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 Economics, Amherst College
Constance Lindsay¶
School of Education, UNC Chapel Hill
Nicholas W. Papageorge‖
Department of Economics, Johns Hopkins University, IZA and NBER

ABSTRACT: We examine the long-run impacts of exposure to a Black teacher for both Black and white students. Leveraging data from the Tennessee STAR class-size experiment, we show that Black students randomly assigned to at least one Black teacher in grades K-3 are 9 percentage points (13%) more likely to graduate from high school and 6 percentage points (19%) more likely to enroll in college than their same-school, same-race peers. No effect is found for white students. We replicate these findings using quasi-experimental methods to analyze a richer administrative data set from North Carolina. The increase in postsecondary enrollments is concentrated in two-year degree programs, which is somewhat concerning because two-year colleges have both lower returns and lower completion rates than four-year colleges and universities. These long-run effects are also concentrated among Black males from disadvantaged backgrounds, which is not evident in short run analyses of same-race teachers' impacts on test scores. These nuanced patterns are of policy relevance themselves and also underscore the importance of directly examining long-run treatment effects as opposed to extrapolating from estimated short-run effects.

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

---

*First draft: March 20, 2017. Current draft February 15, 2021. 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.
§jhyman@amherst.edu.
¶clindsay@unc.edu .
‖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. In particular, we examine the long-run impacts of having a Black teacher in elementary school on students' educational attainment, as measured by high school graduation, college aspirations, and actual college enrollments. Most of our focus is on race match effects, i.e., the impact of Black teachers on Black students. However, we also examine how assignment to a Black teacher affects white students. One reason is that calls to diversify the teaching workforce, in part due to the positive effects of Black teachers on Black students, would increase white students' exposure to Black teachers and it is important to understand how, if at all, they would be affected by a change in the composition of the teaching force.[2]

We identify arguably causal estimates by leveraging the random assignment of students and teachers to classrooms in the Tennessee STAR class size experiment. We find that Black students randomly assigned to at least one Black teacher in grades K-3 are 9 percentage points (13%) more likely to graduate from high school. They are 6 percentage points (19%)

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

[2] We do not examine the impact of having a white teacher because nearly all students we observe have at least one white teacher, which is due to the fact that the teaching force is disproportionately white. This affords us little variation to estimate long-run effects of exposure to white teachers. Put another way, we cannot assess the impact of a same-race teacher on white students as doing so would require a comparison group of white students who never see a same-race teacher and that group is vanishingly small. In contrast, there are many Black students in our sample whom we never observe with a Black teacher. Ideally, moreover, we would like to characterize the impact of race-congruent teachers on students from other minority groups. However, less than one-half of one percent of students in the Tennessee STAR dataset used for our main analyses were from racial groups other than Black and white.

more likely to enroll in college than their same-school, same-cohort Black peers who are not assigned a Black teacher. We also find that effects are concentrated among males and in schools with high proportions of disadvantaged students.

Positive impacts on the long-run outcomes of Black students are broadly consistent with improvements in short-run outcomes discussed in earlier literature, such as end-of-year test scores (Dee, 2004). However, direct examination of long-run effects reveals nuanced and surprising patterns we would otherwise fail to predict. First, increases in postsecondary enrollment are driven almost entirely by enrollments in two-year colleges and we are unable to conclude whether there are increases in degree completion. This merits further investigation as shorter postsecondary degree and certificate programs, and credit accumulation that does not lead to a credential–what the literature often refers to as "some college"–have lower returns in the labor market than bachelor's degrees. We thus connect our results with a burgeoning literature on the returns to "some college," which includes the types of shorter postsecondary pathways that are frequently observed among disadvantaged students, such as the majority of students in the STAR sample. Second, heterogeneity in long-run effects does not necessarily align with what short-run effects portend. Both short-run and long-run effects are concentrated among disadvantaged students. Yet, while short-run estimates suggest (albeit imprecisely) that there are stronger effects of a Black teacher on Black females' outcomes, we find stronger effects on males' longer-run outcomes. Meanwhile, while we find negative contemporaneous effects of exposure to Black teachers on test scores for white students (consistent with Dee (2004)), there is no evidence of negative impacts on white students' long-run educational outcomes. These kinds of nuanced patterns underscore why it is problematic to extrapolate long-run implications from analyses focused solely on short-run treatment effects.

We further investigate the long-run effects of Black teachers by assessing whether our findings replicate outside of the Tennessee STAR context. In particular, we use rich, longitudinal, administrative data on the population of North Carolina public school students. While Tennessee STAR data provide strong internal validity due to the experimental assignment of students to teachers, they are limited in terms of power and external validity. The North Carolina data require non-experimental methods to achieve identification, but sample sizes are larger, there are more background variables with which to examine heterogeneity, and they provide a useful replication of the STAR results for more recent cohorts. First, we replicate the main finding that exposure to a Black teacher in elementary school significantly improves the long-run educational outcomes of Black students, but has no impact on white students. Second, we show that the effects on college intent are entirely driven by the response of persistently disadvantaged students. Third, we document stronger effects of

Black male teachers on Black male students, and of Black female teachers on Black female students. These results provide guidance for the optimal allocation of scarce Black teachers and the importance of intersectionality in discussions of teacher diversity. More generally, these analyses replicate the Tennessee STAR results in another state, in another era, using a different research design.[3]

These results complement mounting evidence that same-race teachers are beneficial to underrepresented minority students on a number of contemporaneous dimensions, such as test scores, attendance, course grades, disciplinary outcomes, and expectations in a variety of educational settings (Dee, 2004, 2005; Fairlie et al., 2014; Gershenson et al., 2016; Lindsay and Hart, 2017; Holt and Gershenson, 2019). They are also consistent with well-established evidence that same-gender teachers and instructors affect educational outcomes: for example, by encouraging women to enter 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 closing stubbornly persistent attainment gaps.[5]

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).[6] 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.

We also highlight several questions that the current study raises and identify some priori-

[3]A recent working paper using data from Texas provides additional evidence that the long-run effects of access to same-race teachers are not specific to the unique STAR experimental data (Delhommer, 2019).

[4]An exception in the context of gender is Lim and Meer (2020), who show that effects of gender match in the 7th grade persist through high school. Similarly, Kofoed et al. (2019) explore same-race and same-gender peer effects on occupational choice.

[5]In this sense, our paper also contributes to a growing literature that revisits older, previously studied interventions to document long-run effects. In labor economics, examples include the long-run impacts of public housing demolition (Chyn, 2018), disruptive peers (Carrell et al., 2018), class size (Chetty et al., 2011; Dynarski et al., 2013), and the Head Start program (Deming, 2009; Garces et al., 2002), to name a few.

[6]More generally, our findings contribute to growing evidence that inputs received in primary school can affect 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).

ties for future work. Rates of college degree receipt are low and possibly undercounted in the STAR data, so we are unable to precisely estimate impacts on college completion. The lack of a clear impact on completion is potentially concerning, as while high school graduation is unambiguously beneficial, postsecondary enrollment absent completion has more modest returns in the labor market. And, again, our college enrollment results are primarily driven by enrollments in two-year degree programs. Thus, a perverse possibility is that Black teachers inspire students to attempt degree programs they are unable to complete, which means they incur many of the costs, but few of the benefits, of postsecondary education. We provide evidence against this case by connecting our results with a growing literature showing that there are indeed returns to non-traditional postsecondary pathways (i.e., those other than a four-year degree), especially for disadvantaged students for whom the alternative is generally no postsecondary education at all. We also find that about half of the Black students who enrolled in college likely enrolled in relatively high-paying programs. Nevertheless, future research should focus on how different school inputs, including teacher demographics, influence postsecondary choices, especially those with relatively low returns in the labor market.

A second concern is that causal estimates of race-match effects do not pinpoint why same-race teachers boost the educational attainment of Black students. The lack of evidence on mechanisms hinders policymakers' ability to effectively and efficiently act on the finding that same-race teachers matter. For example, it is difficult to assess whether Black teachers are more effective at conveying knowledge to Black students, serve as role models, hold and convey high expectations of their Black students, or whether some as yet unknown mechanism is at play. Understanding the channels, which are probably not mutually exclusive, would help in the development of policies that leverage our findings and in particular rethink the pre- and in-service training of white teachers. In Section 2, we describe these potential channels and those that we can safely rule out given the data we have. To draw sharper conclusions about mechanisms, however, we would need to collect additional data.

A third concern is that while our estimates suggest that diversifying the teacher workforce is a reasonable policy objective, it is not clear how that is to be done.[7] Creating a representative teaching workforce would require hiring roughly 250,000 Black teachers. One particular concern is that Black female college graduates who are not teachers earn roughly $4,000 more per year than Black female college graduates who are teachers. Given these pay differences, it is unclear whether and how quickly it is reasonable to expect a pipeline of Black teachers. In the meantime, policies must leverage the teaching workforce we have. The lack of an understanding of channels explaining race-match effects makes doing so difficult. This is another reason that future research and data collection efforts focused on identifying

[7]Gershenson et al. (2021) discuss the challenges and some possible strategies in greater detail.

channels explaining the Black teacher effect is vital.

The paper proceeds as follows. Section 2 provides a brief, multi-disciplinary overview of the pathways through which race-match effects may operate, and may operate differently by student race. Section 3 describes the STAR data and associated analyses. Section 4 presents the main STAR results. Section 5 describes the North Carolina data and associated analyses. Section 6 concludes.

2 Mechanisms: Effectiveness and Role Models

Prior to presenting our empirical analyses, we discuss the channels through which the long-run effects of same-race teachers on Black student outcomes might operate. A straightforward explanation is that in schools serving Black students, Black teachers are simply more effective than their white counterparts. This might occur because white teachers in schools with high shares of Black students 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, where Black teachers have three more years of experience, on average, than their same-school white counterparts. However, we rule out this explanation in the current study by showing that (i) the estimated race-match effect for Black students is robust to adjusting for teachers' observed qualifications, such as experience, and (ii) that random assignment to Black teachers has no impact on white students. The latter result is of policy interest 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. The teaching force in the U.S. is overwhelmingly white, so marginally increasing or decreasing same-race teacher assignments among white students is unlikely to affect their outcomes since they would still be exposed to at least some white teachers over the course of their primary and secondary education.[8]

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 specifically. This idea has received much attention outside of 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 roles and strategies of Black teachers in segregated and majority-Black schools. There is now a growing realization

[8] We do not study Hispanic students because no Hispanic teachers participated in the STAR experiment. However, using data from Texas Delhommer (2019) replicates our main findings and extends those findings to Hispanic students.

that the Brown v. Board of Education ruling and move to integration may have perversely harmed Black students by causing an exodus of Black women from the teaching profession once all-Black schools were legislated out of existence (Thompson, 2021). Kelly (2010) interviewed 44 former Black teachers in North Carolina and argues that while segregated Black schools were severely under-resourced in terms of supplies and physical capital, teachers in these schools were often highly effective, dedicated, and supported by the community.

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 curriculum, 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. The value of these teaching strategies to Black students is consistent with emerging evidence on the effectiveness of ethnic studies coursework (Dee and Penner, 2017) and programs such as the African American Male Achievement (AAMA) program (Dee and Penner, 2019).

Many of these ideas align with, or even motivate, the *identity economics* concepts for improving schools put forth by Akerlof and Kranton (2002).[9]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 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).

