,167	2,167	1,964	1,964
R-squared	0.004	0.338	0.003	0.367	0.006	0.350
School-Cohort FE	No	Yes	No	Yes	No	Yes
N (Students)	4,131	4,131	2,167	2,167	1,964	1,964
N (Classrooms)	447	447	378	378	367	367
R^2	0.004	0.338	0.003	0.367	0.006	0.350
Joint p	0.54	0.74	0.42	0.43	0.49	0.93
$E(Y)$	0.43	0.43	0.43	0.43	0.44	0.44

Notes: The outcome is a binary indicator equal to one if the student was assigned a black teacher in his or her first STAR year, and zero otherwise. Standard errors are clustered by students' first-year classrooms. The reported p values are from joint significance tests of the model's non-FE (fixed effect) covariates. *** p<0.01, ** p<0.05, * p<0.1.

Table 3: Effect of Same-Race Teacher on College Enrollment

Cohort:	Kindergarten		All		All	
	(1)	(2)	(3)	(4)	(5)	(6)
Black Teacher (T)	0.047*	0.062**	0.037**	0.040**	0.047*	0.043*
	(0.028)	(0.026)	(0.017)	(0.017)	(0.028)	(0.026)
Black T × grade 1					-0.002	0.013
					(0.046)	(0.046)
Black T × grade 2					-0.029	-0.028
					(0.044)	(0.043)
Black T × grade 3					-0.028	-0.010
					(0.052)	(0.052)
Controls:	No	Yes	No	Yes	No	Yes
E(Y|Black T = 0)	0.339	0.339	0.314	0.314	0.314	0.314
N Students	2,052	2,052	4,131	4,107	4,131	4,107
N Classrooms	215	215	683	681	683	681
R^2	0.068	0.122	0.099	0.147	0.099	0.147

Notes: OLS estimates of the impact of being randomly assigned to a black teacher in students' first year of STAR on likelihood of enrolling in a college or university. The omitted first-year grade in the interaction models is kindergarten. All models condition on school-by-cohort fixed effects and the randomly assigned class type (small, regular, or regular w/ aide). Controls include student controls for sex and FRL status and teacher controls for a quadratic in experience, highest degree attained, and status on career ladder. Even-numbered column sample sizes are smaller because experience is unobserved for a handful of teachers; missing values of categorical control variables are dummied out to preserve sample size. Standard errors are clustered by students' first-year classrooms. The interaction terms in columns 5 and 6 are jointly insignificant, with p values > 0.7. *** p<0.01, ** p<0.05, * p<0.1.

Table 4: Sensitivity of Estimated Effects of Same-Race Teacher on College Enrollment

Model:	Baseline (w/ controls) (1)	Baseline (Name & DOB Obs.) (2)	FE Logit (3)	Baseline (Logit sample) (4)	Class Size (2SLS) (5)	Class %Black (6)
Black T	0.040**	0.035*	0.208**	0.040**	0.039**	0.040**
by classroom	(0.017)	(0.018)	(0.093)	(0.017)	(0.017)	(0.017)
no clustering	(0.019)					
by school-cohort	(0.020)					
by school	(0.014)					
E(Y|Black T = 0)	0.314	0.352	0.315	0.315	0.314	0.314
N Students	4,107	3,629	4,006	4,006	4,107	4,107
N Classrooms	681	668	595	595	681	681
R^2	0.147	0.174	.	0.127	0.146	0.147

Notes: Column 1 reproduces the baseline OLS estimate from column 4 of Table 3 to facilitate comparisons. Column 1 also reports reasonable alternatives to the preferred standard errors (s.e.) that are clustered by classroom. Column 2 estimates the baseline model for the subset of students whose name and date of birth are observed. Column 3 estimates the FE-Logit version of the baseline linear probability model (LPM), which necessarily excludes students in schools that had no variation in the outcome and precludes clustering the s.e. Column 4 estimates the baseline LPM on the FE-Logit's analytic sample. Column 5 replaces the class-type indicators with an exact count of class size. This model is estimated by 2SLS, where the randomly assigned class type indicators instrument for actual class size (Krueger, 1999). Column 6 adds a control for the share of the class that is black to the set of baseline controls. *** p<0.01, ** p<0.05, * p<0.1.

Table 5: Effect of Same-Race Teacher on Other Postsecondary Outcomes

Outcome:	Took ACT/SAT (1)	Enrolled in Comm. Coll. (2)	Enrolled in Public 4-Yr (3)	Enrolled in Private 4-Yr (4)	Persisted to year 2 (5)	Earned Degree (6)
Black T	0.043***	0.030**	0.002	0.008	0.027*	0.005
	(0.017)	(0.014)	(0.011)	(0.007)	(0.016)	(0.011)
E(Y|Black T = 0)	0.252	0.173	0.105	0.0361	0.236	0.0868
R^2	0.162	0.091	0.140	0.047	0.145	0.132

Notes: All models are akin to the baseline linear probability model estimated in column 4 of Table 3. The analytic sample contains 4,107 students in 681 classrooms. Standard errors are clustered by classroom. *** p<0.01, ** p<0.05, * p<0.1.

