Role Models Revisited: HBCUs, Same-Race Teacher Effects, and Black Student Achievement*

Lavar Edmonds
Stanford University

This Draft: August 15, 2022

Abstract

This paper presents the first analysis of teacher effects for Historically Black College and University (HBCU) graduates. Using multiple estimators that leverage within-student variation in teacher assignment in North Carolina elementary schools, I find Black students score higher on end-of-grade math exams when assigned to an HBCU-trained teacher. Both Black and White HBCU-trained teachers are more effective with Black students than their same-race, non-HBCU peers are. Suggestive evidence indicates students with HBCU-trained teachers benefit from lower suspension rates, particularly Black boys. Effects are unexplained by differences in observable teacher characteristics; I argue they are at least partly the result of differential teacher education practices between colleges.

*Contact: ledmonds@stanford.edu. Thanks to the North Carolina Education Research Data Center for collecting and providing data. Special thanks to Jennifer Jennings for both her research guidance and her generous purchase of the NCERDC data. Likewise, I am grateful to Leah Boustan, Jesse Bruhn, Janet Carlson, Tom Dee, Ellora Derenoncourt, Matthew Desmond, Peter Hinrichs, Isaac Sorkin, Jon Valant, Anne Kat Alexander, Joe Fish, Jennifer Kowalski, Lillian Leung, Renee Louis, João Pugliese, and China Mae Stepter for insightful comments and suggestions on previous proposals and drafts of the paper. I also thank seminar participants from the Princeton Education Research Section, the Eviction Lab, and the Economics of Racism for their helpful feedback. All errors – and there will be some – are all my own.

1 Introduction

Racial representation and concordance studies are widespread across the social sciences, detailing same-race match effects in multiple settings: mentors and impacts on occupational choice and job satisfaction (Grissom and Keiser, 2011; Kofoed and mcGovney, 2019), judges and legal proceedings (Steffensmeier and Britt, 2001; Abrams, Bertrand, and Mullainathan, 2012), police-civilian interactions and criminal justice reform (Headley and Wright, 2020; Ba et al., 2021), physician-patient matches and health outcomes (Alan, Garrick, and Gariziani, 2019; Greenwood et al., 2020), and even basketball referees and game penalties (Price and Wolfers, 2010). Even papers without an explicit focus on racial match have relied on assumptions about same-race pairings and presumed behavior between people (e.g., Arnold, Dobbie, and Yang, 2018; Rose, 2021). Education, too, has delved into this research paradigm, with several papers posing that students benefit from exposure to demographically-congruent teachers (Bartanen and Grissom, 2021; Dee, 2004; Dee, 2005; Egalite, Kisida, and Winters, 2015; Gershenson, Holt, and Papageorge, 2016; Hart, 2020; Holt and Gershenson, 2019; Joshi, Doan, and Springer, 2018; Penney, 2017), a match that may be especially beneficial for racial and ethnic minority students (Villegas and Lucas, 2004).

Many studies have proposed explanations for these positive same-race teacher effects (e.g., role model effects, stereotype threat, and teacher cultural fluency in the classroom), yet they ultimately provide little strong evidence for these mechanisms. In their attempts to unpack these student-teacher dynamics, one feature largely disconnected from common explanations is how teacher effectiveness in same-race matches may be driven by training and background. Building on this, in this paper I revisit same-race teacher effects and their mechanisms via an analysis of Historically Black Colleges and Universities (HBCUs). Specifically, I address the question: what is the impact of having an HBCU-trained teacher on Black student outcomes? HBCUs are a fitting context for studying same-race teacher effects for Black students given their historical and contemporary significance in the education of Black teachers, with approximately 50% of the current supply of Black teachers having attended an HBCU (National Association for Equal Opportunity in Higher Education, 2008).

I study HBCU attendance and teacher quality by estimating teacher effects on end-of-grade math exams using nine years of administrative panel data from elementary schools in North Carolina (2009-2010 to 2017-2018). One empirical concern is that stronger or weaker students might be assigned to classrooms with teachers from

HBCUs. To address this possible source of endogeneity, I employ a two-way fixed effects (TWFE) regression estimator that uses within-student variation in teacher assignment over time to identify differences in average student achievement between years when students are and are not assigned to HBCU-trained teachers.

Correspondingly, I observe several important dynamics. First, I find HBCU-trained teachers have a positive effect on Black student achievement. Different estimators and specifications indicate that, on average, Black students score 0.032-0.052 standard deviations higher on standardized mathematics exams in the years they are assigned to an HBCU-trained teacher. This effect size would represent 5-8% of the estimated math test score gap between Black and White students. Likewise, effect sizes are comparable to, if not larger than, estimates reported in previous papers on same-race teacher matches. Results are robust to alternative specifications that address confounding from teacher experience and biases that may arise from TWFE regression estimators with differential timing in teacher assignment. I show a smaller, yet substantive, result holds for HBCU teacher effects across all students, though the estimate is largely driven by the positive effect for Black students. Importantly, I do not observe a significant negative effect for White students or Hispanic students.

Second, contrary to some prior studies, I find little evidence that Black students in classrooms with Black teachers generally score better on math exams, with the notable exception of the positive effect on test scores for Black teachers who attended an HBCU. Instead, I find HBCU-driven variation in teacher quality within teacher race: Black students score higher when paired with Black HBCU-trained teachers, compared against Black teachers from non-HBCUs. A similar result holds for Black students matched with HBCU-trained White teachers (compared with White teachers from other colleges), bolstering the possibility of race-independent teacher characteristics that improve student outcomes and raising questions about the interpretation of same-race teacher match effects. In particular, I argue that HBCU-trained teacher effectiveness with Black students, rather than solely a function of race, is at least partly the result of these colleges' historical and contemporary commitment to educating teachers in culturally-fluent pedagogical practices that are conducive to Black student academic success.

Third, to probe distinctive practices that may guide their classroom effectiveness, I supplement my focus on test scores by examining, across students, the association between HBCU-trained teacher assignment and the probability that a Black student receives a suspension judgment. Linear probability model (LPM) estimates predict

that Black students assigned to an HBCU-trained teacher have a modestly lower probability of receiving a suspension judgment than their Black peers in other classrooms. Notably, the association appears driven by differential suspension receipts between boys and girls. Both Black and White teachers from HBCUs appear to levy suspensions on Black students less than their same-race counterparts from non-HBCUs do, a correlation entirely driven by differences in suspensions for Black male students. Combined with empirical work showing the deleterious effects of suspensions and the disproportionate impacts on Black male students, these results imply Black boys may especially benefit from assignment to an HBCU-trained teacher.

Finally, in the absence of preservice teacher data to account for differential sorting into HBCUs, I highlight that commonly proposed predictors of teacher ability — teacher certification exams and college-level average SAT scores — are insufficient explanations for HBCU-trained teacher effectiveness with their Black students.

Principally, this paper advances literature on teacher preparation, higher education practices, and teacher value-added. It provides a novel analysis of graduate outcomes for one of the longest standing, historically significant institutions in the United States higher education system and extends the literatures on both same-race teacher effects and racial/ethnic minority teacher recruitment and retention. In doing so, it offers new insights for addressing persistent racial academic achievement gaps driven by resource inequities, including the unequal distribution of high-quality teachers. This work also lays the foundation for revisiting what constitutes an effective teacher and the extent to which our current practices for training and evaluating teachers (e.g., teacher certification exams) optimally prepare and capture teacher effectiveness. Beyond education, this research has implications for how we understand same-race pairings, towards an explicit acknowledgement that a confluence of factors shapes our interactions, including our visible features (e.g., race, gender) as well as our own backgrounds. In this setting, and likely many others, matching individuals on race was not necessarily the defining feature of positive effects. Rather, findings suggest a combination of race, disposition (selection into HBCUs), and training (attendance and teacher education at HBCUs) contribute to teacher success with Black students.

This paper is structured as follows: in Section 2, I briefly discuss the history and relevance of Historically Black Colleges and Universities, especially as they pertain to Black teachers, and how these colleges are well-suited for extending the same-race match results found in previous studies. Section 3 outlines the data used, and Section

4 details the baseline econometric model for the analysis. Section 5 presents results on the effects of HBCU-trained teachers on Black student achievement. I show results are robust to alternative specifications and estimators, and I build on these findings to identify variation in Black student achievement conditional on both the teacher's HBCU attendance and race. I address possible mechanisms for these results in Section 6 before concluding and discussing policy implications in Section 7.

2 Historically Black Colleges and Universities in the United States

Historically Black Colleges and Universities have been staples of Black communities and avenues for Black student access to postsecondary education for nearly two centuries (Harper, 2017). Largely clustered in the U.S. South, but spanning states from Pennsylvania to Florida, there are roughly 100 Historically Black Colleges and Universities operating today in 19 states, D.C., and the U.S. Virgin Islands. These colleges have traditionally served majority Black student populations, though HBCUs specifically refer to colleges founded before 1964 in response to the de jure and de facto exclusion of Black people from higher education. As such, each was founded with the broad mission of educating Black Americans (Roebuck and Murty, 1993, pg. 3).

Transformative institutions for Black college graduates across all fields, HBCUs historically have been especially crucial to the supply of Black teachers in the United States. Many HBCUs were founded, or at least existed at some point, as "normal schools," or institutions designated for preparing teachers for the classroom (Akbar and Sims, 2008). Throughout much of the 20th century, jobs in education were a cornerstone of the Black middle class (Thompson, 2021), driven partially by "separate-but-equal" laws that segregated students by race and propagated a labor market specifically for Black teachers needed for instruction in the all-Black K-12 schools. Remarkably, HBCUs' contributions to the Black teacher supply continued even after the Civil Rights Act of 1964's expansion of collegiate opportunities for Black students at non-HBCUs. Based on 2019-2020 degree conferral data from the Integrated Postsecondary Education Data System (IPEDS), despite HBCUs producing only about 8% of Black college graduates, they graduated just over 20% of the Black college graduates with education degrees, suggesting HBCUs remain a pivotal destination

for Black students interested in teaching.[1] By no means was this year an aberration: previous estimates report HBCUs have produced roughly 50% of the current supply of Black teachers in the U.S. (National Association for Equal Opportunity in Higher Education, 2008). As I will show later in the paper, HBCU-trained teachers in the sample are far more likely to teach in schools with more Black students, indicating their greater potential for impacting racialized inequities through their effects on Black students.

Despite being historical and contemporary powerhouses for Black teacher training, HBCUs have largely been unexplored quantitatively, especially in economics. To date, extant economic research on HBCUs has focused generally on questions related to finance and analyses of college funding (Sav, 2010), accountability, and the long-run returns to their students, defined by employment outcomes (Price, Spriggs, and Swinton, 2011), pecuniary measures like long-run wages (Mykerezi and Mills, 2008), and collegiate outcomes (e.g., likelihood of graduation) (Wilson, 2007). However, seemingly none have published research on HBCUs and teacher preparation or how it relates to the effectiveness of their graduates. This analysis aims to bridge this gap, expanding the body of research on teacher preparation programs and teacher value-added with the first study of HBCU-trained teacher effects on Black student achievement. I am not the first to propose teacher preparation and experiences in preservice education can matter for teacher instructional practices, especially with racial/ethnic minority students (e.g., Kumar and Lauermann, 2018), or that quality and type of teacher preparation may impact student achievement (e.g., Goldhaber, Liddle, and Theobald, 2013; Koedel et al., 2015). However, this would be the first to explicitly study HBCUs and how preservice teacher education at these institutions may directly affect student academic performance.

2.1 Theoretical underpinnings for HBCUs and the link to Black student achievement

Historically Black Colleges and Universities are an ideal setting for studying teacher effects, particularly for research focused on Black student achievement. I pose HBCUs and their approach to teacher education here as an alternative mechanism to those previously discussed and suggested in the growing literature on same-race teacher matches.