Teachers' biases can lead to decisions and behavior that profoundly affect student outcomes in both the short and long run: Card and Giuliano (2016), for example, show that when a large school district shifted from a referral-based system for identifying gifted and talented students to universal screening, this change significantly increased the numbers

[9]Identity economics links a person's sense of self, including their social group or category, to their economic behavior and outcomes.

of poor and minority students identified as eligible for the gifted program. Dougherty et al. (2015) show similar reductions in race-based gaps in identification for eighth-grade algebra under the implementation of policies that reduced teacher discretion in placements by adopting more neutral assignment policies. More broadly, Papageorge et al. (2020) show that biased teacher expectations 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' educational prospects than about Black students'. Finally, teachers' biases may not be implicit. In a different context (Italian teachers' grading and immigrant students' test scores) Alesina et al. (2018) show that some teachers state negative views about immigrants on a survey. They also exhibit bias against immigrant students when grading their tests and, when made aware of these biases, do not take efforts to correct them (even though implicitly biased teachers do). In general, biases can lead to decisions, practices, and behaviors that perpetuate inequality across racial groups.[10]

Another channel through which same-race teachers may matter is 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 translators and intercessors" (p. 51) for Black students and that these functions directly contribute to increased student achievement. Similarly, in his ethnographic work 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. Students who grow up in segregated environments or who have little contact with highly-educated people who look like them may conclude that postsecondary education is simply not available to them and approach their education accordingly. A Black teacher, an educated professional from the middle class, can thus provide students with a crucial counterexample to the view that higher education is out of reach. The potential power of demographic role models in the classroom–who can influence students' understanding of their choice sets and behaviors–is evidenced by a recent experiment in which exposure to a charismatic and successful female economics major increased female students' enrollment in economics (Porter and Serra, 2020).[11]

Effectiveness and role model effects are not mutually exclusive channels and it is entirely possible that both play a role. Yet, they are important to disentangle if the aim is to leverage

[10]In this sense, bias contributes to institutional racism, which refers to teachers' practices and attitudes, including denial of resources or low expectations, that may not be as overtly racist as other behaviors, but still harm Black students. Carroll Massey et al. (1975) elaborate on these ideas by studying a school district over 40 years ago.

[11]A similar intervention in French high schools increased STEM college enrollments (Breda et al., 2018).

race-match effects to develop policy. If Black teachers are more effective teachers for Black students, the focus should be on evaluating what particular practices and attitudes make them so and assessing whether these could be adopted by non-Black teachers. For example, if implicit bias undermines white teachers' effectiveness teaching Black students, identifying and reducing it should be a priority. On the other hand, if Black teachers are role models, there are other considerations. For example, role model effects are theoretically stronger when multiple characteristics are shared (Chung, 2000). This would suggest that Black male teachers might be better than Black female teachers for Black male students, an implication that is consistent with the results of our analysis of North Carolina data. If so, it may be a prudent to intentionally recruit Black men to teach Black male students. When the data allow, we will test for heterogeneity by student race, poverty status, gender, school type, and teacher gender. A role for role model effects also suggests that people other than teachers (e.g., guidance counselors and principals, as well as local business leaders or prominent figures in the community) could help to raise achievement and attainment by inspiring students.

While distinguishing among channels is important, doing so with existing data is difficult and indirect. We return to this point in the conclusion, when we discuss future research, which includes priorities for data collection efforts.

3 Project STAR Data and Methods

3.1 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 throughout the state 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. He 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 has yet to leverage this variation to examine the long-run impacts of having a same-race primary school teacher on educational attainment.[12] Our study thus extends prior work by estimating these long-run effects. 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) together, and concurrent absences and test scores with data on postsecondary educational enrollment and attainment from the National Student Clearinghouse (NSC) collected by Dynarski et al. (2013).

3.2 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 the 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

[12] Footnote 22 of Chetty et al. (2011) reports finding a positive but statistically insignificant effect of having a same-race teacher on earnings, but does not mention testing for heterogeneity by student race. Nor does the paper mention investigating the impact of having a same-race teacher on educational attainment.

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 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 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 3.4. 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. 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 among the Black students in our sample are so low, suggesting we are underpowered to provide conclusive evidence on 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 some 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, the estimate could be biased in either direction. This could be the case, for example, if colleges attended by marginal students are disproportionately undercounted by NSC. To address this possibility, 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 upwardly 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, or 5% of our total estimated effect.

3.3 STAR Data

Table 1 summarizes the main analytic sample of students who participated in Project STAR. Column 1 does so overall, while columns 2 and 3 do so by student race.[13] The main treatment of interest is an indicator for ever having had a Black teacher during Project STAR and the remaining columns of Table 1 summarize the data separately by treatment status.

Panel A of Table 1 summarizes students' baseline characteristics and exposure to Black teachers. Column 1 shows that the sample is 37% Black, 53% male, and 54% were eligible for free or reduced lunch (FRL). More than half entered in cohort 1 (kindergarten). Another baseline characteristic is an indicator for whether the student's name or date of birth (DOB) was missing. Overall, about 11% of the sample has a missing name or DOB, 31% of students had at least one Black teacher, and 20% had a Black teacher in their first year in STAR.

Columns 2 and 3 show some important differences between the Black and white students in the analytic sample. Specifically, Black students are more likely to receive FRL, to have entered in later cohorts, and to have had a Black teacher. The latter is due to teacher sorting,

[13] Less than one half of one percent of STAR students were a race other than Black or white.

with most Black teachers working in schools that have larger Black student enrollments (Dee, 2004). The latter difference, as well as demographic differences in cohorts, motivates aspects of our empirical approach, such as the inclusion of school-by-cohort fixed effects. In general, differences between the Black and white students in the STAR sample motivate us to fully interact all variables in the empirical model with student race.

Columns 4 and 5 split the sample by treatment status, i.e., by whether they ever had a Black teacher during their time in a STAR school. Many of the predetermined variables, such as student sex and cohort entry, are balanced, despite the fact that some student assignments in later years were affected by nonrandom non-compliance and attrition. Other variables, such as student race and FRL status are not balanced, as Black and FRL students are much more likely to be exposed to Black teachers. Again, this is expected due to student and teacher sorting to schools. Indeed, Columns 6 and 7 show that among Black students, the treated and comparison group students are more similar in terms of things like FRL status. Given evidence of non-random non-compliance in later years, we will instrument for treatment and conduct a formal balance test on the instruments rather than the treatment. Notably, our descriptive statistics do show substantial differences in the likelihood of missing an NSC link, with treated Black students much less likely to miss NSC links compared to their non-treated peers. In our balance tests, we show that these differences are effectively eliminated by the use of school-by-cohort fixed effects

Panel B of Table 1 summarizes the classroom and school characteristics. 26% of students were in small classes, and this is fairly similar for Black and white students and for treated and control students. There are some notable differences by student race and treatment status in observed teacher characteristics like holding a graduate degree and teaching experience, though again these differences can by explained by variation by teacher race in these characteristics and to the sorting of teachers and students to schools. Importantly, the main results are robust to controlling for observed teacher qualifications.

Finally, panel C of Table 1 summarizes several long-run educational outcomes of interest. Students' high school graduation status is only observed for 44% of the public-use sample and the missingness of these data is endogenous to treatment, which will shape our empirical approach to and interpretation of evidence on the long-run impact of Black teachers on high school graduation rates. Accordingly, we focus on Black teachers' effects on students' postsecondary educational outcomes. The public-use STAR data includes indicators for whether the student took the SAT or ACT college entrance exams, which are indicators of college intent. Like the NSC data, college entrance exam data is available regardless of whether students attrited from the experimental sample (Krueger and Whitmore, 2001). Just over a third of students in the sample sat for a college entrance exam, though columns

2 and 3 show a significant difference by student race that favors whites. Columns 4 and 5 show that overall, treated and control students took a college entrance exam at similar rates. However, columns 6 and 7 show that among Black students, the treated group was 9 percentage points more likely to take the SAT or ACT. This foreshadows our main results, which show large differences by student race in the impact of having a Black teacher.

The postsecondary educational outcomes from the NSC data, described in section 3.2 and by Dynarski et al. (2013), show that 39% 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, a point we will discuss further when presenting results and return to in the Conclusion. Columns 2 and 3 again show a racial gap in college enrollment that favors whites. And like the college entrance exam patterns discussed above, a notable difference between the enrollment rates of the treatment and control groups only appears in the Black student subsample.

3.4 Identification Strategy

Our empirical approach is motivated by the way in which the STAR experiment was conducted along with concerns about random assignment to treatment. The STAR experiment is notorious for experiencing significant attrition after the first year, which is likely non-random (Ding and Lehrer, 2010; Krueger, 1999). Thus, we cannot simply regress outcomes onto time-varying treatment occurring during grades K-3. However, there is good evidence that randomization was achieved and compliance was not an issue in students' first year in STAR.[14] One possibility, then, is to relate long-run outcomes to Kindergarten (or first year) teachers only.[15] However, students who were not assigned a Black teacher in kindergarten, and would thus be in the control group, may have faced a Black teacher in subsequent STAR years. Accordingly, our preferred approach uses all available years of the STAR data.

Specifically, our treatment of interest is a binary indicator of whether the student *ever* had a Black teacher in grades K-3. Non-random noncompliance after the student's first year in STAR means that this treatment is potentially endogenous. Moreover, endogenous attrition from STAR schools creates measurement error in the treatment variable because the race of students' teachers is only observed while they are in a STAR school. Accordingly, we

[14]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 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.

[15]This was the main approach taken in an earlier draft of the paper (Gershenson et al., 2018). As we explain below, it amounts to the "reduced form" version of our preferred IV specification.

follow Dee (2004) in instrumenting for the endogenous treatment with the expected number of Black teachers students would have had had they complied with the randomly assigned class type and remained in their initial STAR school for the entirety of Project STAR. Note that this expectation is not necessarily a whole number because a student assigned to a small class might be in a school with two small second grade classrooms, one taught by a Black teacher and one taught by a white teacher. In this case, their expected number of Black teachers in grade 2 would be 0.5. In Appendix A we show balance in the instruments using a regression-based balance test similar to that in Table 3 of Dee (2004).

Two other points about the data motivate our empirical approach. First, the outcomes of interest are student-specific and do not vary over time (e.g., college enrollment). As such, we cannot use the panel data models used in previous STAR studies of racial mismatch (Dee, 2004; Penney, 2017). Second, 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 between schools and between the cohorts within schools (Krueger and Whitmore, 2001). The latter is important, as children who enter a school in first grade likely opted out of voluntary kindergarten and children who enter in grades 2 or 3 are experiencing the disruption of a school change.

These concerns lead to a straightforward cross-sectional, instrumental-variables model. Our preferred model uses two instruments, where the "expected number of Black teachers" is split into certain (first year in STAR) and uncertain (subsequent year) components, though the results are robust to the exact specification and functional form of the first stage, including using only the "certain" first-year instrument. For example, for a student who entered STAR in kindergarten, the two instruments are (i) an indicator for whether they had a Black teacher in kindergarten ($Black1$) and (ii) the expected number of Black teachers they would have in grades 1-3 (which ranges continuously from 0-3), assuming perfect compliance with the randomly assigned class type and retention in the initial school ($Expected$). For a student who entered STAR in second grade, the second instrument would only count the expected number of Black teachers in grade 3, which would range continuously from 0 to 1, and so on.

Formally, the first-stage regression is

$$Ever_{igk} = \theta_{gk} + \pi_1 X_i + \pi_2 Black1_i + \pi_3 Expected_i + u_{igk}, \tag{1}$$

where i, g, and k index students, grade of entry, and schools, respectively; $Ever$ is the treatment indicator of having had at least one Black teacher; θ is a school-by-cohort fixed effect (FE), X is a vector of observed student and teacher characteristics including student's sex, race, and FRL status, teacher's experience, education, and certification status, and randomly-assigned class type (i.e., a small indicator). Both the FRL indicator and the teacher characteristic controls are for the student's first year in STAR.[16]

The second-stage (structural) model is

$$y_{igk} = \theta_{gk} + \beta X_i + \delta Ever_i + \epsilon_{igk}, \quad (2)$$

where y is the outcome and the parameter of interest is δ, which represents the Local Average Treatment Effect (LATE) of ever having a Black teacher during a student's time in a STAR school.

A few aspects of the model given by equations (1) and (2) merit further discussion. First, the reduced form effect of $Black1$ on student outcomes is interesting as well, as it shows the impact of having a Black teacher in a student's first year in STAR. Second, we estimate the model by 2SLS, which allows for the straightforward inclusion of the school-by-cohort FE. Finally, we estimate separate models for white and Black students because consistent with theory and previous empirical work, a regression-based Chow test finds the education production function given by equations (1) and (2) to systematically differ by race ($p < 0.001$). The Black teacher-Black student interaction term from the fully interacted model estimated with the pooled sample is reported as well, which represents (and provides a formal statistical significance test of) the difference between the white and Black sample estimates. We cluster standard errors by first-year classroom, as this is the level at which random assignments were made (Abadie et al., 2017), though clustering at higher levels yields similar results.