Table 6: Effect of Same-Race K-3 Teacher on High School Graduation

Model:	LPM	LPM	LPM	FE Logit	LPM	FE Logit
Sample:	Full	Selected	Selected	Selected	Imputed	Imputed
Outcome:	Selected	Enrolled	HS Grad	HS Grad	HS Grad	HS Grad
	(1)	(2)	(3)	(4)	(5)	(6)
Black T	0.042**	0.040	0.049*	0.268*	0.052*	0.271*
	(0.018)	(0.032)	(0.029)	(0.162)	(0.029)	(0.154)
E(Y|Black T = 0)	0.367	0.566	0.677	0.647	0.590	.
N Students	4,118	1535	1,535	1,418	4,118	.
N Classrooms	682	513	513	449	682	.
R^2	0.126	0.234	0.206	.	.	.

Notes: All models estimated in this table are versions of the baseline linear probability model estimated in column 4 of Table 3, on the likelihood of high school (HS) graduation. Standard errors are clustered by students' first-year classrooms (except in the case of FE Logits). In columns 5 and 6 HS graduation is imputed using a probit multiple imputation procedure (40 imputed data sets). The fixed-effects (FE) logit estimator drops school-cohorts with no variation in the outcome, which is why the sample is smaller in column 4 than in column 3. This is also why no sample size is reported in column 6, as the analytic sample varies from 4015 to 3909 in the 40 imputed data sets. *** p<0.01, ** p<0.05, * p<0.1.

Table 7: Heterogeneous Effects of Same-Race Teachers on College Enrollment

Interaction (X):	Male	Free Lunch	Class Size	Class %Black	Poor School	%Black Teachers
	(1)	(2)	(3)	(4)	(5)	(6)
Black T	0.041*	0.048	0.039**	0.039**	0.050	0.040**
	(0.025)	(0.039)	(0.017)	(0.018)	(0.035)	(0.018)
Black T$\times X$	-0.001	-0.012	-0.001	0.002	-0.015	-0.002
	(0.030)	(0.042)	(0.005)	(0.082)	(0.042)	(0.089)
R^2	0.147	0.147	0.146	0.146	0.146	0.146

Notes: All models estimated in this table are versions of the baseline linear probability model estimated in column 4 of Table 3, augmented to interact the black-teacher indicator with a specific student, classroom, or school characteristic. Standard errors are clustered by classroom. The analytic sample includes 4,107 students in 681 classrooms. In the analytic sample, E(Y|Black T = 0) = 0.314. *** p<0.01, ** p<0.05, * p<0.1.

Table 8: Effect of Black Teachers on White Students' College Enrollment

Students:	White Only		Black and White			
	(1)	(2)	(3)	(4)	(5)	(6)
Black Teacher (T)	0.002	0.011	-0.004	0.003		
	(0.034)	(0.035)	(0.031)	(0.031)		
Black Student (S)			0.067***	0.065***	0.071***	0.065***
			(0.019)	(0.020)	(0.022)	(0.022)
Black T×Black S			0.044	0.037	0.044	0.044
			(0.034)	(0.034)	(0.048)	(0.047)
Integration Restriction:	No	Yes	No	Yes	No	Yes
Classroom FE	No	No	No	No	Yes	Yes
E(Y|Black T = 0)	0.435	0.439	0.404	0.383	0.404	0.383
N Students	7,138	3,271	11,245	7,378	11,245	7,378
N Classrooms	972	392	1261	681	1261	681
R^2	0.155	0.163	0.147	0.143	0.220	0.197

Notes: All models estimated in this table are versions of the baseline linear probability model estimated in column 4 of Table 3, augmented to include student-race indicators and their interactions with teacher-race indicators in pooled-race samples. "Integration restriction" refers to restricting the sample to integrated classrooms that contained both white and black students. In columns 5 and 6, the school-by-cohort fixed effects (FE) and teacher-race indicators are subsumed by classroom FE. *** p<0.01, ** p<0.05, * p<0.1.