Education researchers across the social sciences, notably in the past two decades,

[1] Based on the author's calculations.

have established and furthered results showing gains in achievement for students in classrooms with demographically-congruent teachers. As research on these same-race teacher effects expands, so, too, do questions about the underlying causes. The field has yet to coalesce around one particular driver of the racial match effect results, though researchers have proposed several. One such explanation is role model effects. Specific to student achievement, the success of Black teachers with Black students, the theory holds, is a product of Black students revising previously held beliefs about their own ability. Their observation of educated, successful Black adults in the classroom adjusts what they believe is possible and their belief on returns to their own academic achievement. This hypothesis is neither new nor unique to economics[2], and I argue it is ultimately an insufficient explanation. It implies that, on average, Black students lack access to conventionally successful Black adults beyond the classroom, and that adults in the household are not encouraging of the students' success. This is a tenuous assumption, harkening back to a flawed "culture of poverty" ideology (Dudley-Marling and Lucas, 2009). Concurrently, it requires a deficit mindset view of Black students, one that students would have internalized. Alas, this misaligns with empirical research detailing how Black student beliefs in the transformative power of education are "shaped by messages that grandparents, parents, and generations of other family members consistently conveyed to them, their siblings, and their cousins" (Harper and Davis, 2012, pg. 117) and that Black students tend to hold stronger beliefs about the value of schooling relative to their White peers (Harris, 2006; Harris, 2011). It also potentially undersells teacher agency in the student-teacher relationship, inadequately capturing the extent to which teachers may not actively (or even want to) view themselves as role models (Maylor, 2009).

Second, researchers have also invoked stereotype threat in explaining same-race teacher effects for racial/ethnic minority students, drawing from Claude Steele and Joshua Aaronson's foundational work in social psychology (Steele and Aaronson, 1995; Steele, 1997). They argue achievement is hindered when a student's identification with schooling and success in school are disconnected, an event likely when students find themselves in settings that may reinforce negative stereotypes about their group identification (e.g., that Black students have inferior intellectual ability). Teacher expectations matter for student outcomes (Aaronson, 2002; Ferguson, 2003; Papageorge, Gershenson, and Kang, 2020), and prior work has shown Black teachers

[2]See Villegas and Lucas (2004) for a brief discussion on the history and argument for the role model effects hypothesis.

typically have higher expectations for Black students (Gershenson, Holt, and Papageorge, 2016). Same-race teachers, then, may mitigate stereotype threat by alleviating the student's anxiety of having their performance be an indicator of and benchmark for their racial group's perceived acumen. It is unclear, however, how prevalent and salient stereotype threat is, as studies specifically focused on Black elementary school students have argued stereotype threat bears little weight on test performance with students who are unaware of the negative stereotypes about their group (Wasserberg, 2014a; Shelvin, Rivadeneyra, and Zimmerman, 2014).

Third, rather than focus on passive effects like the teacher's demographic characteristics, one hypothesis for same-race match effects centers a teacher's active behavior in the classroom. In particular, teachers may draw from practices that are culturally fluent, relevant, and sustaining (Ladson-Billings, 1995, 2014; Paris and Alim, 2014). These teachers embrace students' diversity and structure content around them in ways that "affirm culture(s) and history" and "link academic tasks to children and families' daily experiences and cultural life(s)." (Acosta, Foster, and Houchen, 2018, pg. 343). Some research has considered this framework in explaining the effectiveness of Black teachers with Black students (e.g., Gershenson et al., 2022), though I contend it is a weaker claim in this broad a setting, overly focusing on teacher race without enough consideration for characteristics that may simply be correlated with race.

Taken together, while I leave open the possibility that these theories suffice in some settings, I maintain they are each insufficient in understanding particularly same-race teacher effects, overly reliant on race alone as the defining feature. Being Black does not automatically endow one with the ability to teach from a sociocultural perspective. Even if a prospective teacher possesses the cultural experiences from which to draw for effective instruction, there is nothing innate about race that guarantees one has the capacity and language to do so. I maintain that, instead, there need be some impetus, like formal teacher education, for thinking about and structuring instruction in this way.[3] As such, a more suitable framework for understanding same-race teacher effects would account for diversity in teacher background and perspective, the guiding principle of this research.

The clearest link connecting historically Black college teacher preparation to Black

[3]For understanding Black teacher connections with Black students in a historical context, we cannot overstate the importance of informal interactions outside of school via these teachers engaging with the local Black community, particularly in the pre-Brown vs. the Board of Education (1954) era. See Ladson-Billings and Anderson (2021) and Smith (1982) for further discussion.

student achievement is their heritage in 20th century Black educator pedagogy theory and practice. They are the last remaining collegiate institutions from a period where Black teachers, principals, and colleges worked explicitly in concert to debate educational philosophy and develop curriculum motivated by and geared directly towards Black students. As detailed in Acosta, Foster, and Houchen (2018) and Givens (2021), the history of Black pedagogical praxis and expertise was not mere happenstance. For the better part of two centuries, but especially following 1930s Jim Crow-era laws, these practices flourished as a purposeful endeavor cultivated and disseminated by a rich network of national, state, and local associations of Black teachers and principals via conferences and publications. Moreover, K-12 schools' connections to HBCUs facilitated a pipeline for recruiting promising Black high school students into teaching, who were trained at these colleges and in turn able to use those relationships to obtain employment in K-12 schools following graduation.

Consequently, the case for HBCU-trained teacher success with Black students may stem from the pedagogical approaches centered during their own preparation, influencing the culturally-fluent classroom practices in their instruction. Building on decades of qualitative work, scholars in recent years have established additional, quantitative support for this dynamic. Closely related to this tradition, Dee and Penner (2017) provide credible causal evidence of the positive effect a high-fidelity curriculum centered around ethnic studies can have on student outcomes, especially for academically-struggling students. In another study, Dee and Penner (2021) explore the impacts of Oakland, California's African American Male Achievement (AAMA) program, an initiative that linked Black male high school students to Black male teachers in classrooms with a curriculum comprising "social-emotional development, African-American history, and culturally relevant pedagogy" (Dee and Penner, 2021, pg. 2). Their quasi-experimental analysis reveals the AAMA led to significant reductions in Black male student dropouts, with smaller but positive reductions on Black female student dropouts as well. In follow-up work on the long-run effects of ethnic studies courses, Bonilla, Dee, and Penner (2021) show students assigned to these courses experienced increased graduation rates, attendance, and likelihoods of enrolling in college. In short, the quantitative research on culturally-fluent education and classroom practices emphatically corroborates its positive impacts on Black student outcomes.

While the argument for culturally-fluent classroom practices has been invoked for Black teachers writ large, it is more appropriate in the HBCU context given their

historical approach to teacher education. For example, notice in Dee and Penner (2021) that, while the instructors are all Black men, they were not drawn at random from the distribution of Black male teachers, but rather were "carefully selected and trained" (Dee and Penner, 2021, pg. 5). How teachers are prepared for the classroom is a nontrivial input in determining their effectiveness. Prior scholarship (e.g., Watkins, 2005; Arroyo and Gasman, 2014) has argued the infusion of culturally-fluent instructional preparation into teaching practices at HBCUs is central to their model of teacher education. Thus, this sociocultural component is more likely a direct feature of education for HBCU graduates than it is for those from other institutions and a more plausible explanation for teacher success with Black students.

3 Data

I employ administrative data for 3rd, 4th, and 5th grade students and teachers in public schools hosted by the North Carolina Education Research Data Center (NCERDC). Students and teachers in the data set are assigned unique IDs that allow researchers to track them across multiple years and multiple data sets. These data contain a host of student-level demographic, socioeconomic, and academic characteristics, such as race, gender, economic disadvantage status, and academic (test scores) and social-behavioral (suspensions, daily attendance) outcomes, as well as information about the classrooms they were in (e.g., courses for English language learners) and the schools they attend (e.g., demographic counts). For teachers, data include demographic characteristics (race, gender, and age), in addition to some preservice history (universities attended, level of education, certification test scores) and years of experience teaching in North Carolina.

As in some prior studies using North Carolina education data (e.g., Jackson, 2018; Hill and Jones, 2021), I use a string-searching algorithm over a course membership file to link students to teachers who taught specific courses and subjects during the year. This method improves fidelity of student-teacher matches over the base approach in early years of the data, which matches students to the teacher who administered the exam. It also comes with tradeoffs, namely that the algorithm is less effective at matching students to teachers at the middle school level where students have more courses and teachers, and specific math teachers are less easily identified. Consequently, this research focuses only on 3rd through 5th grade students.[4]

[4]See the online data appendix for additional details on the matching procedure.

I begin with a data set of 1,336,509 unique students in traditional public schools (e.g., no charter or laboratory schools) with non-missing student IDs in the 2009-2010 through 2017-2018 academic years. Of this group, 1,296,945 students were identified in a math class at some point during their time in the data, and 1,295,145 of those were attached to non-missing teacher ID numbers. Because identification relies on variation in teacher assignment over time, I necessarily focus on students in the data who I observe for at least two school years. This reduces the sample to 1,008,012 unique students, retaining 75% of the students in the original data set. From here, I make several additional restrictions to more closely approximate real-world circumstances. I exclude students without math test scores (0.08% of unique students) and students who attended multiple schools within a year (0.17% of remaining unique students). Similarly, because I am interested in the effect of unique teachers, I remove instances when a student appears in one school in a school year but has more than one math teacher and cannot be matched to a single, unique teacher. This, however, does not ensure that a student appears in one classroom, as I observe cases where a student has one teacher in a school year, but is listed in more than one classroom. Because my modeling strategy in part relies on classroom-level covariates, I use a matching algorithm to assign each student to one classroom in a given year. As in other work on teacher value-added (e.g., Chetty, Friedman, and Rockoff, 2014), I also remove classrooms with fewer than 10 or greater than 50 students (1.9% of unique students). Following these procedures, the data set contains 994,900 unique students. I make similar reductions to the teacher sample. From the initial 36,969 unique teachers, I exclude observations missing pertinent information on educational history (2.2%) and teacher demographics (3.3%). Further, to isolate the effect of attendance at a single historically Black college, I remove teachers if they attended multiple undergraduate institutions and at least once of them was an HBCU (0.1% of the 36,969 teachers). This leaves 34,115 unique teachers in the sample. Any additional exclusions follow from listwise row deletion when there is missingness on variables used in regression models. Crucially, as I show in Table A1, the main result for HBCU-trained teacher effectiveness from my preferred specification is not overly sensitive to these data restrictions. Separate regressions of various indicators on HBCU-trained teacher assignment reveal no relationship exists between teacher assignment and (1) whether a student has a missing test score, (2) missingness on control variables that would drop the student from the sample, or (3) whether an observation is excluded from the sample via the restrictions described above.

3.1 Descriptive statistics

For an overview of the data, I first turn to descriptive statistics for students and teachers in the sample. Panel A of Table 1 shows averages on observable characteristics for HBCU-trained and non-HBCU-trained teachers. HBCU-trained teachers comprise a relatively small share of all teachers (approximately 9%). As expected, these two groups differ considerably along race/ethnicity, with mostly Black HBCU-trained teachers and mostly White non-HBCU-trained teachers (though note almost one-fourth of HBCU-trained teachers are White). HBCU-trained teachers are far more likely to teach in schools with a greater proportion of Black students. Curiously, non-HBCU-trained teachers in this sample are considerably more likely than their HBCU-trained peers to have attended a college outside of North Carolina: nearly one-third of the non-HBCU-trained teachers attended college out of state, compared with just over 10% of HBCU-trained teachers. This difference could be important if, for instance, students at HBCUs are systematically more likely to receive training that directly correlates with North Carolina's curricular standards for student end-of-grade testing, resulting in student achievement results that are a function not necessarily of HBCUs or the teachers who attend them, but rather of the college's geographic location. Lastly, I note a large gap in the average teacher certification test score for HBCU-trained and non-HBCU-trained teachers. Standardized to mean 0 and standard deviation 1 within test-year, non-HBCU-trained teachers in the sample typically score almost three-quarters of a standard deviation higher on these preservice certification exams than their HBCU-trained counterparts do.

Because of this paper's focus on Black student test scores, I also compare Black and non-Black student aggregates to consider any systematic differences between the two groups (Panel B of Table 1). Black students, who comprise 23.1% of all students in the sample, differ from their non-Black peers on several measures, including socioeconomic background, academic performance and classifications, and disciplinary outcomes. Black students are more likely to have had an HBCU-trained teacher, though this is still a fairly rare occurrence: 21.7% of Black students in the sample have at least one teacher from an HBCU; this is the case for only 10.2% of other students.