4 STAR Results

4.1 Main Results

Table 2 reports baseline 2SLS estimates of equation (2) for several outcomes associated with postsecondary educational attainment. Panels A and B estimate the model separately by student race. The interaction terms in panel C depict the differences between the Black and

[16]This is because the subsequent-year teachers are not necessarily randomly assigned due to noncompliance and not necessarily observed due to attrition. Similarly, looking at changes in FRL status is complicated by nonrandom attrition. In any case, the main results are quite robust to how, and even whether, the model adjusts for student and teacher covariates.

white estimates in panels A and B, and provide a robust t test of the significance of those differences. Generally, we see positive and significant effects for Black students, null effects for white students, and significant differences between the two.

In column 1, the outcome is an indicator for whether the student took the ACT or SAT college entrance exam. Taking a college entrance exam indicates college intent during the student's junior or senior year of high school. College intent is a particularly relevant outcome for economically disadvantaged students who comprise the majority of the STAR experiment's student population. However, it is 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). Panel A shows that Black students who have at least one Black teacher are 6.1 percentage points more likely to take a college entrance exam. This effect is statistically significant at the 95% confidence level and large in magnitude: it amounts to a 24% increase from the base test-taking rate. Panel B shows a negative point estimate that is statistically indistinguishable from zero for white students. The interaction term in panel C verifies that the effect of ever having a Black teacher is significantly larger for Black students than for white students.

Column 2 turns attention to an indicator for whether the student ever enrolled in any college (according to the NSC data). These results largely mirror those for test-taking reported in column 1. Black students who ever had a Black teacher are about about six percentage points (19%) more likely to ever attend college than their Black schoolmates who did not. There is no effect of ever having a Black teacher on white students' college enrollment, and once again the Black and white point estimates are significantly different from one another. This shows that college intent led to actual enrollments, and cross-validates the NSC data since the entrance-exam data comes from an independent source (Krueger and Whitmore, 2001).

Columns 3 and 4 re-estimate the college enrollment model separately for two-year and four-year enrollments. While we see positive point estimates for Black students in each type of institution, the main effect is clearly driven by enrollments in two-year colleges. This is intuitive, as institutions offering shorter programs (e.g., community colleges) are the most likely landing spots for students on the margin of pursuing postsecondary education. However, whether these enrollments translate to degree completion is unclear. Columns 5 and 6 of Table 2 show positive but imprecisely estimated effects on Black students' attainment, as measured by semesters enrolled and degree completion, respectively. This could reflect data limitations, i.e., small sample sizes and imperfect NSC coverage of both institutions and degrees, as well as small effect sizes.

The lack of a significant effect on degree completion and the concentration of enrollment effects in two-year colleges could be cause for concern. Compared to four-year degrees, two-year degrees and shorter certificate programs tend to be less lucrative and have lower completion rates. The latter might explain why we find no impact on degree completion. The general concern is that exposure to a Black teacher could lead students into postsecondary programs with few returns in the labor market or that they are ill-prepared to complete, which turn out to be bad investments.[17] However, while completed Bachelor's degrees generate higher returns, associate degrees, certification programs, and even completed community college credits generate wage increases that, on average, more than offset their costs(Liu et al., 2015; Minaya and Scott-Clayton, 2020; Kane and Rouse, 1995; Marcotte et al., 2005; Belfield and Bailey, 2011; Jepsen et al., 2014). For example, Kane and Rouse (1995) write that "A simple cost-benefit analysis shows that, over 30 years, the community college student who completes even only one semester will earn more than enough to compensate him for the cost of the schooling."

The returns to "alternative postsecondary pathways" vary by field of study as well (Bahr, 2019; Liu et al., 2015; Stevens et al., 2019), so it is worth investigating the type of programs that Black STAR students ultimately enrolled in. Unfortunately, the NSC data only record college major for 9% of Black students in the analytic sample, so we cannot do so directly. Instead, we make a back-of-the-envelope calculation by identifying the 20 most popular postsecondary institutions among Black STAR students and analyzing these institutions in the publicly available Integrated Post-secondary Education Data System (IPEDS). Specifically, we tally the degrees and certificates earned by Black students in those colleges in 2010.[18] Almost half (47%) of degrees and certificates earned by Black students at these institutions are in "high earning" fields. This provides additional suggestive evidence that on average, the observed enrollment effects are a positive outcome.

This discussion of how to interpret enrollment effects that are driven by enrollments in two-year colleges and do not accompany clear degree-completion effects encompasses a broader point that should be made explicit: one must be mindful of the relevant counterfactual when studying disadvantaged youth. While a four-year degree from a university is certainly valuable, it is not the modal outcome for the disadvantaged students who comprise the STAR sample and for whom the more likely alternative is no postsecondary education at

[17] These "alternative pathways" are often lumped together under the heading "some college," which includes shorter degree programs (e.g., associate degrees from a community college) and certification programs (e.g., phlebotomy or HVAC installation).

[18] 2010 is the first year of available data. We count all 38 program categories reported by IPEDS. Following (Dynarski et al., 2013), we consider high earning fields to include STEM and business majors, which constitute 12 of the 38 categories.

all. Indeed, some of the STAR students may well have aspired to obtain four-year degrees. This is consistent with our findings on the effects on college entrance-exam taking reported in column 1, which are similar in size to the two-year enrollment effects, since those entrance exams are usually not required for admission to two-year institutions. The reason could be that students were initially motivated to pursue four-year degrees, but for some financial, personal, or academic reason opted to begin their postsecondary experience in a two-year school and then never made the transition to a four-year program. If so, much work remains to be done to adequately support disadvantaged students on their path towards obtaining a four-year Bachelor's degree, should they aspire to do so. However, the existing literature on the returns to community college attendance makes clear that it is inaccurate to conclude that Black teachers provide no benefits to Black students just because they might facilitate alternative postsecondary pathways as opposed to the completion of Bachelor's degrees.

We next conduct several sensitivity analyses prompted by some of the concerns with the STAR data outlined in Section 3.3. In particular, we replicate results on college enrollment using different sample restrictions and modeling assumptions. Results are presented in columns 1-4 of Appendix Table A1. Column 1 presents estimates of the baseline 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 resulting estimates are qualitatively similar to the baseline estimates, reducing concerns that the imperfect coverage of STAR students in the NSC data drives the results.

Column 2 restricts the sample to the inaugural kindergarten cohort. We test this specification because the STAR experimental randomization is cleanest for kindergartners (Krueger, 1999). Indeed, Ding and Lehrer (2010) question whether later STAR entrants were randomly assigned, though we find no evidence that this is an issue in our sample. The estimates here are slightly larger in magnitude, but again show a positive and statistically significant effect for Black students and an imprecise estimate for white students.

Finally, columns 3 and 4 show that the main result is robust to how we control for class size and class composition. Specifically, column 3 replaces the randomly assigned classroom-type indicator with an exact count of class size. Following Krueger (1999), we account for possible endogeneity in exact class size by using the class-type indicators as instruments for realized class size. Column 4 adds the racial composition of the initial classroom to the model as an additional control. Again, both sets of results are nearly identical to the those for the baseline model.

4.2 High-School Graduation

Potential long-run effects of Black teachers on Black students' high school graduation rates are of first-order importance because 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. However, this analysis is hindered by the fact that high school graduation data are missing for more than half the analytic sample. This issue cannot be fully rectified and so these results should be interpreted with a healthy dose of caution.[19]

Table 3 estimates the baseline model (equation 2) for several outcomes associated with high-school graduation. Column 1 of Table 3 takes a sample-selection indicator as the outcome. Panel A shows that for Black students, random assignment to a Black primary school teacher significantly increases the likelihood of their high school graduation data being recorded in the Project STAR database. The point estimate of 0.066 indicates an 18% increase, which is practically significant. However, in Panel B we see no effect of random assignment to a Black teacher on white students' selection into the sample.

Intuitively, this positive selection into the sample among Black students is consistent with the positive impacts on college enrollment documented thus far, as the presence of high-school completion 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 Black students' long-run educational outcomes. Similarly, the lack of an effect on white students is consistent with the null results for white students' college enrollment documented above.

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 for the Black subsample in Panel A. In Panel B, the estimate for the white subsample is once again small and indistinguishable from zero. This suggests that the returns to having a Black teacher are similar for students whose high school graduation status was and was not observed, and lends at least some comfort in the use of these data.

Accordingly, we proceed to column 3 where we estimate the baseline model for high school graduation on the selected sample. These are naive estimates in the sense that no correction for sample selection is made. Consistent with the college enrollment results, we find a large,

[19]Specifically, high-school graduation data are missing for about 52% of white students and 63% of Black students. Appendix Table A2 summarizes the basic student data by high school graduation 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 class-size reductions do not investigate high school graduation (Dynarski et al., 2013; Krueger and Whitmore, 2001).

positive effect for Black students and a null effect for white students. For Black students, the point estimate of about 0.087 suggests that ever having a Black teacher in grades K-3 leads to a 13% increase in the likelihood of graduating high school, though this estimate is not statistically significant at traditional confidence levels. In column 4, we attempt to gain some precision by using a multiple imputation procedure to impute the missing high school graduation outcomes.[20] This yields a similar, yet more precisely estimated effect that is statistically significant. Once again, the effect for white students is smaller and statistically indistinguishable from zero.

Of course, multiple imputation does not eliminate selection bias if the dependent variable is not missing at random, so we also implement an "extreme assumptions" set of regressions in columns 5 and 6 where we replace all missing values with 0 and 1, respectively. As discussed above, students whose high-school outcome information is missing are less likely to have graduated, both because of their socio-demographic backgrounds and because missing these data implies that contact with TN public schools was lost. Thus replacing the missing values with zeroes is the more realistic "extreme imputation" approach. Indeed, the estimate in panel A of column 5 is qualitatively similar to the naive and MI estimates reported in columns 3 and 4, and statistically significant. The other extreme, which assumes that all of these students completed high school, is quite unrealistic and arguably represents a lower bound of the effect of having had at least one Black teacher on the likelihood of graduating from high school. These estimates are reported in column 6, where we see positive point estimates for both Black and white students, though both are smaller and statistically insignificant.

In sum, when combined with the results for college intent and college enrollment presented thus far, the estimates in Table 3 strongly suggest that exposure to a Black teacher in the early elementary grades increases Black students' chances of graduating from high school. We revisit this question and replicate this finding in section 5 using administrative data from North Carolina that are not prone to the missing data problems that plague Project STAR.

4.3 Exploration of Mechanisms

This section discusses additional analyses that help to shed light on the reasons that Black teachers improve the long-run educational outcomes of Black students. We rule out some possibilities (e.g., that Black teachers are more experienced). However, data limitations prohibit a full exploration of the mechanisms discussed in Section 2. We return to this point in the Conclusion when discussing priorities for future work.

One possible explanation of the main results is that Black teachers in STAR schools are

[20] We use a logit formulation of the selection equation and 40 imputations to construct these estimates.

simply more effective teachers than their white colleagues. However, if this were the case, we would expect exposure to Black teachers to boost student outcomes across the board, for white students as well as for their Black peers. Since we consistently find null results for white students, this explanation is not supported by our findings. This result is replicated in column 5 of Appendix Table A1, to serve as a point of reference. Again, this result is important on its own because it suggests that white students are not harmed by increased exposure to Black teachers (and the associated decrease in exposure to same-race teachers).

Another way to test this explanation is to omit the teacher characteristic control variables from the baseline model. Specifically, in column 6 of Appendix Table A1 we re-estimate the baseline model excluding the student and teacher control variables. The resulting estimates are nearly identical to the baseline estimates. Robustness to omitting student controls is to be expected given the random assignment of students to classrooms and is consistent with the balance tests reported in Table A3. However, the robustness to omitting teacher controls suggests that our main results are not driven by within-school racial differences in teachers' observable qualifications (e.g., experience (Wiswall, 2013)).