Table 9: Dosage Effects in Sample of Students who Persisted in STAR for ≥ 2 years

Model	Selection	Baseline	Reduced Form	At Least One	Parametric	Non-Parametric
Estimation:	OLS	OLS	OLS	2SLS	2SLS	2SLS
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. College Entrance Exam Taking						
Black T	0.029*	0.052**				
	(0.015)	(0.022)				
$E[Dose]$			0.168***			
			(0.063)			
$E[Dose]^2$			-0.065**			
			(0.032)			
$1[Dose \geq 1]$				0.098***		
				(0.035)		
$Dose$					0.197**	
					(0.079)	
$Dose^2$					-0.078**	
					(0.039)	
$1[Dose = 1]$						0.118***
						(0.043)
$1[Dose = 2]$						0.080*
						(0.043)
$E[Y\|Dose = 0]$	0.617	0.305	0.305	0.309	0.309	0.309
Adj.-R^2	0.255	0.120	0.122	0.120	0.119	0.119
Joint Sig. Test (p)			0.01		0.01	0.01
Equality Test (p)						0.42
Panel B. College Enrollment						
Black T	0.029*	0.036				
	(0.015)	(0.023)				
$E[Dose]$			0.031			
			(0.059)			
$E[Dose]^2$			0.010			
			(0.029)			
$1[Dose \geq 1]$				0.078**		
				(0.035)		
$Dose$					0.013	
					(0.075)	
$Dose^2$					0.023	
					(0.036)	
$1[Dose = 1]$						0.036
						(0.041)
$1[Dose = 2]$						0.117***
						(0.040)
$E[Y\|Dose = 0]$	0.617	0.364	0.364	0.365	0.365	0.365
Adj.-R^2	0.255	0.123	0.124	0.122	0.121	0.121
Joint Sig. Test (p)			0.04		0.01	0.01
Equality Test (p)						0.06

Notes: Column 1 uses the full baseline sample ($N = 4,107$). The samples in columns 2-6 are restricted to students who persisted in STAR schools for at least two years ($N = 2,445$). However, only the race of the teachers in the first two years are measured in the dosage indicators. The IV estimates use a nonparametric first stage in E(Dose), which is the expected value of the number of Black teachers a student would experience if he or she complied with random assignment in year 2 in STAR. *** p<0.01, ** p<0.05, * p<0.1.

Table 10: Effect of Same-Race K-3 Teachers on Middle School Exam Scores

	Math	Reading	ELA	Battery	Study Skills
	(1)	(2)	(3)	(4)	(5)
A. *Black* OLS	0.055	0.073	0.057	0.055	0.073
	(0.049)	(0.052)	(0.048)	(0.060)	(0.061)
N (students)	2,262	2,255	2,252	1,969	1,690
N (classrooms)	379	379	379	366	356
R-squared	0.224	0.216	0.255	0.252	0.199
B. Quantile Estimates					
$Q_{0.20}$	0.151**	0.095	0.093	0.095	0.154
	(0.071)	(0.071)	(0.062)	(0.072)	(0.104)
$Q_{0.40}$	0.102	0.101**	0.121**	0.102*	0.046
	(0.068)	(0.051)	(0.049)	(0.057)	(0.050)
$Q_{0.60}$	0.074	0.077	0.077	0.108**	0.081
	(0.047)	(0.050)	(0.052)	(0.050)	(0.061)
$Q_{0.80}$	0.041	0.075	0.063	0.072	0.051
	(0.060)	(0.047)	(0.071)	(0.071)	(0.055)
C. $Pr(Y = missing\|X)$					
Black LPM	-0.019	-0.020	-0.018	-0.014	0.004
	(0.020)	(0.019)	(0.020)	(0.021)	(0.019)

Notes: Panels A and B report OLS and Quantile Regression estimates, respectively, of impact of being randomly assigned to a black teacher in students' first year of STAR on middle school exam scores. Panel C reports the impact of being randomly assigned to a black teacher in students' first year of STAR on the likelihood that their middle school exam scores are missing. We standardize grade 6-8 exam scores to have mean zero and SD 1 by grade and subject. We then take each student's average score over all available grade 6-8 scores, though frequently only one grade's score is available per student. All models condition on school-by-cohort fixed effects, the randomly assigned class type (small, regular, or regular w/ aide), student controls for sex and FRL status, and teacher controls for sex, a quadratic in experience, highest degree attained, and status on career ladder. Standard errors are clustered by students' first-year classrooms. *** p<0.01, ** p<0.05, * p<0.1.

Table 11: Effect of Same-Race K-3 Teacher on Subjective Beliefs

4th Grade Teacher Survey Question (1)	Estimate (2)	8th Grade Student Survey Question (3)	Estimate (4)
Effort Related		Values / Information	
"is persistent"	0.110**	"what we learn in school will be useful in job"	-0.025
	(0.053)		(0.031)
"makes effort"	0.086*	"dropping out is a huge mistake"	0.025
	(0.050)		(0.026)
"tries to finish difficult work"	0.116**	"school is more important than people think"	0.010
	(0.058)		(0.021)
Curiosity		Connection w Teacher	
"asks questions"	0.112*	"I can talk to teachers about problems"	0.073**
	(0.064)		(0.028)
"discusses subject matter outside class"	0.087*	"people are interested in what I say"	-0.013
	(0.048)		(0.047)
Behavior		Connection w School	
"works well with others"	-0.001	"I feel proud being part of school"	-0.000
	(0.053)		(0.027)
"needs reprimanding"	-0.068	"school is important part of my life"	0.034
	(0.047)		(0.021)
"annoys others"	-0.027	"I participate in a lot of activities"	-0.046
	(0.041)		(0.035)
"raises hand to talk"	0.024	"school is my favorite place to be"	-0.026
	(0.056)		(0.044)
N (students)	468		920