Figure 1 shows the spatial distribution of HBCU-trained teachers across the 115 North Carolina public school districts in 2016, with the size of the circles corresponding to the number of (Black or White) HBCU-trained teachers working in a district

in that year.[5] For reference, I label each of the HBCUs in the state and their corresponding cities. Indeed, although there are 11 operating HBCUs in North Carolina, all of which are represented in the data set, 78% of the HBCU-trained teachers I study attended one of five universities: Fayetteville State University, North Carolina Central University, North Carolina AT, Winston-Salem State University, and Elizabeth City State University.[6] Given the relatively few HBCU-trained teachers in the sample, there are several school districts (16%) without an HBCU-trained teacher in any year.[7] Still, these teachers are generally well-spread across the state, tending to cluster in districts containing or located near an HBCU.

4 Empirical Strategy

4.1 Modeling student achievement

I derive empirical estimates from a linear and additively separable education production function, the workhorse framework for modeling cumulative student achievement in the economics of education literature (Todd and Wolpin, 2003). Specifically, I employ a two-way fixed effects estimator (TWFE) that relies on within-student variation in teacher assignment over their time in the sample. As with most analyses of teacher effectiveness, a potential source of endogeneity stems from the non-random assignment of students to teachers (Paufler and Amrein-Beardsley, 2014). Teacher effectiveness is obfuscated if these teachers are assigned higher scoring students or, more broadly, students who substantially differ from those in an ostensibly less-effective teacher's classroom. While a within-student estimator does not completely alleviate this concern, empirical claims are stronger when tracking the same student in different classrooms over time, eliminating time-invariant student characteristics that may otherwise bias teacher effectiveness estimates.

$$A_{ijst} = \beta_1(HBCU_{jt}) + X_{it}^{'}\gamma + W_{j}^{'}\delta + Z_{st}^{'}\eta + \theta_i + \lambda_t + \varepsilon_{ijst} \quad (1)$$

[5]For an accurate representation of distribution, I focus on the 2015-2016 academic year, the most recent year with teachers from all three grades.

[6]A variation on the main regression specification that interacts HBCU attendance and college location shows out-of-state HBCU-trained teachers outperform out-of-state, non-HBCU trained teachers, suggesting results are not solely driven by these five large, in-state education programs (0.047; SE 0.015).

[7]Since these districts tend to cluster in the largest cities in the state, they contain the vast majority of students: 90% of all students (and 96% of all Black students). Further, the main result is robust to including only students in these districts (0.034; SE 0.006; N = 339,240).

The corresponding regression equation is presented in Equation (1). The dependent variable is the score on an end-of-grade math exam[8], standardized within grade and year to mean 0 and standard deviation 1, for a student i with teacher j in school s and year t. I regress this test score on a vector X of time-varying student characteristics, a vector W of teacher covariates and a vector Z of time-varying school characteristics, as well as student (θ_i) and year (λ_t) fixed effects. The coefficient of interest is β, which captures the average change in a student's test score in the year when they are assigned an HBCU-trained teacher, relative to years when they were assigned a non-HBCU-trained teacher. Because the data are stacked by student-years, with individual students appearing multiple times, and the level of treatment (teacher assignment) varies with and across multiple students, I follow guidance from Cameron, Gelbach, and Miller (2011) on two-way clustering and cluster standard errors at the student and teacher levels.

In prior studies, typically in a Difference-in-Differences (DD) framework that uses a two-way fixed effects estimator under a conditional parallel trends assumption, researchers have interpreted β as the causal average treatment effect on the treated (ATT) (Bertrand, Duflo, and Mullainathan, 2004). However, recent applied econometrics literature (Borusyak and Jaravel, 2018; Callaway and Sant'Anna, 2020; de Chaisemartin and D'Haultfœuille, 2020a; Goodman-Bacon, 2021) has highlighted challenges to this interpretation that potentially bias the TWFE estimator from the ATT. Conceivably, many of the issues raised may appear in this application – most prominently, differential treatment timing: not all students receive an HBCU-trained teacher in the same year. This would be the case, as an example, for two students who attend elementary school in grades 3, 4, and 5 between 2010 and 2012, where Student A receives an HBCU-trained teacher in the 3rd grade, but not in the two subsequent years, while Student B is assigned an HBCU-trained teacher in the 4th grade. As Goodman-Bacon (2021) elucidates, the TWFE regression mechanically computes estimates as a weighted average of all average treatment effects across all groups and times, where weights are a function of group size (units in the same treatment or control group in a time period) and treatment variance. The comparison between treated and already-treated units in the event of differential treatment timing may lead to negative weights, which could bias TWFE estimates away from the true ATT. A similar problem occurs with the possibility of treatment turning on and

[8]For students who retake a test, I use their median test score and create an indicator for whether a student repeats a grade and whether they retake an exam.

off as students move between "treatment" and "control" groups (i.e., students in the sample who have an HBCU-trained teacher in the 4th grade, but not in the 3rd or 5th). Potentially this is substantively different than having no HBCU-trained teacher in 3rd, 4th, or 5th grade, even though they would both be in the "control" group in the 3rd and 5th grades.

In response to these concerns, I follow the recommendation of de Chaisemartin and D'Haultfœuille (2020a) and calculate the weights associated with TWFE regressions that would bias estimated coefficients using the twowayfeweights package in Stata. Additionally, I consider two alternative specifications to my preferred regression model to ensure robustness of the main results. First, to ameliorate the negative weighting problem with differential treatment timing, I restrict the sample to students who differ on teacher assignment only in the 5th grade. That is, the "treatment" group comprises Black students who are assigned an HBCU-trained teacher exactly once (in the 5th grade), and the "control" group to which they are compared includes Black students who never have an HBCU-trained teacher. Second, I move from the TWFE approach entirely and compute the estimator proposed in de Chaisemartin and D'Haultfœuille (2020b) that, by construction, is robust to heterogeneous treatment effects, differential timing in treatment, and dynamic effects.

5 Results

I estimate the baseline empirical model using ordinary least squares, regressing math student test scores on an indicator for whether a teacher attended an HBCU and a set of student-, teacher- and school-level controls, as well as student and year fixed effects. Regression results appear in Table 2, suppressed to the HBCU attendance variable. Estimates vary from column to column based on the inclusion of additional covariates. Column 1 shows the relationship between having an HBCU-trained teacher and math scores before controlling for additional teacher-, classroom-, and student-level controls. Given test scores are in standard deviation units, this coefficient reflects that a Black student assigned to an HBCU-trained teacher, on average, scores approximately 0.034 standard deviations higher on their end-of-grade math exam than they score in a different year when they have a non-HBCU-trained teacher. Results are reasonably comparable to estimates in prior literature: the coefficient magnitude is larger than the same-race teacher match effects on test scores found previously in North Carolina (0.020 SD; Clotfelter, Ladd, and Vigdor, 2007) and the Black teacher-

Black student match effects reported in Missouri and Tennessee (0.021 SD; Bartanen and Grissom, 2021) and in Florida (0.030 SD; Egalite et al., 2015), almost exactly the Black teacher-Black student effect size reported in another North Carolina study (0.035 SD; Goldhaber and Hansen, 2010), and around half the size reported in Texas (0.104; Hanushek et al., 2005). Observe that because the standard deviation of test scores for Black students in the sample is smaller than one (0.874), a 0.034 standard deviation increase in test scores has an even larger impact for this group. By another metric, the often-discussed racial test score gap, the coefficient is approximately 5% of the math test score gap between Black and White students estimated in prior studies.[9]

Column 2 extends the baseline controls to include covariates for teachers, such as their level of experience[10], graduate school training, teacher licensure test scores[11], the number of classes the teacher taught in a given school year, whether the teacher attended a non-North Carolina college, whether that college was located within the school district in which they teach, and whether that college was historically a normal school. My main result is invariant to these controls, indicating the estimated HBCU teacher effect on student outcomes is not driven by HBCU-trained teachers' higher average teaching experience in my sample, any institutional familiarity with North Carolina education curricula that may have arisen from attending a college in-state or in the district where they teach, or any general comparative advantage from many HBCUs' histories specializing in teacher education. As a secondary point, I find a similar result to previous work indicating teachers with higher teacher certification scores modestly predict higher student test scores (Clotfelter, Ladd, and Vigdor, 2010; Goldhaber, Gratz, and Theobald, 2017). At odds with this, however, is the substantive, positive effect of HBCU-trained teachers, a group with markedly lower average licensure exam scores than their peers. I discuss policy implications of this dissonance in Section 5 later in the paper.

[9]Hill and Jones (2021) report a 0.63 standard deviation unconditional gap between Black and White students among North Carolina high school students, and a 0.70-0.75 standard deviation gap in 2015 NAEP scores. Reardon et al. (2019) estimate an average gap between Black and White students test scores of 0.66 SD across school districts nationally. Specific to this context, from Table 1, the mean group difference in math test scores between Black and non-Black students in the sample is 0.61 standard deviations. Though I do not report it explicitly in the table, the analogous comparison between Black and White students is 0.72 standard deviations.

[10]This specification includes both experience and a demeaned quadratic term, to capture possible nonlinearities in the returns to teacher experience.

[11]As with student test scores, I use the median test score for teachers who take multiple National Teacher Examination and/or Praxis exams, and I standardize all test scores within the year the test was taken to mean 0 and standard deviation 1.

Moving to column 3, I introduce several classroom-level aggregates of student race, gender, and economic disadvantage status, and the class size, to proxy for important peer effects that education researchers have long established are paramount to student academic achievement and behavior (Hanushek et al., 2001; Sacerdote, 2001). As with the inclusion of teacher controls in column 2, the coefficient on the HBCU-trained teacher variable remains essentially unchanged in column 3. Column 4 combines the controls from the two previous columns, slightly reducing the magnitude of the coefficient, but hardly changing the effect size. In column 5, I extend the regression model with richer student-level covariates that theoretically impact test scores, such as whether the student was suspended during the school year, their number of days absent from school, and whether the student had the same teacher in a previous year. These variables do not substantially change the estimates presented in previous columns. They do, however, introduce additional endogeneity concerns, as these variables may also be influenced by a student's teacher. As such, in subsequent specifications, I will build on results from column 4, which will allow for within-student comparisons over time of HBCU- and non-HBCU-trained teacher effects while also accounting for teacher-level and classroom-level differences.

Notwithstanding the focus on achievement for Black students, it is reasonable to consider if these results hold for all students and students from other demographic groups. The strong, positive relationships between HBCU-trained teachers and Black student achievement are illuminating, but ideally this does not come at the expense of other students in the classroom. Incidentally, this is not the case, as shown in Table A2. Using my preferred specification (column 4 from Table 2), I explore HBCU-trained teacher effects on student achievement for all students, White students, and Hispanic students.[12] The coefficient is still positive for all students, though the magnitude is less than half the size of the result for Black students only. Coupled with the statistically insignificant results for regressions with White (and a positive, but smaller association for Hispanic students), these findings suggest the increases in Black student achievement associated with teachers from HBCUs do not come with any meaningful decreases in achievement for other students.

Finally, however robust these results are to additional controls, there still exists the possibility of bias from the estimating strategy altogether (i.e., using a two-way fixed effects estimator in the presence of differential treatment timing). I follow

[12]Following from Table 1, Asian students represent around 2.6% of the students in the sample, and all other race/ethnicity groups sum to less than 6% of students. As such, this disaggregated analysis is limited to White and Hispanic students.

the guidance from recent literature on TWFE bias and implement two additional specifications. First, I estimate the effect of having an HBCU-trained teacher for the first time in the 5th grade, comparing this student's test performance to students who never receive an HBCU-trained teacher. Second, using the multidid package in Stata, I implement the estimator derived in de Chaisemartin and D'Haultfœuille (2020b) that is robust to the differential treatment timing concern.[13] As presented in Table A3, both coefficients are positive, statistically significant and, in fact, larger than the estimated effect sizes in Table 2, suggesting the TWFE specification may be understating the true impact of HBCU-trained teachers on Black student math exam scores.[14]

5.1 Heterogeneous effects: Comparing HBCU-trained and non-HBCU-trained teachers of the same race

Building on the same-race teacher effects literature, a straightforward explanation for these findings is that they merely reflect a correlate with teacher race, with results that are driven by Black teachers who also happen to have attended an HBCU. Admittedly, all prior regressions control for teacher race, so statistically this seems unlikely from the start. Nevertheless, I more explicitly draw further distinction between these channels in the following paragraphs and strengthen the evidence of race-independent HBCU teacher effects in two ways. First, I disaggregate the analysis of HBCU-teacher effectiveness for Black students by their teachers' race. Second, as in previous same-race teacher effects papers, I re-estimate my preferred regression using Black teachers rather that HBCU attendees.