To continue to explore mechanisms, we next consider some intermediate outcomes, though the data we can use to do this are limited. We focus on student absences and test score performance. To begin, we document the effect of being randomly assigned to a Black teacher on both Black and white students' achievement and attendance. There are two reasons for doing so. First, while these effects are carefully documented elsewhere using the STAR data (Dee, 2004; Tran and Gershenson, 2021), it is useful to show that our analytic sample and identification strategy yield similar results. Second, showing these effects alongside those for college enrollment highlights that short-run effects on test scores and attendance do not necessarily imply long-run effects on college enrollment, which suggests that exposure to Black or same-race teachers might affect different student outcomes via different mechanisms. Indeed, 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: 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.

We cannot estimate the effect of Black teachers on test scores or absences using the same treatment used in equation (2), which captured whether each student had a Black teacher at least once in grades K-3, because test scores and absences are annual measures. Therefore, we estimate a cross-sectional model that is essentially the reduced form of the main instrument. Specifically, the treatment ($Black1$) is an indicator equal to one if the student had a Black teacher in their first year in a STAR school, and zero otherwise. We restrict the sample to students' first years in STAR to avoid concerns about noncompliance and attrition in later years, but otherwise control for the same student, classroom, and teacher controls and school-by-cohort fixed effects as the baseline IV model. Formally, we estimate by OLS models of the form

$$y_{igk} = \theta_{gk} + \beta X_i + \gamma Black1_i + \epsilon_{igk}, \tag{3}$$

where interest is in the coefficient γ. As in section 3.4, i, g, k index students, grade of entry, and schools, respectively; θ is a school-by-cohort fixed effect (FE), X is a vector of observed student, teacher, and classroom characteristics, and standard errors are clustered by first-year classroom.

Estimates of equation (3) are presented in Table 4. Panel A reports estimates for Black students and Panel B reports estimates for white students. Column 1 takes the end-of-grade math scale score as the outcome, which has an average score of about 510 and SD of 40. Panel A shows a positive, marginally significant effect of having a Black teacher on Black students' scores of about 5 points. Panel B shows a slightly larger, more precisely estimated negative effect of having a Black teacher on white students' scores. However, because all teachers in our sample are either white or Black, the negative effect in Panel B can equivalently be interpreted as a positive effect of a white (same-race) teacher on white students' math scores.[21] Both results are consistent with Dee (2004)'s analysis of the STAR data, which finds significant positive effects of same-race teachers on both Black and white students' test scores.

Column 2 takes the count of annual absences as the outcome, where the average student is absent about ten times per year.[22] Panel A shows that Black students matched to Black teachers have about 1.2 (13%) fewer absences per year, on average, and that this effect is strongly statistically significant. However, Panel B finds no discernible effect of teacher race on white students' absences. That teacher race affects Black students' attendance but not that of white students is consistent with Tran and Gershenson (2021), who thoroughly analyze the classroom determinants of student absences in Project STAR schools, as well as quasi-experimental evidence from North Carolina (Holt and Gershenson, 2019). This pattern is also consistent with the patterns observed in the main college enrollment results discussed

[21]This is because in the first three years of STAR all teachers were literally either white or Black. In the third grade there were 14 Asian teachers that we exclude from the analytic sample; however, including them does not qualitatively change any of the main results.

[22]The STAR data did not record absences in second grade, so this and subsequent analyses in Table 4 omit the second-grade cohort from the analytic sample.

in section 4.1. We document the enrollment result once again in column 3 of Table 4, using the identification strategy laid out in equation (3). Having a Black teacher in their first year in a STAR school increases Black students' chances of ever enrolling in college by about 4 percentage points (14%), but has no effect on white students' college prospects.[23]

To explore whether changes in attendance and test scores help to explain long-run effects of Black teachers, we next conduct a naive mediation analysis. Here, we include absences and test scores as additional control variables in equation (3). Column 4 conditions on math scores, which are significantly and positively correlated with college enrollment for both Black and white students. For Black students, adding this control reduces the estimated effect of having a Black teacher on the probability of enrolling by about one percentage point (25%). Similarly, Column 5 conditions on absences, which are negatively associated with college enrollment for both Black and white students. Doing so again reduces the estimated effect of having a Black teacher on the probability that Black students enroll in college by about one percentage point (25%). Finally, column 6 shows that for Black students, conditioning on both absences and achievement reduces the Black-teacher effect on college enrollment by 1.5 percentage points (34%).

One must take care in interpreting these results. Taken at face value, the mediation analyses suggests that about one-third of the treatment effect we estimate can be explained by fewer absences and higher test scores. An alternative explanation is that Black teachers affect omitted variables that jointly influence absences, test scores, and long-run educational attainment. For example, Black teachers might serve as role models for their students, which not only increases postsecondary enrollment, but also decreases absences.[24] This could occur even if absences have no bearing on postsecondary education. From a policy standpoint, this distinction is important because it determines whether or not reducing absences is a way to replicate the positive impact of having a Black teacher. Regardless of the interpretation, the results in columns 4-6 suggest that improved achievement and attendance explain at best a modest share of Black teachers' long-run effects on Black students' educational attainment. They also highlight the fact that previous knowledge of how teacher race affects test scores does not perfectly predict how, or even whether, teacher race will affect long-run educational outcomes like college enrollment.

[23] Note that these coefficients differ from our main results because here we are focusing on teacher race in the year of entry and not "ever exposed."

[24] The mediation analysis is prone to bias caused by what Acharya et al. (2016) call intermediate confounders and what Imai et al. (2010) call the failure of sequential ignorability: the mediators (absences and achievement) are themselves potentially affected by other unobserved mediators. The randomization of Project STAR does not eliminate this concern, because neither attendance habits nor academic ability were randomly assigned.

4.4 Heterogeneity

Table 5 explores potential heterogeneity in the effect of Black teachers on students' likelihood of ever enrolling in college by estimating the baseline model separately for different groups of Black and white students. Panel A reports estimates for Black students, where we largely see positive effects similar in size to the baseline estimates reported in column 2 of Table 2, although they are imprecisely estimated for some groups. Panel B reports estimates for white students, where we once again see relatively small, statistically insignificant point estimates.

Columns 1 and 2 of Table 5 estimate the model separately for male and female students, respectively. The effect of having at least one Black teacher on Black boys' probability of ever enrolling in college is almost twice as large as the effect for Black girls, and is strongly statistically significant. The effect for girls is imprecisely estimated, which is likely due to the drop in power, but remains positive and substantively meaningful at 0.05. That the Black boys in relatively disadvantaged STAR schools seem to benefit more from having a same-race teacher than their female counterparts is consistent with arguably causal research that finds gender differences in students' response to schooling inputs and environments (Figlio et al., 2016b).

Columns 3 and 4 estimate the baseline model separately by students' socioeconomic status, as proxied by their eligibility for free or reduced price lunch (FRL) in their initial STAR year. Interestingly, the effect of having a same-race teacher is more than twice as large for non-FRL Black students than for their FRL classmates. This could be because non-FRL students are closer to the college-going margin. However, neither coefficient is precisely estimated, in part because only about 15% of the Black sample is non-FRL. Because FRL is a transitory and imprecise marker of students' socioeconomic background, we also follow Dynarski et al. (2013) in estimating the baseline model separately by schools' socioeconomic status, as measured by the share of FRL students in the school. This arguably provides a broader measure of student background, neighborhood, and general resources available to them. Here, we see a larger and marginally significant effect in the majority-FRL schools.

We next ask whether this heterogeneity is reflected in analyses of short run outcomes. Appendix Table A4 reproduces the estimated effect of having a Black teacher in your first year in STAR on math scores for different student subgroups. For Black students, the race-match effect is larger for female than male students, which is not what we find when examining long-run outcomes. For white students there is no difference by gender. Another difference by race is that for Black students, the effect is driven almost exclusively by FRL students and students in relatively disadvantaged schools, while for white students the effect is observed for both FRL and non-FRL students, but primarily those in disadvantaged schools. This makes the lack of an effect on white students' college going, even in the most disadvantaged schools, all the more surprising and highlights the importance of examining long-run effects directly. Moreover, this underscores that our results are not driven by poverty, but by race.

In sum, Table 5 finds some suggestive evidence of heterogeneity in how Black students benefit in the long-run from having a Black teacher. Specifically, males and students in relatively disadvantaged schools stand to gain the most from having a same-race teacher. However, these differences are relatively small and imprecisely estimated, perhaps due to the small STAR sub-samples and accompanying lack of power; the general lack of variation in student background in the STAR sample, which was purposely composed of disadvantaged schools; or the relatively crude student-level data available in the STAR data. Interestingly, though, the null effect of Black teachers on white students' outcomes is robust across school and student background. Finally, evidence on short-run effects is not entirely consistent with long-run effects, underscoring the dangers of relying solely on short-run outcomes to infer long-run treatment effects. We reassess the question of heterogeneous effects in section 5 using administrative data for the entire population of North Carolina public school students, which provides a larger sample of students across the socioeconomic spectrum.

5 Replicating and Extending the STAR Results

5.1 Data and Methods

We replicate and extend the STAR results using student-level longitudinal administrative data on public school students in North Carolina who entered third grade between the 2000-2001 (2001) and 2004-2005 (2005) school-years.[25] 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., 2016a; 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).[26]

[25] An earlier version of this paper placed greater emphasis on these results (Gershenson et al., 2017).

[26] 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 within the years of data we have available.

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 do 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 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 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 all Black students in the grade, including those who are not in the teacher's self-contained classroom. 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, Black 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 on 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 A5 and find that with the exception of the overall share of Black students, which we directly control for in equation (4), 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

$$y_{ist} = \beta_1 X_i + \beta_2 W_{st} + \delta Share_{st} + \theta_s + \gamma_t + u_{ist}, \quad (4)$$

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).[27] 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 also provide interpretation in which 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 6 summarizes the analytic sample, which contains five cohorts of students in North Carolina who entered third grade for the first time between 2001 and 2005. These means are reported for our full Black (Column 1) and white (Column 2) samples, as well as for the "persistently disadvantaged" Black (Column 3) and white (Column 4) students. Following Michelmore and Dynarski (2017), the "persistently disadvantaged" category is defined as being designated as economically disadvantaged in each year the student is observed from grades 3-8, as these are the years that the economic disadvantage variables are observed for these cohorts of students.[28] The persistently disadvantaged sample is arguably more comparable to the STAR sample, which intentionally recruited schools serving disadvantaged communities. Finally, Columns 5 and 6 report means by sex among the persistently

[27] 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.

[28] The economic disadvantage designation is based on receipt of free or reduced-price lunch.

disadvantaged, Black subsample.

Panel A of Table 6 summarizes students' educational outcomes. The NCERDC data contain two "long run" measures associated with educational attainment, which serve as the dependent variables in equation (4). The first is an indicator for whether students are ever observed as dropping out of high school.[29] Roughly 13% of Black students are recorded as having dropped out of high school compared to 10% of whites, though this racial gap reverses in the persistently disadvantaged subsample. Columns 5 and 6 show a 7 point gender gap in favor of female students in dropout rates in the Black, persistently disadvantaged 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 indicates that the student either declared no intention of attending a four-year college or did not graduate from high school. Roughly 40% of Black students (and 42% of white students) 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 binary proxies for a student's postsecondary educational plans.[30] Consistent with national trends in college enrollment and completion (Bailey and Dynarski, 2011), college intent is lower among persistently disadvantaged students, and higher among females than males for the persistently disadvantaged Black subsample.

Panel B of Table 6 summarizes students' exposure to Black teachers. About 44% of Black students (but only 14% of white students) have at least one Black classroom teacher in grades 3-5. The modal number of Black teacher exposures in these grades is zero. The majority of students who do have a Black teacher have exactly one (about 30% for Black students). Only about 14% of Black students (and 2% of white students) have multiple Black teachers in grades 3-5. The persistently disadvantaged Black and white subsamples are exposed to Black teachers at about the same rates as the full samples. 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 (4). That said, there is a strong, mechanical

[29] 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.