Notes: OLS estimates of impact of being randomly assigned to a black teacher in students' first year of STAR on subjective survey responses. The binary outcome in column 1 is coded as a one if the teacher responded "Always" or "Almost always" to the statement (regarding student i). The binary outcome in column 3 is coded as a one if the student reported "Agree" or "Strongly agree" with the statement. All models condition on school-by-cohort fixed effects, student and teacher controls, and the randomly assigned class type (small, regular, or regular w/ aide). Controls include student controls for sex and FRL status and teacher controls for a quadratic in experience, highest degree attained, and status on career ladder. Standard errors are clustered by students' first-year classrooms. *** p<0.01, ** p<0.05, * p<0.1.

Table 12: North Carolina Summary Statistics

	All Black Students			Persistently Poor Black Students		
	All	Male	Female	All	Male	Female
	(1)	(2)	(3)	(4)	(5)	(6)
A. Outcomes (in%)						
H.S. Dropout	13.00	16.50	9.65	14.30	17.84	11.10
H.S. Grad, No 4-Year Intent	46.60	48.86	44.45	52.23	53.61	50.99
H.S. Grad, 4-Year Intent	40.22	34.35	45.79	33.25	28.21	37.77
B. Exposure to Black Teachers in Grades 3-5						
Exposure to $\geq$ 1 Black T	43.8%	43.7%	44.0%	45.6%	45.3%	45.8%
0 Black Teachers	56.2%	56.3%	56.0%	54.5%	54.8%	54.2%
1 Black Teacher	29.7%	29.3%	30.0%	30.2%	29.8%	30.5%
2 Black Teachers	11.2%	11.3%	11.0%	12.0%	12.0%	12.0%
3 Black Teachers	2.9%	3.0%	2.9%	3.3%	3.4%	3.3%
% Cohort's Teachers Black	25.5%	25.4%	25.5%	26.9%	26.8%	27.0%
	(24.94)	(24.99)	(24.88)	(25.93)	(25.98)	(25.88)
Within-School SD	[9.20]	[9.17]	[9.22]	[9.47]	[9.43]	[9.50]
C. Student Characteristics						
Persistently Low Income	45.44	44.22	46.60	100.00	100.00	100.00
Ever Low Income	85.79	85.63	85.95	100.00	100.00	100.00
Ever LEP	0.29	0.34	0.24	0.33	0.41	0.26
Ever Exceptional	10.75	14.47	7.19	11.78	15.88	8.08
Parent Ed: HS Dropout	10.99	11.18	10.81	14.53	14.69	14.39
Parent Ed: College Grad	12.20	12.33	12.09	4.02	4.10	3.96
Unique students	106,373	51,928	54,445	48,335	22,962	25,373

Notes: Standard deviations (SD) presented in parentheses. FRL=Free/Reduced-price Lunch. Teacher composition variables capture students and teachers in grades 3-5. Sample includes black students entering 3rd grade in NC Public Schools from 2001 to 2005. 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 13: North Carolina Reduced Form Estimates