Empirically, the first approach interacts HBCU attendance and teacher race and takes the form

[13] I also use their twowayfeweights command in Stata to probe any negative weighting issues among the various ATTs that TWFE estimators average over. Accordingly, only around 14% of the weights were negative, summing to approximately 0.031, further indicating negative weights are unlikely to be an empirical issue in this setting.

[14] Point estimates are also largely robust to a series of other, alternative specifications, which are also presented in Table A3: the inclusion of school-by-year fixed effects, explicitly controlling for prior student achievement, the restriction of the sample to students observed over all grades, and, following the empirical strategy of Bettinger and Long (2005), instrumenting for teacher assignment with the share of available HBCU-trained teachers in a given school year.

$$A_{ijst} = \beta_1(HBCU, Black)_{jt} + \beta_2(HBCU, nonBlack)_{jt} + \beta_3(nonHBCU, nonBlack)_{jt} + X_{it}'\Gamma + U_j'\Delta + V_c'\xi + Z_{st}'H + \theta_i + \lambda_t + \epsilon_{ijst} \quad (2)$$

Vectors of student-level, teacher-level, classroom-level, and school-level controls are represented by X, U, V, and Z, respectively. The β coefficients capture the relationships between teacher assignment and student test scores for different types of teachers relative to a Black teacher who did not attend an HBCU. Of these, I focus on β_1, which compares directly Black teachers who went to HBCUs to those who did not, exploring any heterogeneity in same-race teacher matches.

Column 1 of Table 3 contains results for Black HBCU-trained and non-HBCU-trained teachers, where non-HBCU-trained Black teachers are the reference category. I find evidence consistent with the proposition that Black teacher effectiveness in part varies with HBCU attendance: Black students in classrooms with Black teachers who attended an HBCU score on average 0.026 standard deviations higher on math exams, compared with their performance in years with a non-HBCU-trained Black teacher. This result is key to a point discussed in Section 2 regarding Black teachers and contemporary teaching practices. Culturally-fluent pedagogy is a broad descriptive, encompassing both hard and soft skills that a teacher may utilize in the classroom. Given the historical record of advances in this praxis being driven by Black teachers and principals, on its face it seems reasonable to view contemporary Black teachers together as inheritors of this educational thought. Were this pedagogical style broadly applicable to all Black teachers in the sample, however, it is unlikely we would observe such a stark contrast in effectiveness between different subgroups. Instead, the evidence provided suggests there is something about HBCU attendance that signifies an important distinction between Black teachers, in ways that should be interrogated further to understand same-race teacher effects.

Building on this point, I extend this specification to White teachers in the sample, repeating the prior regression but instead focusing on comparing HBCU-trained White teachers to White teachers who attended other institutions (column 2). As with Black teachers, I find that Black students in classrooms with White teachers who attended an HBCU score on average 0.053 standard deviations higher on math exams than they do with non-HBCU-trained White teachers. In addition to reinforcing the mounting evidence of the important roles HBCUs play in Black student achievement,

these results imply HBCUs have a race-independent effect: for students, the relevant characteristic for improving academic achievement may not simply be a racial match with their teacher, but rather the teachers' prior experiences, including their teacher preparation.

Continuing the exploration of teacher effects by HBCU attendance and race, I also specify regressions that model the relationship between all Black teachers and Black student test scores.

$$A_{ijst} = \mu(BlackTeacher)_{jt} + X_{it}^{'}\tilde{\Gamma} + U_{j}^{'}\tilde{\Delta} + V_{c}^{'}\tilde{\xi} + Z_{st}^{'}\tilde{H} + \theta_i + \lambda_t + u_{ijst} \quad (3)$$

Control vectors X, U, V, and H are the same as with Equation (2). Analogous to Equation (1) with an HBCU-trained teacher indicator, I now instead use an indicator for whether the teacher is Black. Results in Table 4 differ considerably from those shown for HBCU-trained teachers in Table 2: not only is the coefficient magnitude much smaller than the estimated HBCU-trained teacher effect, but the point estimate is statistically insignificant, showing no discernible test score improvement for Black students paired with Black teachers. Additional regressions (shown in columns 2 and 3) unpack this further, where I replace the indicator for all Black teachers with indicators for HBCU-trained Black teacher and non-HBCU-trained Black teacher assignment, respectively. The relationship between HBCU-trained Black teachers and Black student test scores mirrors that shown previously for all HBCU-trained teachers (albeit with a smaller magnitude), while I observe a larger, negative impact from assignment to a non-HBCU-trained Black teacher. Taken together, I find a positive same-race match effect for Black teachers and students only when I condition on that Black teacher having attended an HBCU.

As a final point, note that in additional regressions (available upon request) that link Hispanic teachers with Hispanic students, I similarly find no evidence that Hispanic students score better on math exams when assigned to a same-race teacher. Combined with the previous result for Black teacher-student matches, I find evidence for academic returns to same-race teacher pairings for racial/ethnic minority students in this setting is weak, at best.[15]

[15]That I find no general positive effect for Black teachers on Black students may result from several key differences from other works. For instance, I consider effects on a short-run outcome (testing), rather than one in the long-run like educational attainment (e.g., Gershenson et al., 2022), so it is possible the mechanisms driving the results are different. Other differences may arise from varying model specifications. I consider level changes in student test scores, while Hanushek et al.

5.2 Dynamic sorting by teacher experience

The prior results adjust for differences in teacher experience — typically found relevant to teacher quality (e.g., Wiswall, 2013; Kraft and Papay, 2015) — by controlling for a teacher's years teaching in the state, but this may not fully capture the role experience plays in the classroom if there is student sorting to teachers conditioned on the teachers' experience. Previous research has suggested this could be especially true for more experienced teachers, who may have more autonomy over choosing the students in their classrooms or which classes to teach (Kalogrides and Loeb, 2013). Concurrently, less experienced teachers are often assigned students with lower previous academic achievement. Given that HBCU-trained teachers in the sample are, on average, more experienced than their peers, this practice could systematically confound estimates of HBCU-trained teacher effects.

Relatedly, teacher experience may complicate estimates through on-the-job learning and any skills teachers may pick up post-graduation from their preservice teacher training. By controlling for teacher experience, we can make claims about HBCU-trained teacher effectiveness relative to non-HBCU-trained teachers holding experience fixed, but, if what we are actually interested in is the unique effect of HBCU preparation itself, this specification is unable to disentangle the effects of an HBCU education from the effects of their graduates spending more time in the classroom and sharpening their craft.

Without richer data about teacher education programs and student sorting into different colleges, I am unable to test for differences directly. However, for an approximation in that direction, I use another specification whereby I interact the HBCU attendance variable with indicators for whether the teacher is in their first three years teaching. In addition to assuaging concern that prior regressions conflated HBCU-trained teacher effects with their ability to choose students who would have likely scored higher on the math exam regardless of their teacher's college attendance, these regressions allow for a better comparison of teachers relatively fresh out of their preservice training, gauging which teachers are most effective when they are relying solely or mostly on preparation rather than any on-the-job learning.

(2005) use changes in test score growth as the outcome variable; my main regressions use student fixed effects, whereas Bartanen and Grissom (2021) compare results across students with school fixed effects; and Egalite et al. (2015) control for teacher and time-varying school characteristics, all of which are included in my model in addition to more teacher (e.g., teacher age, certification exam scores) and classroom-level (e.g., class size) controls. That said, I am also not the first to find a null or negative same-race teacher effect for Black students (e.g., Ehrenberg et al., 1995; Jennings and DiPrete, 2010).

For these regressions, I use indicators to partition the data set into four categories based on teacher experience. The coefficient in column 1 of Table 5 results from a specification evaluating teacher effects for four different groups: HBCU-trained first-years, non HBCU-trained first-years, HBCU-trained teachers with more than one year of experience, and non HBCU-trained with more than one year of experience (the non HBCU-trained first year teacher indicator is omitted as the reference group). I show a significant (though imprecisely estimated) difference in expected test scores for Black students with first-year HBCU-trained teachers, with average test scores 0.010 standard deviations higher than their exam scores in years with other first-year teachers. Subsequent columns report even larger, statistically significant coefficients with expanded samples that include second- and third-year teachers (0.040 and 0.033, respectively). The evidence here suggests that HBCU-trained teachers are not only more effective with Black students in general, but they are also effective virtually from the start, independent of any learning on the job and peer effects that could artificially inflate HBCU impacts. This effectiveness may have additional importance following research finding greater success early in teachers' careers may correlate with staying in the classroom longer (Henry, Bastian, and Fortner, 2011).

5.3 Classroom dynamics beyond test scores: The relationship between HBCU-trained teachers and student suspensions

The positive, HBCU-trained teacher effects on achievement I observe may arise from how teachers behave in the classroom, with differences in how teachers relate to their students and how they conduct classroom management. Hence, one avenue for investigating heterogeneous effects of teachers on Black student achievement may lie with differences in disciplinary actions. Researchers have consistently showed the detrimental impacts suspensions can have on students, including lower test scores for those suspended (Pope and Zuo, 2019), increased rates of school dropouts (Lee, Cornell, Gregory, and Fan, 2011), and lower graduation rates and higher rates of being arrested and incarcerated (Bacher-Hicks, Billings, and Deming, 2019). Further, empirical evidence demonstrates these impacts are disproportionately realized, with disparities in suspension rates for Black students (Barrett et al., 2019). In 2014, data from the U.S. Department of Education Office for Civil Rights showed Black K-12 students were suspended and expelled "at a rate three times greater than White students" (U.S. Department of Education Office for Civil Rights, 2014). Analyzing this channel is indispensable given this research's focus on Black students, and especially

so with over one-fifth of Black students in the sample receiving a suspension judgment at some point. Evaluating suspensions can be a key element in understanding student academic success, not only in that those not suspended have the capacity to spend more time in the classroom, but they may also speak to the teacher's connection with the student, critical to a student's academic performance.

$$Suspension_{ijcgst} = \rho(HBCU)_{jt} + \ddot{X}_{it}^{'}\tau + \ddot{W}_{j}^{'}\phi + \ddot{Z}_{st}^{'}\chi + \psi_g + \omega_{st} + \zeta_{ijcgst} \quad (4)$$

As specified in Equation (4), I use a linear probability model to predict the probability that a Black student i with teacher j in classroom c and grade g receives a suspension judgment in a given school-year st. Whereas previous specifications used within-student variation in teacher assignment, I compare suspension probabilities across students. Specifically, I regress whether the student has a suspension judgment on an indicator for teacher assignment j (HBCU-trained or not), time-varying student-, teacher, and classroom-level covariates ($\ddot{X}$, $\ddot{Y}$, and $\ddot{Z}$, respectively) — which include any suspension receipt in the prior year — and grade and school-by-year fixed effects.[16] The school-by-year fixed effect in particular is vital to accounting for any changes in school administration and leadership over time that may influence school suspension rates.

Figure 2 plots coefficients representing the association between HBCU-trained teacher assignment and the probability of being suspended for Black students, with the analogous correlations for Black and White teachers included for reference. The signs on each coefficient are as expected: while Black students have a higher probability of suspension when assigned to a White teacher, the change in suspension likelihood is negative with Black teachers and with teachers from HBCUs, even after controlling for the teacher's race.