[30] Ideally, we would like to use a measure of college test-taking in this sample to directly compare to the entrance-exam results in Project STAR. However, ACT data is available only for the final cohort, which prevents us from doing so. In contrast, the STAR data captures taking either the ACT or the SAT. Moreover, North Carolina adopted a policy that required all students to take the ACT in 2012-2013, which would affect our final cohort.

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. For Black students, the average cohort's teacher pool was about 25% Black, with an in-school standard deviation of about ten percentage points.[31] By comparison, for white students the average cohort's teacher pool was only about 8% Black. Again, for each race group, the persistently disadvantaged group looks similar to the full sample on this measure.

Panel C of Table 6 summarizes the students themselves. About 45% of Black students, and 12% of white students, were persistently economically disadvantaged. For Black (white) students, 85% (38%) were considered economically disadvantaged at least once between grades 3-8. About 11% of both Black and white students had exceptionalities, proxied by the presence of an IEP (individualized education plan). While the rates of exceptionality are slightly higher in the persistently disadvantaged subgroups, the most notable differences are based on gender disparities: Compared to their female counterparts, persistently disadvantaged Black males are about twice as likely to be identified with learning exceptionalities. About 12% of Black students (and 35% of white 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 Results

Table 7 presents estimates of equation (4), which identify the reduced-form effect of the racial composition of schools' teaching staffs on students' long-run educational outcomes.[32] 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, Black teachers

[31] 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.

[32] Appendix Table A6 shows that these results are robust to a variety of modeling decisions. Panel A replicates the main results for ease of comparison and also reports standard errors clustered at the school level. Panel B restricts the sample to schools that exhibited variation in the share of Black teachers. This excludes about 6,000 students from the Black student sample and, given the large share of schools serving white children with no Black teachers, cuts the sample size for white students by more than half. The point estimates are robust, which is to be expected given that the baseline model conditions on school fixed effects. Panel C 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 robust, suggesting that unobserved trends are not driving the results. Finally, panel D reports the FE-logit version of equation (4) that accounts for the binary nature of the outcomes. Once again, for both the Black and white persistently disadvantaged samples, the main results are robust.

could plausibly affect student outcomes beyond their own classrooms as well, by acting as mentors and advocates for Black students throughout the grade level and by supporting their fellow teachers.[33]

Panel A of Table 7 estimates equation (4) 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. Panel B replicates the same specification for the white student sample. Consistent with the STAR sample, we see null effects across both outcomes and by sex for white students. Panel C modifies the basic specification, pooling the Black and white student samples and including the interaction of each covariate with a Black student indicator. As in the STAR results, the regression-based Chow test of the joint significance of these interaction terms supports estimating separate models for Black and white students ($p < 0.001$), and so moving forward we stratify by race.

While the results from Panels A-C show effects on Black and white students from all economic strata statewide, recall that the STAR experiment targeted disadvantaged schools. To replicate the STAR findings, we now turn to the subset of persistently disadvantaged students. Panel D of Table 7 restricts the sample to Black students who were considered economically disadvantaged in each of grades 3-8. In columns 1-3 we see a larger effect on high school dropout than in the full sample of Black students, and once again the effect on dropout is entirely driven by male students. Columns 4-6 show significant and positive 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 2.2%, respectively, off the base rates of 17.8% and 28.2% for these measures).

In contrast, we continue to see null effects when we look at the persistently disadvantaged sample of white students (Panel E). Because the primary goal here is to replicate the STAR results, and because the effects for Black students seem to be concentrated among persistently disadvantaged students, all subsequent analyses are restricted to the persistently disadvantaged sample.

Panel F estimates an augmented version of equation (4) that distinguishes the share

[33] 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 learn to more effectively educate Black students from Black teachers in their grade or school.

of Black male teachers from the share of Black female teachers for our Black, persistently disadvantaged subsample. This was not possible in the STAR data because nearly all teachers were female. As in the baseline model, columns 1-3 of panel F 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 D. 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 the baseline estimate reported in panel D. However, unlike the high-school dropout results, columns 5 and 6 show stark differences by student sex in how students' college intent responds 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 Black male students' college intent, a difference that is marginally significant. Similarly, column 6 shows that the Black female students' college intent is only affected by the share of Black-female teachers.

The gender differences observed in columns 4-6, especially for male students, suggest at least some role for the role-model phenomenon presented in section 2. The reason is that the ability to teach using culturally-relevant pedagogy or hidden curricula 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.

Finally, panel G provides another heterogeneity analysis that might provide some suggestive evidence on the channels through which Black teachers improve Black students' long-run educational outcomes. Specifically, we test whether such effects were larger in counties with higher unemployment rates.[34] There are two potential, non-mutually exclusive reasons that unemployment rates may moderate the effects of exposure to Black teachers. First, the salience of Black teachers as role models could be greater in areas with higher unemployment rates, as students in these areas might see fewer successful professionals and more adults struggling to find employment. Second, it could be that the impact of Black teachers,

[34] We thank an anonymous reviewer for suggesting this analysis. Unemployment data come from Local Area Unemployment Statistics collected by the Bureau of Labor Statistics. Average county-level unemployment data from September 2000 (when the first cohort is entering third grade) to August 2007 (when the final cohort is about to enter sixth grade) are averaged across school years. "High" and "Low" unemployment counties are defined by whether their time-averaged unemployment rate was above or below the median county-level unemployment rate.

and schooling inputs more generally, are moderated by local economic conditions, as students are known to seek postsecondary education when economic opportunities are limited (Clark, 2011). For persistently disadvantaged Black boys, column 2 shows that the effect on high school dropout in high-unemployment counties was 13 percentage points, more than double the effect in low-unemployment rate counties, and this difference is marginally statistically significant. Similarly, in column 5 the effect on boys' college intent is twice as large in high- relative to low-unemployment counties, though this difference is imprecisely estimated. While not conclusive, these patterns are consistent with those in panel F suggesting that role-modeling plays at least some role in explaining the main results. More generally, this is an interesting source of heterogeneity that merits consideration in future research on the impacts of educational interventions and the channels through which those interventions operate.

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 8. 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 7, 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 0.10, which is slightly larger than the effects of around 0.06 on SAT/ACT taking and college enrollment reported in Table 2. Taken together, these North Carolina results corroborate the basic finding in the STAR analyses: exposure to even one Black teacher in primary school significantly increases the odds that economically disadvantaged Black students aspire to, and enroll in, college.[35]

How credible are the IV estimates? Black et al. (2017) describe an intuitive test, which

[35]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 8, are reported in Appendix Table A7. The results are qualitatively similar, suggesting that the IV results are not driven by a linear functional form.

amounts to estimating the reduced form (equation 4) 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 of Table 8. Intuitively, if the instrument is valid, the share of Black teachers should not significantly affect the outcome among individuals who are not treated.[36] 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 A5, 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 7 and even so, the IV estimates might not be too far off the mark. Indeed, the similarity with the STAR estimates suggests as much.

6 Conclusion

We provide causal evidence that Black students who have at least one Black teacher in elementary school are 9 percentage points (13%) more likely to graduate high school and 6 percentage points (19%) 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 and the magnitudes are large enough to be economically relevant. We generate similar results using administrative data from North Carolina, at a later time period, and with 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 students' exposure to Black teachers. While each data set and identification strategy has its weaknesses, together they suggest a meaningful impact of same-race teachers for Black students. Moreover, replication means key findings are not limited to a specific state, time, or experimental setting.

These findings suggest some cause for optimism, as they suggest a path to reducing stubbornly persistent racial attainment gaps. However, they raise at least three concerns that require further research.

First, while our findings on high school completion can be viewed as an unqualified

[36] 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.

benefit, our findings on postsecondary enrollment are less straightforward. While we see increased enrollment, our results are too imprecise to detect whether there are corresponding increases in the likelihood of completing a college degree. The enrollment result is driven by enrollment in two-year programs, which tend to have lower returns than four-year degrees and also lower degree-completion rates (Minaya and Scott-Clayton, 2020), though there are likely some modest returns to college coursework that does not culminate in a degree (Jepsen et al., 2014; Liu et al., 2015). Thus, it is possible that exposure to Black teachers encourages some Black students to make costly educational investments that do not pay off, which is a potential downside that deserves further exploration.

Unfortunately, the data do not allow us to examine the factors that might further contextualize the college enrollment results. Earlier research suggests that there are benefits to alternative pathways (e.g., associate degrees or coursework without a degree), but labor market returns vary by field and credential. For example, the return on investments in two-year college attendance and completion may depend on whether students enrolled in community colleges or for-profit institutions; the latter are more costly for students (Cellini, 2012) and may have lower returns than do non-profit institutions (Deming et al., 2012). Given that alternative postsecondary pathways have been (and continue to be) a likely outcome for the types of disadvantaged Black students most affected by Black teachers in the STAR context, future data collection and research efforts should focus on which particular programs students sort into and, if needed, explore policies that could leverage same-race teacher benefits in a manner that helps guide students towards pathways with higher returns.

Second, while we provide compelling evidence that some exposure to Black teachers improves Black students' long-term academic outcomes, identifying the exact mechanisms through which these effects operate is an important exercise that lies outside the scope of the current paper. Future work should further explore these mechanisms, as specific policy recommendations ultimately hinge on the mechanisms at play. For instance, if Black teachers primarily improve student outcomes by serving as role models, policies should provide students with more exposure to Black teachers, and to Black professionals more generally. Indeed, role models need not be teachers, but could include other professionals in the community and college graduates from the school who can cause students to update their beliefs.[37]

If effectiveness teaching Black students is the main channel, other sets of policies could be explored, including training the existing, largely non-Black teaching workforce to better

[37]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 increase the likelihood that female students apply to selective science majors in college (Breda et al., 2018).

serve Black students and students of color more generally. 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. Of course, this should not come at the expense of efforts to diversify the teaching force and make it more representative of the student body it serves. These policies are not mutually exclusive and will likely work well in concert. Extant data do not allow us to distinguish between these two channels (or among other channels that are as yet unknown). Future data collection should focus on identifying these channels. This is by no means a straightforward task since it is not clear how to measure role model effects versus other channels. An initial effort could involve methods more typically used in other fields, such as ethnographic data collection via observation or open-ended interviewing of students and teachers.

Third, our findings raise questions surrounding efforts to diversify the teaching workforce. For example, while our study provides support for the idea that diversifying the teaching work force could *ceteris paribus* increase high school completion and college enrollment rates, a pipeline that could fulfill massive, near-term growth in the number of Black teachers is not currently in place (Putman et al., 2016; Gershenson et al., 2021). Hiring practices that attempt to diversify the teaching force while maintaining high teacher quality would thus necessitate, for example, re-allocating college-educated Black professionals 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 get us close to aligning the racial composition of the work force with the student body they 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 Black workers who are not teachers is roughly $49,000 while median earnings for Black teachers is $45,000 (Ruggles et al., 2018).[38] 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 Black workers by more than one

[38]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).

billion dollars.[39].

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 worked in North Carolina, where a \$1,800 bonus reduced teacher turnover rates by 14% (Clotfelter et al., 2008). Broadly, policies based on our findings must be evaluated in light of their benefits and their costs, especially if the costs are borne largely by Black college graduates who would become teachers.

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.

Acharya, Avidit, Matthew Blackwell, and Maya Sen, "Explaining causal findings without bias: Detecting and assessing direct effects," *American Political Science Review*, 2016, *110* (3), 512–529.

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.

Alesina, Alberto, Michela Carlana, Eliana La Ferrara, and Paolo Pinotti, "Revealing Stereotypes: Evidence from Immigrants in Schools," Mimeo, Bocconi University 2018.

Bahr, Peter Riley, "The labor market returns to a community college education for non-completing students," *The Journal of Higher Education*, 2019, *90* (2), 210–243.

Bailey, Martha J. and Susan M. Dynarski, "Inequality in postsecondary attainment," in "Whither Opportunity: Rising Inequality, Schools, and Children's Life Chances," New York: Russell Sage Foundation, 2011, pp. 117–132.