Outcome:	High School Dropout			College Intent		
Sample:	All	Male	Female	All	Male	Female
	(1)	(2)	(3)	(4)	(5)	(6)
A. Full Sample						
$\hat{\delta}$	-0.022**	-0.049***	0.003	0.011	-0.006	0.019
(by school-cohort)	(0.011)	(0.016)	(0.014)	(0.016)	(0.021)	(0.023)
(by school)	(0.012)	(0.019)	(0.016)	(0.018)	(0.024)	(0.026)
N	105,155	51,321	53,834	103,693	50,458	53,235
B. Persistently Disadvantaged Sample						
$\hat{\delta}$	-0.036**	-0.082***	0.009	0.069***	0.062**	0.066**
(by school-cohort)	(0.015)	(0.023)	(0.019)	(0.023)	(0.029)	(0.033)
(by school)	(0.017)	(0.027)	(0.023)	(0.025)	(0.033)	(0.038)
N	47,900	22,747	25,153	47,164	22,323	24,841
C. Persistently Disadvantaged Sample, Drop "No-Variation Schools"						
$\hat{\delta}$	-0.036**	-0.081***	0.008	0.070***	0.067**	0.065**
	(0.015)	(0.023)	(0.019)	(0.022)	(0.029)	(0.033)
N	41,471	19,607	21,864	40,837	19,244	21,593
D. Persistently Disadvantaged Sample, w/ School-Specific Linear Time Trends						
$\hat{\delta}$	-0.035**	-0.053*	-0.012	0.083***	0.062*	0.082**
	(0.018)	(0.030)	(0.023)	(0.026)	(0.036)	(0.039)
N	47,900	22,747	25,153	47,164	22,323	24,841
E. Persistently Disadvantaged Sample, FE Logit Coefficient Estimates						
$\hat{\delta}$	-0.307**	-0.563***	0.128	0.327***	0.337**	0.287**
	(0.146)	(0.194)	(0.226)	(0.109)	(0.168)	(0.146)
N	46,614	21,611	22,498	46,890	21,828	24,577
F. White Student Sample						
$\hat{\delta}$	-0.008	-0.003	-0.012	0.006	0.006	0.003
	(0.011)	(0.017)	(0.015)	(0.017)	(0.023)	(0.024)
N	209,924	107,038	102,886	207,804	105,756	102,048
G. Persistently Disadvantaged Sample, by Teacher Gender						
$\hat{\delta}_{Male}$	-0.030	-0.111*	0.037	0.063	0.178**	-0.041
	(0.038)	(0.064)	(0.049)	(0.054)	(0.077)	(0.080)
$\hat{\delta}_{Female}$	-0.036**	-0.078***	0.005	0.069***	0.049	0.079**
	(0.015)	(0.024)	(0.020)	(0.023)	(0.030)	(0.035)
$H_0 : \delta_{Male} = \delta_{Female}$ (p value)	0.866	0.623	0.523	0.905	0.103	0.146

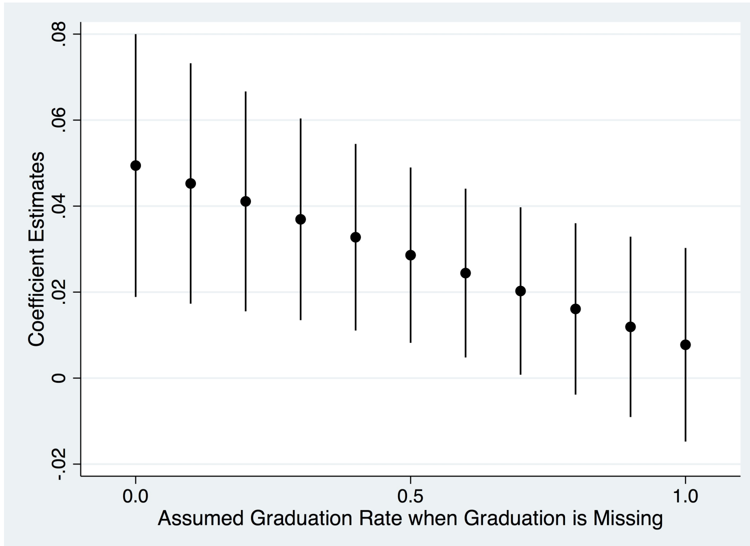
Notes: Standard errors reported in parentheses. Baseline standard errors clustered by the level of treatment variation: school-cohort. Persistently disadvantaged refers to students eligible for free or reduced lunch in each of grades 3-5. All models control for time-varying school characteristics and observed student socio-demographics. *** p<0.01, ** p<0.05, * p<0.1.

Table 14: North Carolina Instrumental Variables Estimates

Outcome:	High School Dropout			College Intent		
Sample:	All	Male	Female	All	Male	Female
	(1)	(2)	(3)	(4)	(5)	(6)
A. First Stage Effect of Share on having $1[\geq 1 Black\ Teacher]$						
	0.698***	0.671***	0.721***	0.699***	0.670***	0.725***
	(0.028)	(0.035)	(0.033)	(0.028)	(0.035)	(0.032)
B. IV (2SLS) Estimates for Persistently Disadvantaged Sample						
$1[\geq 1 BlackTeacher]$	-0.051**	-0.122***	0.012	0.098***	0.093**	0.091**
	(0.021)	(0.034)	(0.026)	(0.032)	(0.043)	(0.046)
N	47,900	22,747	25,153	47,164	22,323	24,841
C. Reduced Form Estimates for Treated Sample ($\geq$ 1 *Black Teachers*)						
	-0.049**	-0.103***	0.002	0.073**	0.070*	0.065
	(0.020)	(0.033)	(0.028)	(0.031)	(0.040)	(0.045)
N	21,811	10,297	11,514	21,488	10,112	11,376
D. Reduced Form Estimates for Non-Treated Sample (0 *Black Teachers*)						
	-0.035	-0.044	-0.016	0.091***	0.038	0.125**
	(0.025)	(0.043)	(0.032)	(0.035)	(0.049)	(0.053)
N	26,089	12,450	13,639	25,676	12,211	13,465

Notes: Standard errors reported in parentheses. Baseline standard errors clustered by the level of treatment variation: school-cohort. Persistently disadvantaged refers to students eligible for free or reduced lunch in each of grades 3-5. All models control for time-varying school characteristics and observed student socio-demographics. *** p<0.01, ** p<0.05, * p<0.1.