Notably, this result on its own masks important, gender-specific variation in suspension rates.[17] The association between HBCU-trained teacher assignment and suspension probability for Black female students is essentially zero, obscuring that the association for Black male students is sizable: the magnitude is over twice as large as the coefficient for all students (-0.015, SE 0.005). As a point of reference, the mean

[16]Each of the student-, teacher-, and classroom-level covariates is dichotomized to zero or one.

[17]For robustness, I return to my preferred specification (column 4 of Table 2) and disaggregate by student gender. Unlike the regressions focused on suspension outcomes, it does not appear that effects on test scores are entirely driven by results for Black boys, though the point estimate is larger for Black boys than the estimate for Black girls (0.035, SE 0.007 for boys; 0.030, SE 0.007 for girls).

predicted probability of suspension for Black male students is approximately 0.23, with a standard deviation of 0.42. Thus, assignment to an HBCU-trained teacher is associated with roughly a 0.036 SD change in the probability of being suspended for Black boys.

I also disaggregate the relationship between teacher race and Black student suspension probability by whether the teacher attended an HBCU (Figure 3). As with the test score analysis discussed above, the likelihood of suspension for Black boys declines with assignment to an HBCU-trained teacher, even among teachers of the same race. This is particularly stark when comparing White teachers to HBCU-trained White teachers: whereas the probability of suspension generally increases when a Black student has a White teacher, White teachers who went to an HBCU are less likely to suspend their Black male students, compared with other White teachers (-0.026, SE 0.010).

Combining these results with Figure 4, which shows associations between HBCU-trained teachers and suspension for all, White, and Hispanic students are insignificant and/or negligible, the presented evidence is consistent with the proposal that teachers from these universities have a particular approach to teaching (and classroom management) that uniquely correlates with higher achievement for Black students. Irvine (1989) outlines the predicament Black students can often find themselves in in schools due to cultural conflict: students are disadvantaged when placed with teachers who "do not understand minority student's behavior, physical movements, verbal and nonverbal language, values, worldview, home environment, and learning styles" (pgs. 55-56). It is challenging empirically to separate student behavior when assigned to certain teachers from general teacher behavioral expectations in the classroom. Nevertheless, to some extent, the degree to which the success of Black students stems from access to teachers who serve as cultural translators, or at the very least teachers with the capacity and willingness to guide Black students in bridging these gaps, may be reflected in differences in suspension rates.

6 Discussion of Potential Mechanisms

Why are HBCU-trained teachers more effective with Black students than their non-HBCU graduate peers? In section 2, I argued HBCU-trained teachers are likely stronger teachers for Black students due to a teacher education at these institutions that is historically rooted in pedagogical practices particularly well-suited for Black

students. I then supported this theoretical foundation with empirical evidence from North Carolina, detailed in section 4, ruling out a same-race teacher match explanation, as HBCU trained teachers show greater effectiveness even among teachers of the same race, and ruling out HBCU graduate-specific gains from experience, as even novice, HBCU-trained teachers are more effective with Black students than similarly inexperienced, non-HBCU-trained teachers (Table 5).

One unexplored explanation is selection of students into these colleges: who goes to an HBCU? For example, it is fairly uncommon for White students to attend HBCUs; the motivations for attending these colleges and attitudes towards working with Black people may well likely vary in important, unobservable ways from other White college graduates. Some qualitative work (e.g., Freeman, 1999) has documented a series of possible factors influencing one's choice to attend an HBCU — including a desire for greater cultural awareness about and cultural connections to the Black community — but little research has explicitly focused on prospective teachers who choose HBCUs. More generally, perhaps it is less about the institutions themselves and more about the types of students who attend them. Students with propensities to be better teachers may systematically sort into HBCUs. Here, difficulty arises from imprecision in how to measure these traits. The skills and characteristics of a high school student most likely to be an effective teacher are not well understood. Moreover, the data set I have does not include any measures of teacher performance or ability prior to college matriculation. Some insights may be gleaned, however, when comparing group-level average statistics for the colleges teachers attended.

I obtained data for the percent of students admitted to college for each year between 2001 and 2013 and the 25th and 75th percentile scores for the math and reading sections of the SAT from the publicly-available Integrated Postsecondary Education Data System (IPEDS). I average these measures within-college during these time frames, then compute the averages across colleges conditional on being an HBCU or not. I present means and standard deviations in Table A4. On average, whereas HBCUs in the sample are more selective than the non-HBCUs, the middle 50% of students admitted to HBCUs score significantly lower on both the math and reading sections of the SATs.

This incongruence between college entrance exam scores and teacher effectiveness draws attention to the metrics used for admissions to HBCUs and how they may deviate from traditional measures of achievement, which could be relevant for understanding the types of students who select into these colleges. Recalling an earlier

finding that the typical HBCU-trained teacher in the analytic sample scores significantly lower on teacher certification exams than the typical teacher who attended a non-HBCU, it is unclear how to reconcile the gaps in supposed measures of college quality and classroom readiness with findings showing HBCU-trained teachers from "weaker" backgrounds are outperforming their peers with Black students. Scholars for decades have raised concerns regarding certification testing's negative impact on racial and ethnic minority teacher recruitment (Goertiz and Picher, 1985; G.P. Smith, 1987), as well as the extent to which they meaningfully predict teacher quality (Garcia, 1986; Popham, 1986), a sentiment also reflected in contemporary research (e.g., Goldhaber and Hansen, 2010). At a minimum, this challenges what licensure exams are measuring and whether they accurately capture proficiency in skills critical to nurturing student academic success, particularly for Black students.

7 Conclusion

This is, to my knowledge, the first quantitative paper studying the effect of HBCU-trained teachers on Black student achievement. Through a variety of econometric approaches, I show not only do Black elementary school students in North Carolina perform better on math exams and are less likely to receive disciplinary action with an HBCU-trained teacher, but that this relationship extends beyond same-race teacher matches. While I focus on an educational setting, the results here illuminate the interacting effects of race, outlook, and background on outcomes apply more broadly and further contribute to our intuition of the gains from same-race pairings.

Conventional wisdom on same-race teacher matches can potentially lead to an increased burden placed on racial/ethnic minority teachers. Through this lens, it may be easy for non-Black teachers to undersell their capacity for contributing to Black student excellence as a function of their racial mismatch. My findings argue the contrary. Whether it is the result of training or the teachers themselves, there is a clear pattern of teachers from Historically Black Colleges and Universities, independent of race, connecting with Black students in a way that advances academic achievement and improves behavioral outcomes. Racial achievement gaps stem not from inherent cognitive or cultural deficits among Black students, but from historical and contemporary disparities in resources allotted to students and the schools they attend (Ladson-Billings, 2006; Anderson, 2007). Closing them, then, necessarily requires that Black students are afforded the resources requisite of an enriching

educational environment. Policy often discusses economic inequities, but these gaps can also extend to personnel disparities: students need access to high-quality teachers committed to both their academic success and mental and emotional well-being. As evident in this study, teachers from historically Black colleges can play an extraordinary role in filling these positions for Black students.

For the policy-minded, this evidence highlights the potential benefits to hiring and retaining more teachers trained at Historically Black Colleges and Universities.[18] HBCUs already produce a substantial portion of Black teachers, making them distinctly qualified to help mitigate the undersupply of Black teachers in the U.S. Earlier scholars have proposed this initiative (e.g., Irvine and Fenwick, 2011); this paper quantifies and supports the effects underlying their argument. In the face of concerns regarding teacher recruitment and retention and their effect on students, these results speak to a simple, effective, and targeted approach for raising student achievement largely omitted from prior teacher labor market discussions. Incidentally, this is at odds with the declining enrollment shares (Hinrichs, 2015) and financial woes (Smith-Barrow, 2019) that have plagued many HBCUs in recent years, a macrolevel challenge at best overlooked by policymakers and, at worst, outright dismissed. Indeed, as discussed in Harper (2017), a contingency of Predominantly White Institution (PWI) advocates maintains that HBCUs are relics of the past, no longer relevant with increased resource and network opportunities for Black students at PWIs. On the contrary, this paper suggests HBCUs are uniquely producing teachers capable of elevating Black academic success and, consequently, policymakers and elected officials should instead grapple with what gains may be lost in the absence of these institutions.

[18] Any discussion on teacher recruitment necessitates one on teacher retention (Ingersoll, May, and Collins, 2019). Policymakers should be mindful of not only attracting HBCU graduates, but also keeping these teachers in the classroom.

References

Abrams, D. S., M. Bertrand, and S. Mullainathan (2012). Do Judges Vary in their Treatment of Race? *Journal of Legal Studies 41*, 347–384.

Acosta, M. M., M. Foster, and D. F. Houchen (2018). 'Why Seek the Living Among the Dead?' African American Pedagogical Excellence: Exemplar Practice for Teacher Education. *Journal of Teacher Education 69*(4), 341–353.

Akbar, R. and M. J. Sims (2008). Surviving Katrina and Keeping Our Eyes on the Prize: The Strength of Legacy and Tradition in New Orleans's HBCU Teacher Preparation Programs. *Urban Education 43*(4), 445–462.

Anderson, J. (2007). The Historical Context for Understanding the Test Score Gap. *National Journal of Urban Education and Practice 1*, 1–21.

Aronson, J. (2002). Chapter 14 – Stereotype Threat: Contending and Coping with Unnerving Expectations. In *Improving Academic Achievement: Impact of Psychological Factors on Education*, pp. 279–301.

Arroyo, A. T. and M. Gasman (2014). An HBCU-Based Educational Approach for Black College Student Success: Toward a Framework with Implications for All Institutions. *American Journal of Education 121*(1), 57–85.

Ba, B. A., D. Knox, J. Mummolo, and R. Rivera (2021). The role of officer race and gender in police-civilian interactions in Chicago. *Science 371*, 696–702.

Bacher-Hicks, A., S. B. Billings, and D. J. Deming (2019). The School to Prison Pipeline: Long-Run Impacts of School Suspensions on Adult Crime. *NBER Working Paper* (w26257).

Barrett, N., A. McEachin, J. N. Mills, and J. Valant (2020). Disparities and Discrimination in Student Discipline by Race and Family Income. *The Journal of Human Resources 55*(4), 1–51.

Bartanen, B. and J. Grissom (2021). School Principal Race and the Hiring and Retention of Racially Diverse Teachers. *Journal of Human Resources*, 1–68.

Bertrand, M., E. Duflo, and S. Mullainathan (2004). How Much Should We Trust Differences-in-Differences Estimates? *Quarterly Journal of Economics 119*, 249–275.

Boyd, D. J., P. L. Grossman, H. Lankford, S. Loeb, and J. Wyckoff (2009). Teacher Preparation and Student Achievement. *Educational Evaluation and Policy Analysis 31*(4), 416–440.

Cameron, C., J. Gelbach, and D. L. Miller (2011). Robust Inference with Multi-way clustering. *Journal of Business and Economic Statistics 21*(2), 238–49.

Chaisemartin. Difference-in-Differences Estimators of Intertemporal Treatment Effects. In *SSRN (2020b*.

Chaisemartin, Clement, and X. D'Haultføeuille. Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects. *American Economic Review 110*(2020a), 2964–2996.

Chetty, R., J. N. Friedman, and J. E. Rockoff (2014). Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates. *American Economic Review 104*(9), 2593–2632.

Clotfelter, C. T., H. F. Ladd, and J. L. Vigdor (2007). How and Why Do Teacher Credentials Matter for Student Achievement? *NBER Working Paper* (w12828).

Clotfelter, C. T., H. F. Ladd, and J. L. Vigdor (2010). Teacher Credentials and Student Achievement in High School: A Cross-Subject Analysis with Student Fixed Effects. *The Journal of Human Resources 45*(3), 655–681.

Dan, G., T. Gratz, and R. Theobald (2017). What's in a teacher test? Assessing the relationship between teacher licensure test scores and student STEM achievement and course-taking. *Economics of Education Review 61*, 112–129.

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

Dee, T. (2005). A Teacher Like Me: Does Race, Ethnicity or Gender Matter? *American Economic Review 95*(2).

Dee, T. and E. Penner (2017). The Causal Effects of Cultural Relevance: Evidence From an Ethnic Studies Curriculum. *American Educational Research Journal 54*(1), 127–166.

Dudley-Marling, C. and K. Lucas (2009). Pathologizing the Language and Culture of Poor Children. *Language Arts 86*(5), 362–370.