[39] $\$4{,}000 \times 256{,}000$ = \$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)

Belfield, Clive R. and Thomas Bailey, "The benefits of attending community college: A review of the evidence," *Community College Review*, 2011, *39* (1), 46–68.

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.

_ and Laura Giuliano, "Universal screening increases the representation of low-income and minority students in gifted education," *Proceedings of the National Academy of Sciences*, 2016, *113* (48), 13678–13683.

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.

_ , Mark Hoekstra, and Elira Kuka, "The long-run effects of disruptive peers," *American Economic Review*, 2018, *108* (11), 3377–3415.

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.

Cellini, Stephanie Riegg, "For-profit higher education: An assessment of costs and benefits," *National Tax Journal*, 2012, *65* (1), 153.

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.

Chyn, Eric, "Moved to opportunity: The long-run effects of public housing demolition on children," *American Economic Review*, 2018, *108* (10), 3028–56.

Clark, Damon, "Do recessions keep students in school? The impact of youth unemployment on enrolment in post-compulsory education in England," *Economica*, 2011, *78* (311), 523–545.

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.

Dee, Thomas and Emily Penner, "My Brothers Keeper? The Impact of Targeted Educational Supports," Technical Report, National Bureau of Economic Research 2019.

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 Emily K. Penner, "The causal effects of cultural relevance: Evidence from an ethnic studies curriculum," *American Educational Research Journal*, 2017, *54* (1), 127–166.

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

Delhommer, Scott, "Role Models and Minority College Achievement," Mimeo, University of Texas at Austin 2019.

Deming, David, "Early childhood intervention and life-cycle skill development: Evidence from Head Start," *American Economic Journal: Applied Economics*, 2009, *1* (3), 111–34.

Deming, David J., Claudia Goldin, and Lawrence F Katz, "The for-profit postsecondary school sector: Nimble critters or agile predators?," *Journal of Economic Perspectives*, 2012, *26* (1), 139–64.

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.

Dougherty, Shaun M., Joshua S. Goodman, Darryl V. Hill, Erica G. Litke, and Lindsay C. Page, "Middle school math acceleration and equitable access to eighth-grade algebra: Evidence from the Wake County Public School System," *Educational Evaluation and Policy Analysis*, 2015, *37* (1_suppl), 80S–101S.

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.

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.

_ , _ , Jeffrey Roth, Melanie Wasserman et al., "School quality and the gender gap in educational achievement," *American Economic Review*, 2016, *106* (5), 289–95.

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.

Garces, Eliana, Duncan Thomas, and Janet Currie, "Longer-term effects of Head Start," *American Economic Review*, 2002, *92* (4), 999–1012.

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

— , — , Joshua Hyman, Constance Lindsay, and Nicholas W Papageorge, "The long-run impacts of same-race teachers," Technical Report, National Bureau of Economic Research 2018.

— , Michael Hansen, and Constance Lindsay, *Teacher diversity and student success*, Cambridge: Harvard Education Press, 2021.

— , 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.

Holt, Stephen B. and Seth Gershenson, "The impact of demographic representation on absences and suspensions," *Policy Studies Journal*, 2019, *47* (4).

Imai, Kosuke, Luke Keele, and Dustin Tingley, "A general approach to causal mediation analysis.," *Psychological Methods*, 2010, *15* (4), 309.

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.

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.

Jepsen, Christopher, Kenneth Troske, and Paul Coomes, "The labor-market returns to community college degrees, diplomas, and certificates," *Journal of Labor Economics*, 2014, *32* (1), 95–121.

Kane, Thomas J. and Cecilia Elena Rouse, "Labor-market returns to two-and four-year college," *The American Economic Review*, 1995, *85* (3), 600–614.

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

Kofoed, Michael S. et al., "The Effect of Same-Gender or Same-Race Role Models on Occupation Choice Evidence from Randomly Assigned Mentors at West Point," *Journal of Human Resources*, 2019, *54* (2), 430–467.

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.

Lim, Jaegeum and Jonathan Meer, "Persistent effects of teacher-student gender matches," *Journal of Human Resources*, 2020, *55* (3), 809–835.

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.

Liu, Vivian YT, Clive R. Belfield, and Madeline J. Trimble, "The medium-term labor market returns to community college awards: Evidence from North Carolina," *Economics of Education Review*, 2015, *44*, 42–55.

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.

Marcotte, Dave E., Thomas Bailey, Carey Borkoski, and Greg S. Kienzl, "The returns of a community college education: Evidence from the National Education Longitudinal Survey," *Educational Evaluation and Policy Analysis*, 2005, *27* (2), 157–175.

Massey, Grace Carroll, Mona Vaughn Scott, and Sanford M. Dornbusch, "Racism without racists: Institutional racism in urban schools," *The Black Scholar*, 1975, *7* (3), 10–19.

Michelmore, Katherine and Susan Dynarski, "The gap within the gap: Using longitudinal data to understand income differences in student achievement," *AERA Open*, 2017, *3* (1), 1–18.

Minaya, Veronica and Judith Scott-Clayton, "Labor Market Trajectories for Community College Graduates: How Returns to Certificates and Associate's degrees Evolve Over Time," *Education Finance and Policy*, 2020, pp. 1–62.

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.

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

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," *Review of Economics and Statistics*, 2020, *102* (2), 234–251.

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

Porter, Catherine and Danila Serra, "Gender differences in the choice of major: The importance of female role models," *American Economic Journal: Applied Economics*, 2020, *12* (3), 226–54.

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.

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

Stevens, Ann Huff, Michal Kurlaender, and Michel Grosz, "Career Technical Education and Labor Market Outcomes Evidence from California Community Colleges," *Journal of Human Resources*, 2019, *54* (4), 986–1036.

Thompson, Owen, "School desegregation and Black teacher employment," *The Review of Economics and Statistics*, 2021.

Tran, Long and Seth Gershenson, "Experimental Estimates of the Student-Attendance Production Function," *Educational Evaluation and Policy Analysis*, 2021.

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.

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

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

Figures and Tables

Table 1: Analytic Sample Means

Students:	All (1)	Black (2)	White (3)	Treated All (4)	Control All (5)	Treated Black (6)	Control Black (7)
A. Student Characteristics							
Black	0.37	1.00	0.00	0.73	0.20	1.00	1.00
Male	0.53	0.52	0.53	0.52	0.53	0.52	0.53
FRL	0.54	0.82	0.38	0.68	0.47	0.81	0.83
Missing NSC link	0.11	0.12	0.11	0.07	0.13	0.07	0.19
cohort 1	0.55	0.50	0.59	0.54	0.56	0.52	0.47
cohort 2	0.20	0.21	0.19	0.21	0.19	0.21	0.21
cohort 3	0.14	0.17	0.12	0.16	0.13	0.17	0.16
cohort 4	0.11	0.12	0.10	0.09	0.12	0.10	0.15
$\geq$ 1 Black T (treated)	0.31	0.63	0.13	1.00	0.00	1.00	0.00
Black T in year 1	0.20	0.44	0.06	0.62	0.00	0.69	0.00
Expected Black Ts yrs 2-4	0.44	0.94	0.16	0.93	0.21	1.07	0.70
B. Classroom & School Characteristics							
Small class	0.26	0.24	0.27	0.24	0.27	0.24	0.24
Regular class	0.37	0.38	0.37	0.39	0.37	0.39	0.38
Regular + Aide	0.36	0.38	0.36	0.36	0.36	0.37	0.38
Class size	21.01	21.43	20.77	21.32	20.87	21.35	21.58
T Grad degree	0.36	0.31	0.39	0.29	0.39	0.25	0.40
T Experience	10.76	10.56	10.85	11.01	10.64	11.18	9.49
Low-income school	0.49	0.81	0.30	0.74	0.37	0.87	0.71
C. Long-Run Outcomes							
HS observed	0.44	0.37	0.47	0.42	0.45	0.40	0.32
HS grad	0.77	0.67	0.82	0.72	0.80	0.69	0.65
Took SAT/ACT	0.34	0.27	0.38	0.33	0.34	0.30	0.21
College Enrollment	0.39	0.32	0.43	0.38	0.40	0.35	0.28
Two-year Enrollment	0.27	0.22	0.30	0.26	0.28	0.24	0.19
Four-year Enrollment	0.25	0.20	0.27	0.24	0.25	0.22	0.17
Semesters Attempted	3.14	2.56	3.48	2.98	3.21	2.79	2.14
Graduated	0.16	0.09	0.19	0.12	0.17	0.10	0.08
N	11,245	4,064	7,135	3,522	7,723	2,578	1,486

Notes: Sample size (N) refers to full analytic sample; means for high school (HS) graduation only reported for those whose HS records are observed. T stands for teacher. Low-income school is defined as more than 48% (sample median) of a school's students being eligible for free lunch (FRL) as in Dynarski et al. (2013). Missing NSC link refers to missing the student's name or date of birth, which complicates the National Student Clearinghouse data merge.

Table 2: Long-Run Effects of Ever Having a Black Teacher on Educational Attainment

Outcome:	Took SAT/ACT (1)	Ever College (2)	Ever 2-yr (3)	Ever 4-yr (4)	Semesters (5)	Degree (6)
A. Black Students						
$\geq$ 1 Black T	0.061**	0.059**	0.062**	0.015	0.279	0.004
	(0.026)	(0.027)	(0.025)	(0.021)	(0.300)	(0.018)
N (students)	4,064	4,064	4,064	4,064	4,064	4,064
R^2	0.063	0.058	0.034	0.047	0.066	0.049
$E(y)$	0.252	0.313	0.212	0.194	2.434	0.0861
N (classrooms)	638	638	638	638	638	638
B. White Students						
$\geq$ 1 Black T	-0.029	-0.019	-0.016	-0.024	-0.443	-0.030
	(0.032)	(0.035)	(0.027)	(0.027)	(0.321)	(0.024)
N (Students)	7,135	7,135	7,135	7,135	7,135	7,135
R^2	0.095	0.075	0.048	0.053	0.069	0.048
$E(y)$	0.384	0.435	0.303	0.276	3.526	0.197
N (classrooms)	969	969	969	969	969	969
C. All Students						
$\geq$ 1 Black T	-0.029	-0.019	-0.016	-0.024	-0.443	-0.030
	(0.031)	(0.034)	(0.027)	(0.027)	(0.315)	(0.023)
$\geq$ 1 Black T $\times$ Black S	0.089**	0.078*	0.078**	0.039	0.722*	0.034
	(0.041)	(0.043)	(0.036)	(0.033)	(0.417)	(0.029)
N (students)	11,245	11,245	11,245	11,245	11,245	11,245
R^2	0.189	0.162	0.116	0.141	0.156	0.137
$E(y)$	0.350	0.404	0.280	0.255	3.248	0.168
N (classrooms)	1261	1261	1261	1261	1261	1261

Notes: 2SLS estimates of the impact of ever having a Black teacher (Black T) in grades K-3, as described in equations (1) and (2). 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 free-lunch 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. The pooled models in panel C fully interact all covariates and school-by-year fixed effects with the Black student (Black S) indicator; a Chow (joint F) test of these interaction terms finds them to be strongly significant ($p < 0.001$) in all six models, suggesting that the education production function is systematically different for white and Black students in the STAR schools. We do not report the coefficient on the Black S variable because it is not directly interpretable, due to these interactions. *** p<0.01, ** p<0.05, * p<0.1.