Figure 1: Effect of Same-Race K-3 Teacher on HS Graduation



Notes: OLS estimates of *Black* effect with 95% confidence intervals. Each estimate generated from sample that replaced missing dropout values with the "assumed graduation probability" listed on the horizontal axis. All models condition on school-by-cohort fixed effects and the randomly assigned class type (small, regular, or regular w/ aide). Controls include student controls for sex and FRL status and teacher controls for a quadratic in experience, highest degree attained, and status on career ladder. Standard errors are clustered by students' first-year classrooms.

Appendix A Additional Results

Table A1: Sensitivity Analysis for Selected Name & DOB Sample

Model:	Baseline	FE Logit	Baseline (FE Logit Sample)	Class Size (2SLS)	Class %Black
	(1)	(2)	(3)	(4)	(5)
Black T	0.035*	0.177*	0.035*	0.034*	0.035*
by classroom	(0.018)	(0.097)	(0.018)	(0.018)	(0.018)
no clustering	(0.020)				
by school-cohort	(0.021)				
by school	(0.013)				
E(Y|Black T = 0)	0.352	0.353	0.353	0.352	0.352
N Students	3,629	3,533	3,533	3,629	3,629
N Classrooms	668	587	587	668	668
R-squared	0.174		0.154	0.173	0.174

Notes: This table conducts the set of sensitivity analyses conducted in Table 4 for the subset of students whose name and date of birth (DOB) are observed. Column 1 reproduces the baseline OLS estimate from column 2 of Table 4 to facilitate comparisons, along with several reasonable estimates of the associated standard error. Column 2 estimates the FE-Logit version of the baseline linear probability model (LPM), which necessarily excludes students in schools that had no variation in the outcome and precludes clustering the s.e. Column 3 estimates the baseline LPM on the FE-Logit's analytic sample. Column 4 replaces the class-type indicators with an exact count of class size. This model is estimated by 2SLS, where the randomly assigned class type indicators instrument for actual class size (Krueger, 1999). Column 5 adds a control for the share of the class that is black to the set of baseline controls. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table A2: Effect of Same-Race K-3 Teacher on College Entrance Exam Performance

	Naive OLS (1)	Heckit No Exclusion (2)	Heckit Exclude FE (3)	Lee (2009) Bounds (4)
A. Full Sample				
ACT Score	-0.292	-0.194	-0.225	.
	(0.323)	(0.493)	(0.232)	
Selection	0.039**	0.152**	0.152**	.
	(0.019)	(0.062)	(0.062)	
$\hat{\lambda}$	.	0.954	-0.598	.
		(3.90)	(0.395)	
$\hat{\sigma}$	.	3.257	3.499	.
$\hat{\rho}$		0.293	-0.171	.
N (selected)	1001	1001	1001	.
N (not selected)	.	2740	2740	.
$E(ACT \mid Black = 0)$	16.6	16.6	16.6	
$E(Selection \mid Black = 0)$	0.25	0.25	0.25	.
B. Bounds Sample				
ACT Score	-0.347	-0.226	-0.197	[-0.34, 0.49]
	(0.312)	(0.631)	(0.293)	
Selection	0.039*	0.137**	0.137**	.
	(0.020)	(0.066)	(0.066)	
$\hat{\lambda}$		1.308	-1.247	.
		(5.99)	(0.643)	
$\hat{\sigma}$		3.442	3.591	
$\hat{\rho}$		0.380	-0.347	
N (selected)	719	719	719	719
N (not selected)	.	1823	1823	1823
$E(ACT \mid Black = 0)$	16.5	16.5	16.5	16.5
$E(Selection \mid Black = 0)$	0.26	0.26	0.26	0.26

Notes: The "ACT Score" and "Selection" coefficients refer to the estimated coefficient on the binary indicator for being randomly assigned to a black teacher in students' first year of STAR. The ACT score is the true ACT score for students who took the ACT college-entrance exam and the converted equivalent for students who took the SAT college-entrance exam. The selection coefficients reported in columns 2 and 3 are probit coefficients and therefore do not have a direct "partial effect" interpretation; however, using an approximate scale factor of 0.4, these probit coefficients map into partial effects of about 0.05, which are quite similar to their linear counterparts reported in Column 1. The models in columns 1-3 condition on school-by-cohort fixed effects, student and teacher controls, and the randomly assigned class type (small, regular, or regular w/ aide). Specific controls include student controls for sex and FRL status and teacher controls for a quadratic in experience, highest degree attained, and status on career ladder. In column 3, the school-by-cohort fixed effects are excluded from the structural model and are only included in the selection equation. In column 4, the Lee (2009) bounds are tightened by school-by-cohort fixed effects, but no other controls. The sample in panel B is restricted to school-specific cohorts that contained students who were and were not exposed to a black teacher in their first year, and who did and did not take a college entrance exam. Standard errors are clustered by students' first-year classrooms. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table A3: Sample Means by High School (HS) Completion Status