Egalite, A. J. and B. Kisida (2018). The Effects of Teacher Match on Students' Academic Perceptions and Attitudes. *Educational Evaluation and Policy Analysis 40*(1), 59–81.

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

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

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

Freeman, K. (1999). HBCs or PWIs? African American High School Students' Consideration of Higher Education Institution Types. *The Review of Higher Education 23*(1), 91–106.

Garcia, P. (1986). The impact of national testing on ethnic minorities: With proposed solutions. *Journal of Negro Education 55*(3), 347–357.

Gay, G. and K. Kirkland (2003). Developing Cultural Critical Consciousness and Self-Reflection in Preservice Teacher Education. *Theory Into Practice 42*(3), 181–187.

Gershenson, S., C. M. Hart, J. Hyman, N. W. Papageorge, and C. A. Lindsay (2022). The Long-Run Impacts of Same Race Teachers. *American Economic Journal: Economic Policy*.

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

Ginsberg, A., M. Gasman, A. C. Samoya, and F. Ramos (2015). *Re-assessing Teacher Education Quality: What NCTQ Isn't Telling Us about the Impact and Value of HBCUs*. Teachers College Record.

Goertz, M. and B. Picher (1985). *The impact of NTE use by states on teacher selection*. Princeton. NJ: Educational Testing Service.

Goldhaber, D. (2017). Does the Match Matter? Exploring Whether Student Teaching Experiences Affect Teacher Effectiveness. *American Educational Research Journal 54*(2), 325–359.

Goldhaber, D., S. Liddle, and R. Theobald (2013). The Gateway to the Profession: Assessing Teacher Preparation Programs Based on Student Achievement. *Economics of Education Review 34*, 29–44.

Greenwood, B. N., R. R. Hardeman, L. Huang, and A. Sojourner (2020). Physician-patient racial concordance and disparities in birthing mortality for newborns. *Proceedings of the National Academy of Sciences of the United States of America 117* (35), 21194–21200.

Grissom, J. A. and L. R. Keiser (2011). A Supervisor Like Me: Race, Representation, and the Satisfaction and Turnover Decision of Public Sector Employees. *Journal of Policy Analysis and Management 30*(3), 557–580.

Hanushek, E. A., J. F. Kain, J. M. Markman, and S. G. Rivkin (2001). Does Peer Ability Affect Student Achievement? *NBER Working Paper* (w8502).

Hanushek, E. A., J. F. Kain, D. M. O'Brien, and S. G. Rivkin (2005). The Market for Teacher Quality. *NBER Working Paper* (w11154).

Harper, B. (2017). Teacher Preparation at Historically Black Colleges and Universities: Remaining Relevant in a Climate of Accountability. In M. Lynnette and P. Emery (Eds.), *Teacher Education across Minority-Serving Institutions: Programs, Policies, and Social Justice*, pp. 161–76. New Brunswick, Camden, Newark, New Jersey; London: Rutgers University Press.

Harper, S. R. and C. H. Davis, III (2012). They (Don't) Care about Education: A Counternarrative on Black Male Students' Responses to Inequitable Schooling. *Educational Foundations 26*(1-2), 103–120.

Harris, A. L. (2008). Optimism in the Face of Despair: Black-White Differences in Beliefs About School as a Means for Upward Social Mobility. *Social Science Quarterly 89*(3), 608–630.

Harris, A. L. (2011). *Kids Don't Want to Fail*. Cambridge, Massachusetts; London, England: Harvard University Press.

Hart, C. M. (2020). An Honors Teacher Like Me: Effects of Access to Same-Race Teachers on Black Students' Advanced-Track Enrollment and Performance. *Educational Evaluation and Policy Analysis 42*(2), 163–187.

Headley, A. M. and J. E. Wright, II (2020). Is Representation Enough? Racial Disparities in Levels of Force and Arrests by Police. *Public Administration Review 80*(6), 1051–1062.

Henry, G. T., K. C. Bastian, and C. Fortner (2011). Stayers and Leavers: Early-Career Teacher Effectiveness and Attrition. *Educational Researcher 40*(6), 271–280.

Hill, A. J. and D. B. Jones (2021). Paying for Whose Performance? Teacher Incentive Pay and the Black-White Test Score Gap. *Educational Evaluation and Policy Analysis 43*(3), 445–471.

Hinrichs, P. (2015). An Empirical Analysis of Racial Segregation in Higher Education. *NBER Working Paper* (w21831).

Holt, S. B. and S. Gershenson (2019). The Impact of Demographic Representation on Absences and Suspensions. *Policy Studies Journal 47*(4).

Ingersoll, R., H. May, and G. Collins (2017). Minority teacher recruitment, employment and retention: 1987 to 2013. Technical report, Learning Policy Institute.

Irvine, J. J. (1989). Beyond Role Models: An Examination of Cultural Influences on the Pedagogical Perspectives of Black Teachers. *Peabody Journal of Education 66*(4), 51–63.

Irvine, J. J. (1990). Black and white students' role models.

Irvine, J. J. and L. T. Fenwick (2011). Teachers and Teaching for the New Millennium: The Role of HBCUs. *The Journal of Negro Education 80*(3), 197–208.

Jackson, K. C. (2018). What Do Test Scores Miss? The Importance of Teacher Effects on Non–Test Score Outcomes. *Journal of Political Economy 126*(5), 2072–2107.

Jennings, J. L. and T. A. DiPrete (2010). Teacher Effects on Social and Behavioral Skills in Early Elementary School. *Sociology of Education 83*(2), 135–159.

Joshi, E., S. Doan, and M. G. Springer (2018). Student-Teacher Race Congruence: New Evidence and Insight from Tennessee. *AERA Open 4*(4), 1–25.

Kalogrides, D. and S. Loeb (2013). Different Teachers, Different Peers. *Educational Researcher 42*(6), 304–316.

Kalogrides, D., S. Loeb, and T. Béteille (2012). Systematic Sorting: Teacher Characteristics and Class Assignments. *Sociology of Education 86*(2), 103–123.

Klopfenstein, K. (2005). Beyond Test Scores: The Impact of Black Teacher Role Models on Rigorous Math Taking. *Contemporary Economic Policy 23*(3), 416–428.

Koedel, C., K. Mihaly, and J. E. Rockoff (2015). Value-Added Modeling: A Review. *Economics of Education Review 47*(ue C), 180–195.

Koedel, C., E. Parsons, M. Podgursky, and M. Ehlert (2015). Teacher Preparation Programs and Teacher Quality: Are there Real Differences Across Programs? *Education Finance & Policy 10*(4), 508–534.

Koefoed, M. S. and E. mcGovney (2019). The Effect of Same-Gender or Same-Race Models on Occupation Choice: Evidence from Randomly Assigned Mentors at West Point. *Journal of Human Resources 54*(2), 430–467.

Krueger, A. B. (1999). Experimental Estimates of Education Production Functions. *The Quarterly Journal of Economics 114*(2), 497–532.

Kumar, R. and F. Lauermann (2018). Cultural Beliefs and Instructional Intentions: Do Experiences in Teacher Education Institutions Matter? *American Educational Research Journal 55*(3), 419–452.

Ladson-Billings, G. (1995). Toward a Theory of Culturally Relevant Pedagogy. *American Educational Research Journal 32*(3), 465–491.

Ladson-Billings, G. (2006). From the Achievement Gap to the Education Debt: Understanding Achievement in U.S. Schools. *Educational Researcher 35*(7), 3–12.

Ladson-Billings, G. (2014). Culturally Relevant Pedagogy 2.0: a.k.a. the Remix. *Harvard Educational Review 84*(1), 74–84.

Ladson-Billings, G. and J. D. Anderson (2021). Policy Dialogue: Black Teachers of the Past, Present, and Future. *History of Education Quarterly 61*(1), 94–102.

Lawrence, J. S., B. T. Marks, and J. S. Jackson (2010). Domain identification predicts black students' underperformance on moderately-difficult tests. *Motivation and Emotion 34*(2), 105–109.

Lee, T., D. Cornell, A. Anne Gregory, and X. Fan (2001). High suspension schools and dropout rates for Black and White students. *Education & Treatment of Children 34*(2), 167–192.

Lindsay, C. A. and C. M. Hart (2017). Exposure to Same-Race Teachers and Student Disciplinary Outcomes for Black Students in North Carolina. *Educational Evaluation and Policy Analysis 39*(3), 485–510.

Maylor, U. (2009). 'They do not relate to Black people like us': Black teachers as role models for Black pupils. *Journal of Education Policy 24*(1), 1–21.

Mihaly, K., D. McCaffrey, T. R. Sass, and L. J.R (2013). Where You Come from or Where You Go? Distinguishing between School Quality and the Effectiveness of Teacher Preparation Program Graduates. *Education Finance and Policy 8*(4), 459–493.

Mykerezi, E. and B. F. Mills (2008). The Wage Earnings Impact of Historically Black Colleges and Universities. *Southern Economic Journal 75*(1), 173–187.

of Education Office for Civil Rights, U. D. (2016a). 2013-14 Civil Rights Data Collection: A first look.

of Education Office for Civil Rights, U. D. (2016b). The State of Racial Diversity in the Educator Workforce. Place: Washington, D.C.

Papageorge, N. W., S. Gershenson, and K. M. Kang (2020). Teacher Expectations Matter. *The Review of Economics and Statistics 102*(2), 234–251.

Papay, J. P. and M. A. Kraft (2015). Productivity Returns to Experience in the Teacher Labor Market: Methodological Challenges and New Evidence on Long-Term Career Improvement. *Journal of Public Economics 130*, 105–119.

Paris, D. (2012). Culturally Sustaining Pedagogy: A Needed Change in Stance, Terminology, and Practice. *Educational Researcher 41*(3), 93–97.

Paufler, N. A. and A. Amrein-Beardsley (2014). The Random Assignment of Students Into Elementary Classrooms: Implications for Value-Added Analyses and Interpretations. *American Educational Research Journal 51*(2), 328–362.

Penney, J. (2017). Racial Interaction Effects and Student Achievement. *Education Finance and Policy 12*(4), 447–467.

Pope, N. and G. W. Zuo (2020). Suspending Suspensions: The Education Production Consequences of School Suspension Policies.

Popham, W. J. (1986). Teacher competency testing: The devil's dilemma. *Journal of Negro Education 55*(3), 379–386.

Price, G. N., W. Spriggs, and O. H. Swinton (2011). The Relative Returns to Graduating from a Historically Black College/University: Propensity Score Matching Estimates from the National Survey of Black Americans Price. *Review of Black Economy 38*(2), 103–130.

Price, J. and J. Wolfers (2010). Racial Discrimination Among NBA Referees. *The Quarterly Journal of Economics*, 1859–1887.

Reardon, S. F., D. Kalogrides, and K. Shores (2019). The Geography of Racial/Ethnic Test Score Gaps. *American Journal of Sociology 124*(4), 1164–1121.

Roebuck, J. B. and K. Murty (1993). Historically Black Colleges and Universities: Their Place in American Higher Education.

Rose, E. K. (2021). Who Gets a Second Chance? Effectiveness and Equity in Supervision of Criminal Offenders. *The Quarterly Journal of Economics*, 1199–1253.

Sacerdote, B. L. (2001). Peer effects with random assignment: Results for Dartmouth roommates. *Quarterly Journal of Economics 116*(2), 681–704.

Sav, G. T. (2010). Funding Historically Black Colleges and Universities: Progress toward Equality? *Journal of Education Finance 35*(3), 295–307.

Shelvin, K. H., R. Rivadeneyra, and C. Zimmerman (2014). Stereotype threat in African American children: The role of Black identity and stereotype awareness. *Revue Internationale de Psychologie Sociale 27*(3-4), 175–204.

Smith, G. (1982). Increasing the number of minority teachers: Recommendations for a call for action. In *Paper prepared for the Quality Education for Minorities Project*. Boston: Massachusetts Institute of Technology.

Smith, G. (1987). *The effects of competency testing on the supply of minority teachers. A report prepared for the Quality Education for Minorities Project*. Cambridge, MA: Massachusetts institute of Technology.