Table 3: Effect of Ever having a Black Teacher on High School Graduation

Outcome:	Selected	College	HS Grad	HS Grad	HS Grad	HS Grad
Imputation:			none	MI	all 0	all 1
	(1)	(2)	(3)	(4)	(5)	(6)
A. Black Students						
$\geq$ 1 Black Teacher	0.066**	0.068	0.087	0.098**	0.077***	0.011
	(0.028)	(0.062)	(0.056)	(0.049)	(0.025)	(0.018)
$E(y)$	0.368	0.541	0.672	0.569	0.248	0.881
N (students)	4,064	1,496	1,496	4,100	4,064	4,064
N (classrooms)	638	474	474		638	638
R^2	0.034	0.068	0.048		0.053	0.003
B. White Students						
$\geq$ 1 Black Teacher	0.031	-0.021	0.018	0.042	0.041	0.010
	(0.028)	(0.059)	(0.043)	(0.053)	(0.026)	(0.018)
$E(y)$	0.480	0.563	0.823	0.751	0.394	0.915
N (students)	7,135	3,366	3,366	7,134	7,135	7,135
N (classrooms)	969	758	758		969	969
R^2	0.029	0.081	0.072		0.055	0.015

Notes: All models in this table are estimated by 2SLS as described in equations (1) and (2). The outcome in column 1 is a selection indicator that equals one if the high-school (HS) graduation outcome is observed, and zero otherwise. The outcome in column 2 is the main college enrollment indicator, where the baseline model is estimated on the selected sample. The outcome in columns 3-6 is an indicator for HS graduation. Standard errors are clustered by students' first-year classrooms. In column 4 HS graduation is imputed using a logit multiple imputation (MI) procedure (40 imputed data sets). *** p<0.01, ** p<0.05, * p<0.1.

Table 4: Do Short-Run Effects of Black Teachers Predict Long-Run Effects?

Outcome:	Math Score (1)	Absences (2)	College (3)	College (4)	College (5)	College (6)
A. Black Students						
Black T	4.993*	-1.210***	0.044**	0.033*	0.034*	0.029
	(2.565)	(0.301)	(0.019)	(0.018)	(0.019)	(0.018)
Math score				0.002***		0.002***
				(0.000)		(0.000)
Absences					-0.006***	-0.006***
					(0.001)	(0.001)
N (students)	3,664	3,310	3,383	3,103	3,310	3,066
R^2	0.646	0.094	0.135	0.176	0.152	0.188
$E(y)$	499.9	8.866	0.320	0.324	0.320	0.323
N (classrooms)	620	496	501	494	496	490
B. White Students						
Black T	-6.748**	-0.449	0.014	0.027	0.011	0.021
	(3.274)	(0.622)	(0.040)	(0.040)	(0.041)	(0.039)
Math score				0.003***		0.003***
				(0.000)		(0.000)
Absences					-0.004***	-0.003***
					(0.001)	(0.001)
N (students)	6,400	6,157	6,296	5,778	6,157	5,698
R^2	0.588	0.118	0.149	0.208	0.157	0.211
$E(y)$	519.8	10.11	0.442	0.452	0.446	0.454
N (classrooms)	920	724	735	709	724	700

Notes: All models in this table are estimated by OLS, where the Black T (teacher) variable is an indicator equal to one if the student had a Black teacher in their first year in a STAR classroom, and zero otherwise. The outcome in columns 1 and 2 are the scaled math score and count of annual absences, respectively, for the student's first year in a STAR classroom; in columns 3-6 the outcome is an indicator for ever having enrolled in college. Standard errors are clustered by students' first-year classrooms. All models control for school-by-cohort fixed effects and the full set of student and teacher contorls. Absences are not observed for the second-grade cohort, so columns 2-6 exclude the second-grade cohort from the analytic sample. *** p<0.01, ** p<0.05, * p<0.1.

Table 5: Heterogeneity in the Effect of Having a Black Teacher on College Enrollment

Sample:	Male (1)	Female (2)	FRL (3)	Non-FRL (4)	FRL school (5)	non-FRL school (6)
A. Black Students						
$\geq$ 1 Black T	0.096***	0.054	0.045	0.117	0.059*	0.018
	(0.033)	(0.040)	(0.030)	(0.081)	(0.031)	(0.057)
N (students)	2,112	1,908	3,319	659	3,308	756
R^2	0.033	0.039	0.034	0.081	0.056	0.080
$E(y)$	0.234	0.397	0.272	0.531	0.281	0.416
N (classrooms)	531	498	574	291	409	229
B. White Students						
$\geq$ 1 Black T	-0.057	0.003	-0.013	-0.037	0.016	-0.024
	(0.044)	(0.050)	(0.064)	(0.050)	(0.054)	(0.044)
N (Students)	3,778	3,348	2,708	4,307	2,169	4,966
R-squared	0.052	0.080	0.016	0.017	0.073	0.078
ymean	0.386	0.491	0.245	0.559	0.370	0.462
N (classrooms)	866	821	780	796	305	664

Notes: 2SLS estimates of the impact of ever having a Black teacher (Black T) in grades K-3, as described in equations (1) and (2), on the probability of ever enrolling in college. 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 free-lunch 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 classroom. Each column estimates this baseline model on a different subsample. *** p<0.01, ** p<0.05, * p<0.1.

Table 6: North Carolina Summary Statistics

	All Students		Persist. Disadv. Students		Persist. Disadv. Black Students	
	Black	White	Black	White	Male	Female
	(1)	(2)	(3)	(4)	(5)	(6)
A. Outcomes (in%)						
H.S. Dropout	13.00	9.97	14.30	23.87	17.84	11.10
H.S. Grad, No 4-Year Intent	46.60	47.87	52.23	60.82	53.61	50.99
H.S. Grad, 4-Year Intent	40.22	42.06	33.25	15.02	28.21	37.77
B. Exposure to Black Teachers in Grades 3-5						
Exposure to $\geq$ 1 Black T	43.8%	13.9%	45.6%	14.3%	45.3%	45.8%
0 Black Teachers	56.2%	86.1%	54.5%	85.7%	54.8%	54.2%
1 Black Teacher	29.7%	12.3%	30.2%	12.18%	29.8%	30.5%
2 Black Teachers	11.2%	1.5%	12.0%	1.9%	12.0%	12.0%
3 Black Teachers	2.9%	0.1%	3.3%	0.2%	3.4%	3.3%
% Cohort's Teachers Black	25.5%	7.7%	26.9%	7.6%	26.8%	27.0%
	(24.94)	(12.42)	(25.93)	(13.78)	(25.98)	(25.88)
Within-School SD	[9.23]	[5.42]	[9.50]	[5.58]	[9.46]	[9.54]
C. Student Characteristics						
Persistently Disadvantaged	45.44	11.96	100.00	100.00	100.00	100.00
Ever Economically Disadvantaged	85.79	37.90	100.00	100.00	100.00	100.00
Ever LEP	0.29	0.22	0.33	0.63	0.41	0.26
Ever Exceptional	10.75	10.82	11.78	16.14	15.88	8.08
Parent Ed: HS Dropout	10.99	5.73	14.53	20.40	14.69	14.39
Parent Ed: College Grad	12.20	35.11	4.02	3.56	4.10	3.96
Unique students	106,373	211,207	48,335	25,254	22,962	25,373

Notes: Standard deviations (SD) presented in parentheses. Economic disadvantage (ED) designated by free or reduced-price lunch use. Persistent disadvantaged indicated if designated ED for each of grades 3-8. Teacher composition variables capture students and teachers in grades 3-5. Sample includes 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 7: 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. Black Student Sample						
$\hat{\delta}$	-0.022**	-0.049***	0.003	0.011	-0.006	0.019
	(0.011)	(0.016)	(0.014)	(0.016)	(0.021)	(0.023)
N	105,155	51,321	53,834	103,693	50,458	53,235
B. 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
C. Pooled Sample of Black and White Students: Fully Interacted						
Share	-0.008	-0.003	-0.012	0.006	0.006	0.003
	(0.011)	(0.017)	(0.015)	(0.017)	(0.023)	(0.024)
Share × *Black*	-0.014	-0.046*	0.015	0.005	-0.012	0.017
	(0.015)	(0.024)	(0.019)	(0.023)	(0.031)	(0.033)
N	315,079	158,359	156,720	311,497	156,214	155,283
Chow Test Results, Pooled Models:						
F value	47.588	22.143	21.067	25.645	14.794	14.855
P value	0.000	0.000	0.000	0.000	0.000	0.000
D. Persistently Economically Disadvantaged Black Student Sample						
$\hat{\delta}$	-0.036**	-0.082***	0.009	0.069***	0.062**	0.066**
	(0.015)	(0.023)	(0.019)	(0.023)	(0.029)	(0.033)
N	47,900	22,747	25,153	47,164	22,323	24,841
E. Persistently Economically Disadvantaged White Student Sample						
$\hat{\delta}$	0.001	0.005	0.010	-0.025	-0.022	-0.063
	(0.045)	(0.076)	(0.063)	(0.037)	(0.050)	(0.062)
N	25,209	12,754	12,455	24,914	12,578	12,336
F. 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)
N	47,900	22,747	25,153	47,164	22,323	24,841
$H_0 : \delta_{Male} = \delta_{Female}$ (p value)	0.866	0.623	0.523	0.905	0.103	0.146
G. Persistently Disadvantaged Sample, by County Unemployment						
$\hat{\delta}_{LowUnemployment}$	-0.032*	-0.047	-0.014	0.062**	0.035	0.078*
	(0.019)	(0.031)	(0.026)	(0.030)	(0.038)	(0.045)
$\hat{\delta}_{HighUnemployment}$	-0.042*	-0.133***	0.045	0.080**	0.104**	0.048
	(0.023)	(0.036)	(0.028)	(0.034)	(0.046)	(0.050)
N	47,815	22,708	25,107	47,080	22,284	24,796
$H_0 : \delta_{Low} = \delta_{High}$ (p value)	0.757	0.068	0.122	0.682	0.247	0.660

Notes: Standard errors reported in parentheses. Baseline standard errors clustered by the level of treatment variation: school-cohort. Persistently disadvantaged refers to students designated as economically disadvantaged in each of grades 3-8. All models control for time-varying school characteristics and observed student socio-demographics. *** p<0.01, ** p<0.05, * p<0.1.

Table 8: 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 (≥ 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.

Appendix A Additional Results

Table A1: Sensitivity of Baseline STAR Estimates

Model:	Name and DOB (1)	K Only (2)	Class size (3)	Pct Black (4)	Baseline (5)	No Controls (6)
A. Black Students						
$\geq$ 1 Black T	0.054*	0.108**	0.057**	0.059**	0.059**	0.060**
	(0.031)	(0.049)	(0.027)	(0.027)	(0.027)	(0.027)
N (Students)	3,590	2,043	4,064	4,064	4,064	4,088
R^2	0.052	0.073	0.057	0.058	0.058	0.006
$E(y)$	0.35	0.339	0.313	0.313	0.313	0.313
N (classrooms)	629	206	638	638	638	640
B. White Students						
$\geq$ 1 Black T	-0.037	-0.069	-0.02	-0.019	-0.019	-0.016
	(0.036)	(0.049)	(0.035)	(0.035)	(0.035)	(0.034)
N (Students)	6,355	4,182	7,135	7,135	7,135	7,135
R^2	0.072	0.092	0.074	0.075	0.075	-0.001
$E(y)$	0.469	0.481	0.435	0.435	0.435	0.435
N (classrooms)	968	251	969	969	969	969
C. All Students						
$\geq$ 1 Black T	-0.037	-0.069	-0.02	-0.019	-0.019	-0.016
	(0.036)	(0.048)	(0.035)	(0.034)	(0.034)	(0.034)
$\geq$ 1 Black T $\times$ Black S	0.091*	0.177***	0.078*	0.078*	0.078*	0.075*
	(0.047)	(0.068)	(0.044)	(0.043)	(0.043)	(0.042)
N (Students)	9,987	6,235	11,245	11,245	11,245	11,269
R^2	0.184	0.147	0.162	0.162	0.162	0.102
$E(y)$	0.44	0.448	0.404	0.404	0.404	0.404
N (classrooms)	1261	323	1261	1261	1261	1263

Notes: 2SLS estimates of the impact of ever having a Black teacher (Black T) in grades K-3, as described in equations (1) and (2), on the probability of ever enrolling in college. All models condition on school-by-cohort fixed effects. Column 1 is restricted to the sample of students who name and date of birth are observed, column 2 is restricted to the kindergarten cohort, column 3 changes the class type dummies to the count of class size, and instruments for size with the type dummy, column 4 adds a control for the share of the class that is Black to the set of baseline controls, column 5 is the baseline specification, and column 6 omits student and teacher controls. Baseline controls include student controls for sex and free-lunch 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. The pooled models in panel C fully interact all covariates and school-by-year fixed effects with the Black student (Black S) indicator; a Chow (joint F) test of these interaction terms finds them to be strongly significant ($p < 0.001$) in all six models, suggesting that the education production function is systematically different for white and Black students in the STAR schools. We do not report the coefficient on the Black S variable because it is not directly interpretable, due to these interactions. *** p<0.01, ** p<0.05, * p<0.1.