	All	HS Observed	HS Grad	HS Not Grad	HS Missing
	(1)	(2)	(3)	(4)	(5)
Male	0.525	0.449	0.395	0.562	0.570
Free Lunch	0.816	0.742	0.688	0.854	0.860
Missing Name or DOB	0.116	0.027	0.021	0.038	0.169
College Enrollment	0.326	0.568	0.714	0.264	0.181
Enrolled 2-Yr	0.176	0.290	0.330	0.204	0.107
Enrolled 4-Yr Public	0.112	0.212	0.295	0.040	0.053
Enrolled 4-Yr Private	0.038	0.066	0.088	0.020	0.021
Persisted	0.244	0.440	0.572	0.164	0.126
Earned Degree	0.092	0.182	0.252	0.036	0.039
N	4131	1544	1044	500	2587

Notes: HS Grad/Not Grad refers to a graduation record in the state of Tennessee. Students who graduated HS in other states could be counted in either column 4 or 5.

Table A4: North Carolina Balance Test

	Base	District-by-Year FE	Linear School Time Trends
	(1)	(2)	(3)
% Students FRL	-0.008	0.003	-0.004
	(0.014)	(0.026)	(0.033)
% Students Black	0.266***	0.262***	0.190*
	(0.066)	(0.068)	(0.102)
% Students Hispanic	0.019	0.004	-0.087
	(0.103)	(0.108)	(0.156)
School Average EOG	-2.991*	-3.080	0.058
	(1.777)	(1.979)	(2.538)
Pupil-Teacher Ratio	-0.080	-0.161	-0.131
	(0.092)	(0.111)	(0.134)
Log Enrollment	-1.027	-1.483	-1.389
	(2.234)	(2.436)	(2.765)

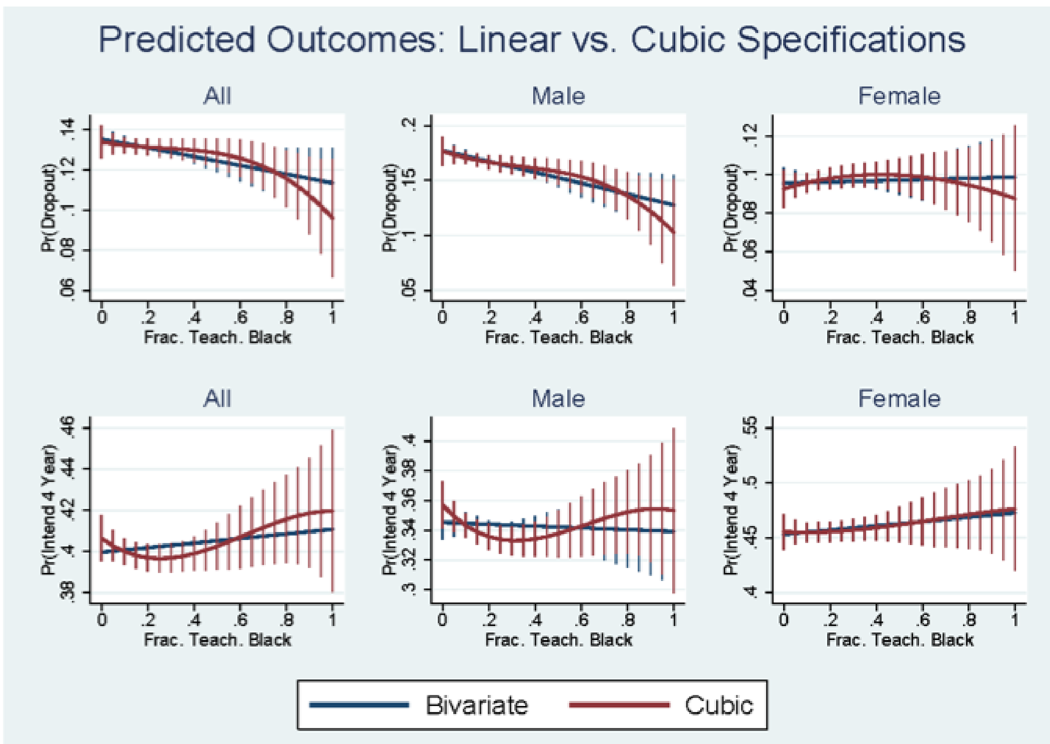
Notes: School-level panel regressions condition on school fixed effects (FE) and cluster standard errors by school. Dependent variable is the fraction of teachers for a school-cohort who are black, multiplied by 100 to be comparable in scale to school characteristics.