Smith-Barrow, D. (2019, October). Many HBCUs are teetering between surviving and thriving. The Hechinger Report,.

Steffensmeier, D. and C. L. Britt (2001). Judges' Race and Judicial Decision Making: Do Black Judges Sentence Differently? *Social Science Quarterly 82*(4), 749–764.

Thompson, O. (2021). School Desegregation and Black Teacher Employment. *The Review of Economics and Statistics*.

Villegas, A. M. and T. F. Lucas (2004). Diversifying the Teacher Workforce: A Retrospective and Prospective Analysis. *Yearbook of the National Society for the Study of Education 103*(1), 70–104.

Wasserberg, M. J. Stereotype threat effects on African American children in an urban elementary school. *The Journal of Experimental Education 4*(2014a), 502–517.

Watkins, A. F. (2005). Cultivating the Education of African American College Students: A Learning Styles Approach. In L. B. Gallien and M. S. Peterson (Eds.), *Instructing and Mentoring the African American College Student: Strategies for Success in Higher Education*, pp. 122–27. Boston: Pearson Education.

Wayne, A. J. and P. Youngs (2003). Teacher Characteristics and Student Achievement Gains. *A Review." 73*(1), 89–122.

Wilson, V. R. (2007). The Effect of Attending an HBCU on Persistence and Graduation Outcomes of African-American College Students. *The Review of Black Political Economy 34*(1–2), 11–52.

Wiswall, M. (2013). The Dynamics of Teacher Quality. *Journal of Public Economics 100*, 61–78.

Yarnell, L. M. and G. W. Bohrnstedt (2018). Student- teacher racial match and its association with Black student achievement: An exploration using multilevel structural equation modeling. *American Education Research Journal 5*(2), 287–324.

Table 1. Descriptive statistics for teachers and students, 2010-2018[12]

Panel A: Teachers	HBCU-trained	Non-HBCU-trained	Absolute Difference
Race/ethnicity			
Black	75.0	7.10	67.9***
White	22.9	89.2	66.3***
Hispanic	0.88	1.20	0.32
Asian	0.21	0.71	0.50
Female teacher	93.0	92.7	0.3
Attended college out of state	10.4	31.0	20.6***
Graduate degree	44.6	38.6	6.0***
Mean years of experience	13.9	11.9	2.0***
Mean age	42.8	39.6	3.1***
Mean classes taught in year	1.46	1.40	0.06
Mean NTE/Praxis z score	-0.49	0.23	0.72***
Mean share of Black students in school	47.5	27.5	20.0***
N	1,937	19,553	

Panel B: Students	Black	Non-Black	Absolute Difference
Share of total students	23.2	52.7 (White) 16.2 (Hispanic) 2.57 (Asian)	
Female student	50.4	49.2	1.2
Economically disadvantaged	84.0	53.3	30.7***
Disability services-eligible	25.4	28.8	3.4***
English Language-Learner	1.05	14.8	13.3***
Received suspension judgment	20.8	5.7	15.1***
Mean number of days absent	5.39	5.89	0.50***
Retake exam	27.5	13.6	13.9***
Repeat grade	3.95	1.66	2.29***
Number of HBCU-trained teachers			
Zero	70.9	87.9	17.0***
One	21.8	10.2	11.5***
Two or more	7.3	1.90	5.60***
Mean math exam score	-0.35	0.26	0.61***
N	155,272	515,003	

All values, except those associated with experience, age, NTE/Praxis score, number of days absent, and median math test score are expressed as percent. NTE/Praxis scores are standardized by exam year. Both number of days absent and math exam scores are medians and represent the medians across all students after taking the median value within students over time. Math exam scores are standardized by grade and year. *** denotes statistical difference at 0.01 level.

[1] This is for teachers who ever have any Black students in the sample period. Based on unreported tabulations, this group of teachers is not statistically different from the full sample of teachers on any key variables.

[2] Because the status of certain variables varies across academic years, these values represent proportions if a given student ever falls into a given category. For example, a student who is categorized as economically disadvantaged in 3rd grade may not necessarily also be labelled as such in 4th grade. Thus, the table reflects that 84.0% of Black students in the sample are *ever* categorized as economically disadvantaged during their academic tenure.

Table 2. TWFE regression estimates for 3rd, 4th, and 5th grade math scores, HBCU-trained teachers

	(1)	(2)	(3)	(4)	(5)
HBCU-trained teacher	0.034***	0.033***	0.035***	0.033***	0.032***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Student fixed effects	Y	Y	Y	Y	Y
Year fixed effects	Y	Y	Y	Y	Y
Teacher-level controls	N	Y	N	Y	Y
Classroom-level controls	N	N	Y	Y	Y
Additional student-level controls	N	N	N	N	Y
Test score mean	-0.352	-0.352	-0.352	-0.352	-0.352
Test score SD	0.874	0.874	0.874	0.874	0.874
R^2	0.87	0.87	0.87	0.87	0.88
N	370,965	370,965	370,965	370,965	370,965

All regressions include controls for student, teacher, and school characteristics. Baseline model controls for a student's economic disadvantage status, disability status, and English language learner status (all in a given year), teacher race/ethnicity, gender, and age, and school-level proportions of student race/ethnicity, gender, students receiving free or reduced-price lunch, and whether a school was classified as Title 1 fund eligible. Additional teacher-level controls include years of experience and its corresponding square term, median score on NTE/Praxis teacher certification exams, the number of classes a teacher taught during the school year, and indicators for having a graduate degree, whether the teacher attended college out of state, whether their undergraduate college is located in the school district where they teach, and whether their college was historically a normal school. Classroom-level controls include class size and proportion variables for how many students were Black, White, Hispanic, female, economically disadvantaged. Additional student-level controls in Column 5 include repeating a grade, retaking a test, receipt of a suspension judgment, number of days absent during the school year, and whether the student had the teacher in a previous year. Heteroskedasticity-consistent standard errors clustered at the student-teacher level in parentheses. *** $p < 0.01$.

Table 3. TWFE regression estimates of 3rd, 4th, and 5th grade math scores for Black students with interaction term for HBCU attendance and teacher race

	(1) Reference category: Non-HBCU-trained Black teacher	(2) Reference category: Non-HBCU-trained White teacher
HBCU-trained Black teacher	0.026***	
	(0.007)	
HBCU-trained White teacher		0.053***
		(0.011)
Student fixed effects	Y	Y
Year fixed effects	Y	Y
Test score mean	-0.352	-0.352
Test score SD	0.874	0.874
R^2	0.87	0.87
N	370,965	370,965

All regressions control for a student's economic disadvantage status, disability status, and English language learner status (all in a given year), teacher race/ethnicity, gender, and age, years of experience and its corresponding square term, median score on NTE/Praxis teacher certification exams, the number of classes a teacher taught during the school year, and indicators for having a graduate degree, whether the teacher attended college out of state, whether their undergraduate college is located in the school district where they teach, class size, proportion variables for how many students were Black, White, Hispanic, female, economically disadvantaged, school-level proportions of student race/ethnicity, gender, students receiving free or reduced-price lunch, and whether a school was classified as Title 1 fund eligible. Heteroskedasticity-consistent standard errors clustered at the student-teacher level in parentheses. *** $p < 0.01$.

Table 4. TWFE regression estimates for 3rd, 4th, and 5th grade math scores, Black teachers

	(1)	(2)	(3)
All Black teachers	-0.005		
	(0.005)		
HBCU-trained Black teacher		0.013**	
		(0.006)	
Non-HBCU-trained Black teacher			-0.019***
			(0.005)
Student fixed effects	Y	Y	Y
Year fixed effects	Y	Y	Y
Test score mean	-0.352	-0.352	-0.352
Test score SD	0.874	0.874	0.874
R^2	0.87	0.87	0.87
N	370,965	370,965	370,965

All regressions include controls for student, teacher, and school characteristics. Baseline model controls for a student's economic disadvantage status, disability status, and English language learner status (all in a given year), teacher race/ethnicity, gender, and age, and school-level proportions of student race/ethnicity, gender, students receiving free or reduced-price lunch, and whether a school was classified as Title 1 fund eligible. Additional teacher-level controls include years of experience and its corresponding square term, median score on NTE/Praxis teacher certification exams, the number of classes a teacher taught during the school year, and indicators for having a graduate degree, whether the teacher attended college out of state, whether their undergraduate college is located in the school district where they teach, and whether their college was historically a normal school. Classroom-level controls include class size and proportion variables for how many students were Black, White, Hispanic, female, economically disadvantaged. Additional student-level controls in Column 5 include repeating a grade, retaking a test, receipt of a suspension judgment, number of days absent during the school year, and whether the student had the teacher in a previous year. Heteroskedasticity-consistent standard errors clustered at the student-teacher level in parentheses. *** $p < 0.01$.

Table 5. TWFE regression estimates for 3rd,4th, and 5th grade math scores with interactions between HBCU training and teacher experience levels

	(1)	(2)	(3)
HBCU-trained teacher, First year	0.006		
	(0.017)		
HBCU-trained teacher, First two years		0.038***	
		(0.013)	
HBCU-trained teacher, First three years			0.032***
			(0.011)
Student fixed effects	Y	Y	Y
Year fixed effects	Y	Y	Y
Test score mean	-0.352	-0.352	-0.352
Test score SD	0.874	0.874	0.874
R^2	0.87	0.87	0.87
N	370,965	370,965	370,965

The coefficient for each column is relative to a non HBCU-trained teacher with an equal amount of experience. All regressions control for a student's economic disadvantage status, disability status, and English language learner status (all in a given year), teacher race/ethnicity, gender, and age, years of experience and its corresponding square term, median score on NTE/Praxis teacher certification exams, the number of classes a teacher taught during the school year, and indicators for having a graduate degree, whether the teacher attended college out of state, whether their undergraduate college is located in the school district where they teach, class size, proportion variables for how many students were Black, White, Hispanic, female, economically disadvantaged, school-level proportions of student race/ethnicity, gender, students receiving free or reduced-price lunch, and whether a school was classified as Title 1 fund eligible. Heteroskedasticity-consistent standard errors clustered at the student-teacher level in parentheses.*** $p < 0.01$.

Figure 1. Distribution of HBCU-trained teachers across NC school districts, 2015-2016