Table A2: Sample Means by High School (HS) Completion Status

	All (1)	HS Observed (2)	HS Grad (3)	HS Not Grad (4)	HS Missing (5)
A. Black Students					
Male	0.525	0.449	0.393	0.564	0.570
FRL	0.819	0.746	0.693	0.855	0.863
Missing NSC Link	0.116	0.026	0.020	0.038	0.170
Low income school	0.814	0.779	0.738	0.863	0.835
Took SAT/ACT	0.269	0.529	0.723	0.129	0.113
College enrollment	0.324	0.566	0.714	0.259	0.181
Two-year enrollment	0.221	0.382	0.461	0.218	0.126
Four-year enrollment	0.200	0.363	0.498	0.085	0.103
Semesters attempted	2.556	4.761	6.413	1.352	1.242
Graduated	0.092	0.179	0.250	0.034	0.039
N	4,064	1,517	1,022	495	2,547
B. White Students					
Male	0.530	0.517	0.495	0.619	0.541
FRL	0.381	0.315	0.257	0.580	0.439
Missing NSC Link	0.109	0.039	0.037	0.048	0.172
Low income school	0.304	0.312	0.301	0.358	0.297
Took SAT/ACT	0.380	0.551	0.659	0.059	0.227
College enrollment	0.432	0.564	0.643	0.205	0.313
Two-year enrollment	0.299	0.391	0.441	0.164	0.217
Four-year enrollment	0.274	0.373	0.438	0.074	0.186
Semesters attempted	3.476	4.738	5.581	0.905	2.343
Graduated	0.193	0.271	0.325	0.023	0.123
N	7,135	3,377	2,768	609	3,758

Notes: HS Grad/Not Grad refers to a high school graduation record in the state of Tennessee. Students who graduated HS in other states could be counted in either column 4 or 5. FRL is free or reduced price lunch. NSC links are names and birth dates.

Table A3: Balance Test

	IV 1: Black T in first year				IV 2: Expected Black Ts in years 2-4			
	All	All	Male	Female	All	All	Male	Female
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A. Black Students								
Male	-0.008	0.001			-0.050**	0.002		
	(0.016)	(0.012)			(0.022)	(0.009)		
FRL	-0.019	-0.014	-0.042	0.016	0.102	-0.016	-0.023	-0.018
	(0.038)	(0.021)	(0.027)	(0.030)	(0.062)	(0.014)	(0.016)	(0.022)
Small class	-0.051	-0.021	-0.016	-0.025	0.260**	0.198***	0.208***	0.211***
	(0.058)	(0.058)	(0.059)	(0.066)	(0.103)	(0.053)	(0.055)	(0.057)
Missing NSC link	-0.031	-0.024	-0.035	-0.018	0.414***	-0.010	-0.015	0.006
	(0.040)	(0.021)	(0.029)	(0.029)	(0.051)	(0.017)	(0.024)	(0.022)
N	4,107	4,064	2,112	1,908	4,107	4,064	2,112	1,908
R^2	0.004	0.329	0.355	0.335	0.069	0.875	0.879	0.879
$E(y)$	0.433	0.437	0.437	0.447	0.926	0.935	0.919	0.972
B. White Students								
Male	0.003	0.004			-0.015*	-0.003		
	(0.005)	(0.003)			(0.008)	(0.003)		
FRL	-0.007	-0.006	-0.004	-0.012*	-0.032**	0.004	0.006	0.001
	(0.009)	(0.004)	(0.006)	(0.006)	(0.014)	(0.004)	(0.005)	(0.006)
Small class	-0.015	-0.009	-0.012	-0.006	-0.030	-0.030	-0.034*	-0.027
	(0.017)	(0.013)	(0.015)	(0.013)	(0.027)	(0.020)	(0.020)	(0.021)
Missing NSC link	0.013	-0.006	0.002	-0.013	0.087***	0.001	-0.005	0.011
	(0.016)	(0.006)	(0.008)	(0.009)	(0.025)	(0.007)	(0.008)	(0.011)
N	7,138	7,135	3,778	3,348	7,138	7,135	3,778	3,348
R^2	0.003	0.484	0.492	0.516	0.012	0.797	0.794	0.814
$E(y)$	0.0590	0.0587	0.0601	0.0568	0.159	0.158	0.152	0.165
Fixed Effects	No	Yes	Yes	Yes	No	Yes	Yes	Yes

Notes: Instrumental variable (IV) 1 is a binary indicator for having had a Black teacher (T) in the student's first year in STAR. IV 2 is the the expected number of Black teachers the student would have had, had they complied with random assignment and remained in that school for the remaining STAR years. FRL refers to free or reduced price lunch. Missing NSC link refers to missing the student's name or date of birth, which complicates the National Student Clearinghouse data merge. Fixed effects are at the school-by-cohort level, as in the main model. *** p<0.01, ** p<0.05, * p<0.1.

Table A4: Heterogeneous Effects of Black Teacher in First Year on Math Scores

	Male (1)	Female (2)	FRL (3)	Non-FRL (4)	FRL School (5)	Non-FRL School (6)
A. Black Students						
Black Teacher	2.989	8.381**	6.374**	-1.647	5.523*	-0.739
	(2.808)	(3.425)	(2.769)	(4.769)	(2.881)	(6.074)
N (Students)	1,907	1,720	3,008	599	2,976	690
R^2	0.688	0.627	0.654	0.638	0.649	0.643
$E(y)$	498.2	500.8	497.4	509.3	497.3	508.6
N (classrooms)	503	482	554	269	400	222
B. White Students						
Black Teacher	-7.398*	-7.471	-9.715**	-5.917	-13.934***	-2.920
	(4.108)	(4.599)	(4.773)	(4.032)	(4.554)	(4.176)
N (students)	3,387	2,999	2,357	3,972	1,936	4,467
R^2	0.616	0.590	0.648	0.567	0.553	0.605
$E(y)$	519	520.4	513.9	522.2	521.3	519.2
N (classrooms)	802	762	703	752	289	631

Notes: OLS estimates of equation (3) on end of grade math scores. All models condition on school-by-cohort fixed effects. Controls include student controls for sex and free-lunch (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 A5: North Carolina Balance Test Regressions

	Base	District-by-Year FE	Linear School Time Trends
	(1)	(2)	(3)
% Students Econ. Disadv.	-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)
% Black Gr. 3 Cohort Persist. Disadv.	-0.010	-0.012	-0.010
	(0.007)	(0.007)	(0.009)
% 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. Persistently disadvantaged refers to students designated as economically disadvantaged in each of grades 3-8. Each predictor entered in separate models.

Table A6: North Carolina Sensitivity Analyses

Outcome:	High School Dropout			College Intent		
Sample:	All	Male	Female	All	Male	Female
	(1)	(2)	(3)	(4)	(5)	(6)
A. Replicate Main Results: Persistently Disadvantaged Students						
$\hat{\delta}: BlackSample$	-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
$\hat{\delta}: WhiteSample$	0.001	0.005	0.010	-0.025	-0.022	-0.063
(by school-cohort)	(0.045)	(0.076)	(0.063)	(0.037)	(0.050)	(0.062)
(by school)	[0.050]	[0.090]	[0.071]	[0.042]	[0.057]	[0.073]
N	25,209	12,754	12,455	24,914	12,578	12,336
B. Drop "No-Variation Schools"						
$\hat{\delta}: BlackSample$	-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	41,475	19,609	21,866	40,841	19,246	21,595
$\hat{\delta}: WhiteSample$	0.011	0.021	0.011	-0.027	-0.023	-0.063
	(0.046)	(0.078)	(0.065)	(0.038)	(0.051)	(0.063)
N	11,919	5,998	5,921	11,767	5,906	5,861
C. Include School-Specific Linear Time Trends						
$\hat{\delta}: BlackSample$	-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
$\hat{\delta}: WhiteSample$	0.037	0.108	-0.040	0.032	0.070	-0.002
	(0.056)	(0.092)	(0.086)	(0.046)	(0.067)	(0.084)
N	25,209	12,754	12,455	24,914	12,578	12,336
D. FE Logit Coefficient Estimates						
$\hat{\delta}: BlackSample$	-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
$\hat{\delta}: WhiteSample$	0.024	0.064	0.007	-0.199	-0.071	-0.466
	(0.284)	(0.419)	(0.420)	(0.346)	(0.580)	(0.466)
N	24,372	11,770	11,322	23,311	9,918	10,961

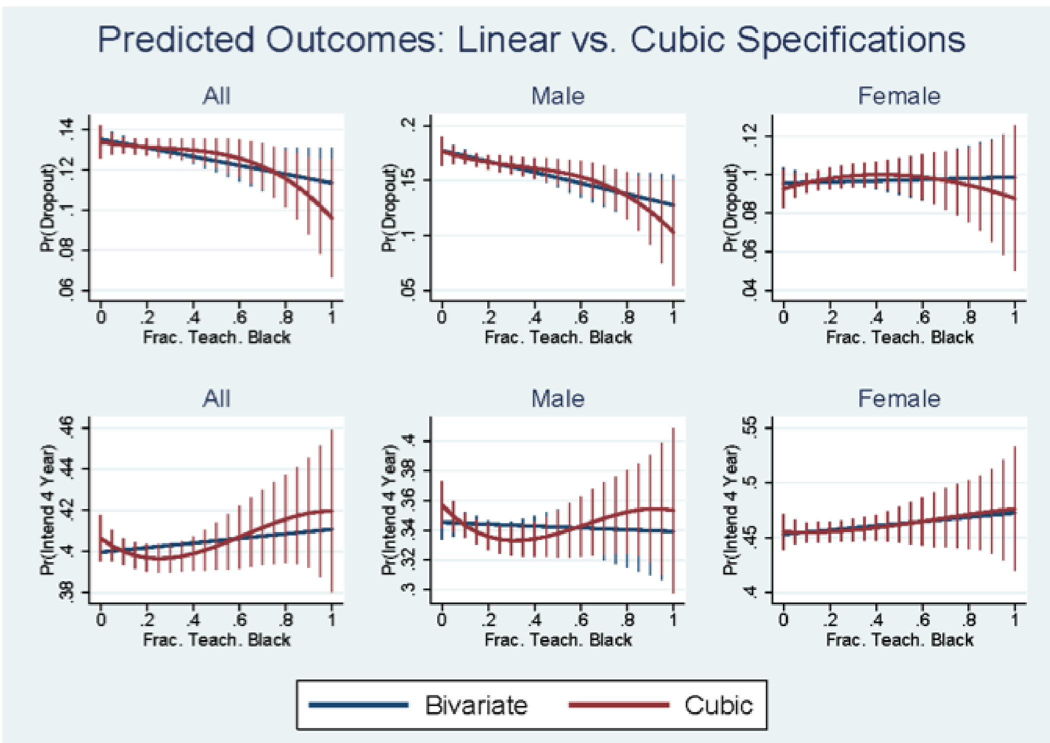
Notes: Standard errors reported in parentheses. Baseline standard errors in Panels A, B and C clustered by school-cohort. In Panel D, errors are unclustered. Persistently disadvantaged refers to students designated as economically disadvantaged in each of grades 3-8. All models control for time-varying school characteristics and observed student sociodemographics. No variation schools include those with always-100% or always-0% Black teaching staffs. *** p<0.01, ** p<0.05, * p<0.1.

Table A7: 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*** (0.050)	-0.037*** (0.011)	0.059*** (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 8. 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 4 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).