Table A5: Mixed Process Bi-Probit Model Estimates

	Coefficient (1)	APE (Dropout) (2)	APE (Intent) (3)
A. Probit First Stage			
Share	2.273***		
	(0.090)		
B. Ordered-Probit			
$1[\geq 1 BlackT]$	0.175***	-0.037***	0.059***
	(0.050)	(0.011)	(0.017)

Notes: $N = 47,164$ persistently FRL students. A first-stage probit and second-stage ordered probit are jointly estimated as a mixed process, as in Roodman (2011). The ordinal outcome takes one of three values: high school (HS) drop out, HS graduate, or HS graduate with college intent. The model is otherwise identical to the linear models estimated by 2SLS described in Table 14. The models control for school fixed effects, which are manually dummied out, and thus might introduce incidental parameters bias. However, this bias is likely minimal, as there tend to be many students per school (Greene, 2004).

Figure A1: Effect of Same-Race K-3 Teacher on HS Graduation



Notes: Fitted values from equation 10 using either linear of cubic specification of *Share* with 95% confidence intervals clustered by school-cohort.

Appendix B Calculations for Cost-Benefit Analysis

This paper shows that there are long-run benefits for black students of having a black teacher. This result is often used as motivation for calls to diversify the teacher workforce (i.e., to hire more black teachers). Currently, there are approximately 3.8 million K-12 teachers in the U.S., and only 256,000, or 6.7%, of them are black (NCES, 2017). One way to relatively quickly increase the fraction of teachers who are black is to induce black college graduates who are not teachers to become teachers. However, there are costs to such a policy that are sometimes overlooked by advocates of such policies. On average, black college graduates who are not teachers earn higher wages than those who are teachers, suggesting that if policymakers were able to somehow induce some of these individuals into teaching, they would suffer an income loss. Alternatively, we can view the difference in wages as the amount it would cost to induce such workers into teaching (i.e., a compensating wage differential).

Suppose the goal was to double the fraction of teachers who are black from 6.7%, or 256,000 to 13.4%, or 512,000. To calculate income distributions for black workers, we use data from the 2018 March CPS (Ruggles et al., 2018). We include all black individuals ages 21-65 who have at least a Bachelor's degree, worked for at least 26 weeks in 2017, whose primary occupation in 2017 was not in the armed forces, and who earned at least $1,000 and less than the top-coded value of $1,099,999 in their primary occupation in 2017. In this sample, the fraction of college educated blacks who are teachers is 8.3%. We next calculate average wage and salary income for blacks in our sample by occupation (i.e., teacher versus non-teacher). Average income for teachers is $51,129, for non-teachers is $65,888, and overall is $64,663. The income gap between black teachers and black non-teachers is $14,759, or 28.9%. Given this $14,759 gap between black teachers and non-teachers, and the current number of 256,000 black teachers, doubling the fraction of teachers would lead to a yearly loss of income of $3,778,302,000 from black college graduates, or $151,132,160,000 over a 40-year work life. This could be viewed as the amount of money it would take to double the number of black teachers over a 40-year long career.

There are a few reasons this basic calculation is likely an overestimate. First, average income of non-teachers includes those with doctoral degrees and professional graduate degrees who earn far more than teachers (for whom 88% have either a Bachelor's or Master's degree (NCES, 2017)), and would be unlikely to switch into teaching. Second, average income is skewed right by very high-income earners who disproportionately affect non-teacher average income, while teacher salaries tend to be compressed. Third, over three quarters of teachers are female (NCES, 2017), and females earn less than males, so the average income of non-teachers is higher because more of them are men.

We thus recalculate our statistics using median income for female workers who earned a Bachelor's degree but not higher than a Master's degree. Among blacks, median income for teachers is $45,000, for non-teachers is $49,000, and overall is $48,000. Given this difference in median income of $4,000, doubling the fraction of teachers who are black would lead to approximately $4,000 lower income for 256,000 black workers, or a total of $1,024,000,000 from black college graduates, or $40,960,000,000 over a 40-year work life.

This back-of-the-envelope calculation suggests that it would cost approximately $4,000 per year to induce (or compensate) one extra black college graduate into teaching. However, there are certainly many concerns with this simple calculation. For example, we do not attempt to focus on some subset of the non-teachers who may be most likely to switch into teaching. A more serious attempt at calculating this number might attempt to match teachers to non-teachers based on their observable characteristics. We leave such attempts to future research, though note that researchers have attempted similar calculations in the past, albeit not explicitly focused on black teachers, and come up with estimates similar to those reported here (Goldhaber, 2010).