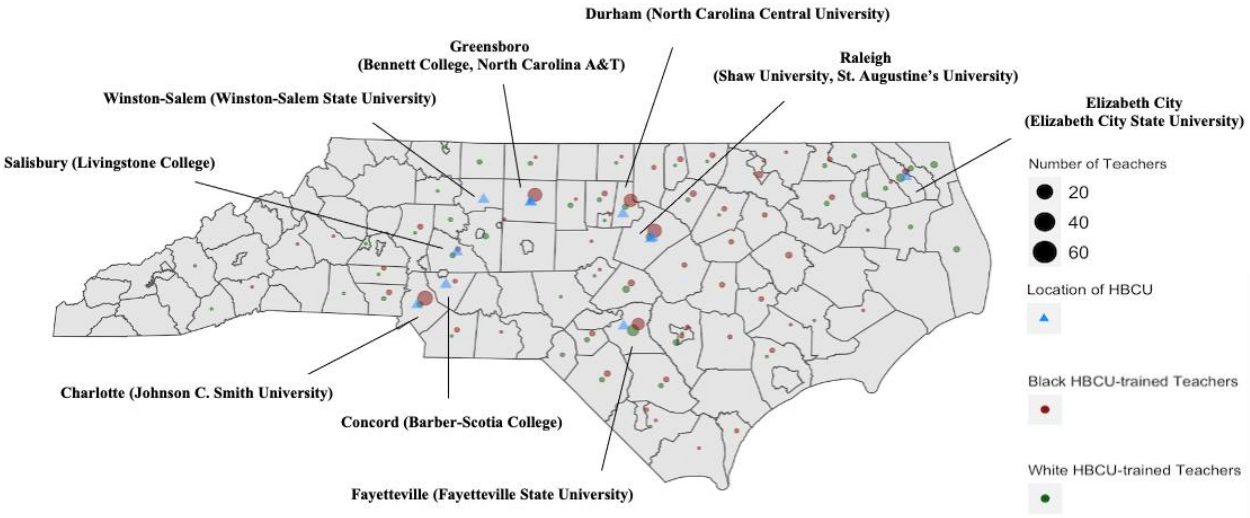
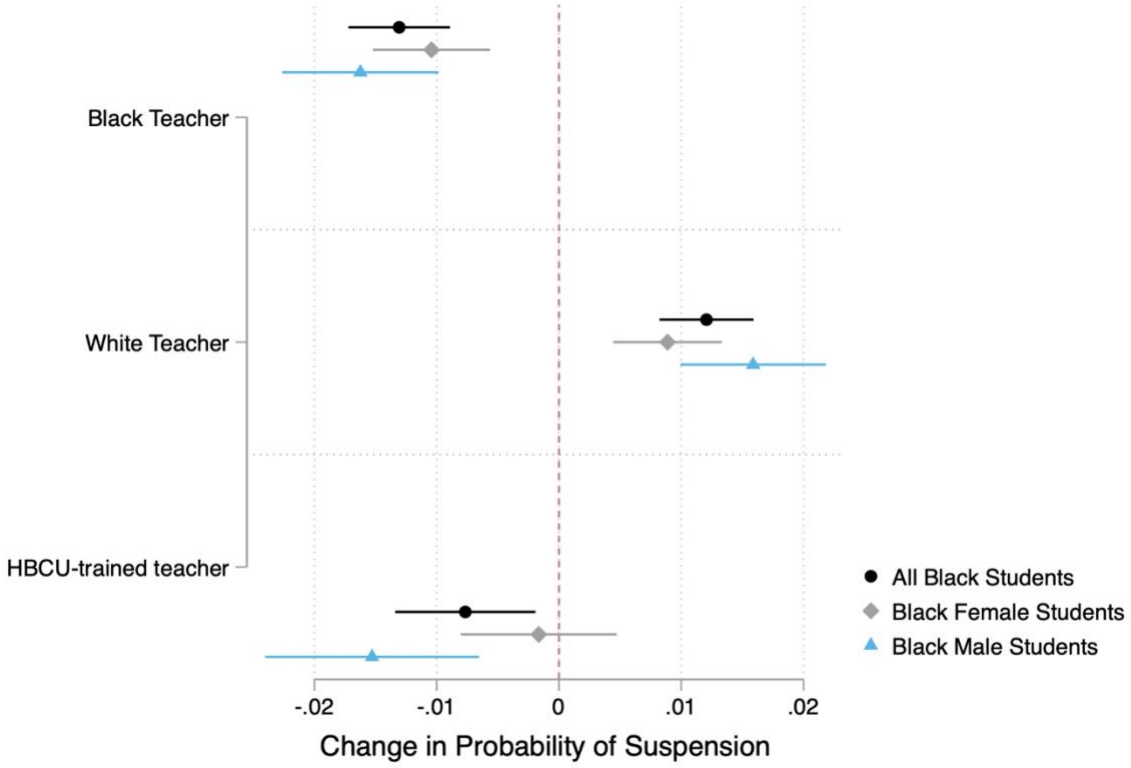
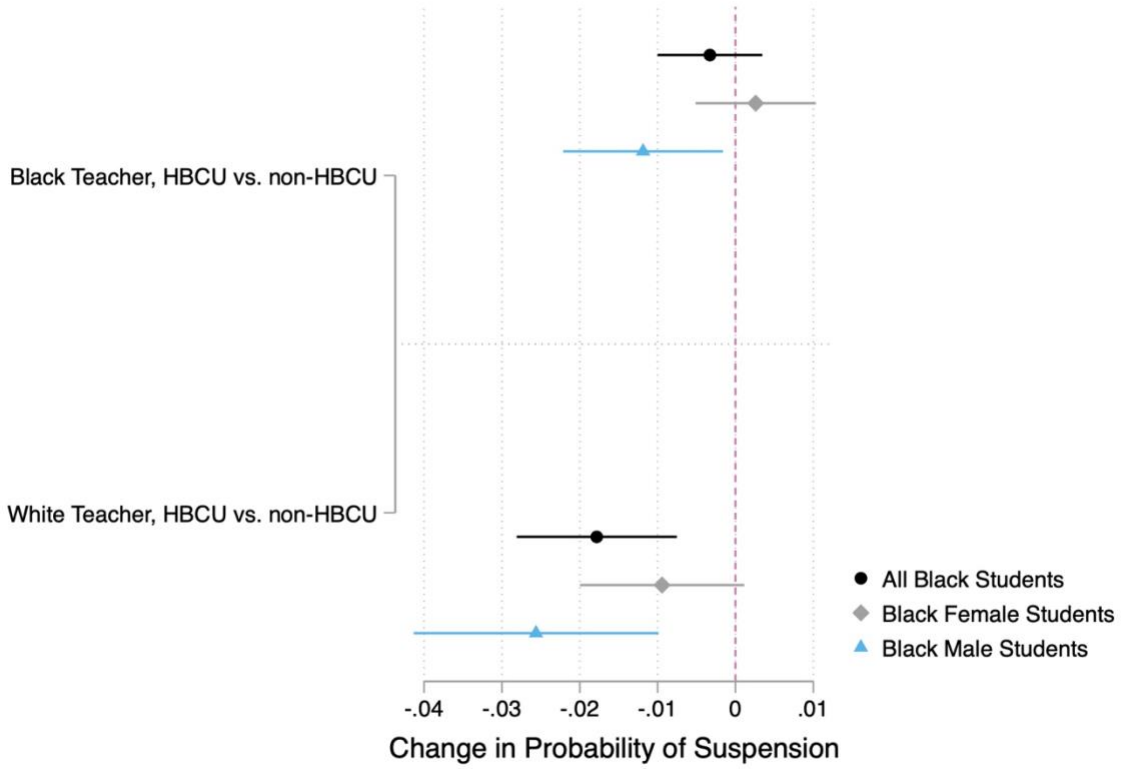


Figure 2. Linear probability model estimates predicting suspension for Black students



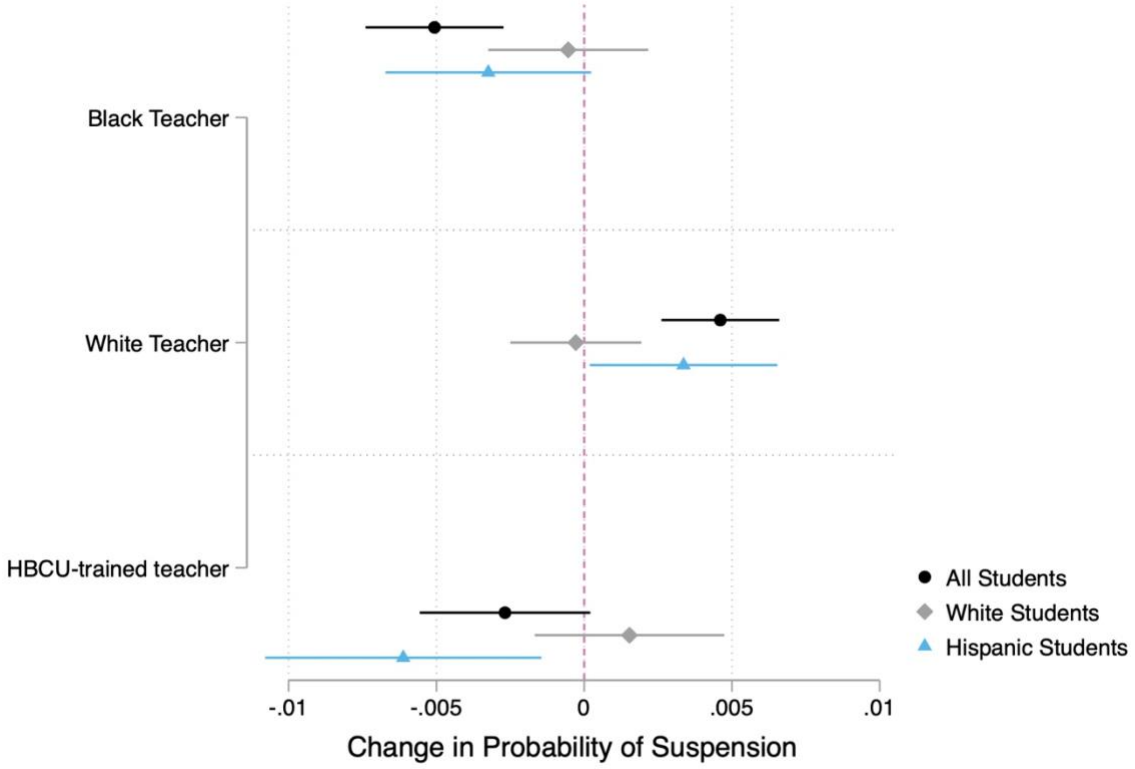
Regressions include controls for student gender, quartile rank of previous year's test score, economic disadvantage status, disability status, English language learner status, received suspension judgment during previous school year, appears in special education classroom, teacher race, teacher gender, teaching experience quartile indicators, age quartile indicators, indicator for teaching multiple classes in a year, majority Black classroom, majority Female students in classroom, majority economically disadvantaged students in classroom, indicator for large class size (greater than 25 students), and grade and school-by-year fixed effects. Zero line corresponds to 90% confidence interval. Heteroskedasticity-consistent standard errors clustered at the student-teacher level.

Figure 3. Linear probability model estimates predicting suspension for Black students, within-race teacher comparisons



Regressions include controls for student gender, quartile rank of previous year's test score, economic disadvantage status, disability status, English language learner status, received suspension judgment during previous school year, appears in special education classroom, teacher race, teacher gender, teaching experience quartile indicators, age quartile indicators, indicator for teaching multiple classes in a year, majority Black classroom, majority Female students in classroom, majority economically disadvantaged students in classroom, indicator for large class size (greater than 25 students), and grade and school-by-year fixed effects. Zero line corresponds to 90% confidence interval. Heteroskedasticity-consistent standard errors clustered at the student-teacher level.

Figure 4. Linear probability model estimates predicting suspension for all, White, and Hispanic students



Regressions include controls for student gender, quartile rank of previous year's test score, economic disadvantage status, disability status, English language learner status, received suspension judgment during previous school year, appears in special education classroom, teacher race, teacher gender, teaching experience quartile indicators, age quartile indicators, indicator for teaching multiple classes in a year, majority Black classroom, majority Female students in classroom, majority economically disadvantaged students in classroom, indicator for large class size (greater than 25 students), and grade and school-by-year fixed effects. Zero line corresponds to 90% confidence interval. Heteroskedasticity-consistent standard errors clustered at the student-teacher level.

APPENDIX

Table A1. Association between teacher assignment and missing/restricted data

	Missing test score	Missing control variables	Missing via sample restrictions
Assigned to HBCU-trained teacher	-0.0003 (0.001)	0.002 (0.001)	-0.001 (0.003)
Student fixed effects	Y	Y	Y
Year fixed effects	Y	Y	Y
R^2	0.76	0.53	0.63
N	651,725	651,725	651,725

All columns represent regressions of indicators for types of missingness on an indicator for whether a Black student was assigned to an HBCU-trained teacher in a given school year. The "Missing test score" regression takes as the dependent variable an indicator for whether an observation was excluded from the analytic sample due to the student having a missing test score in a given year. The "Missing control variables" regression takes as the dependent variable an indicator for whether an observation was excluded from the analytic sample due to having missingness on one of the control variables (as the main regressions in the paper use listwise deletion across all variables included). The "Missing via sample restrictions" regression takes as the dependent variable an indicator for whether an observation was excluded from the analytic sample due to data restrictions described in Section 3.

Table A2. TWFE estimation results for 3rd, 4th, and 5th grade math scores for other demographic groups

	All students	White students	Hispanic students
HBCU-trained teacher	0.012** (0.005)	-0.003 (0.007)	0.009 (0.008)
Student fixed effects	Y	Y	Y
Year fixed effects	Y	Y	Y
Teacher-level controls	Y	Y	Y
Classroom-level controls	Y	Y	Y
Test score mean	0.114	0.370	-0.122
Test score SD	0.956	0.907	0.893
R^2	0.90	0.89	0.88
N	1,626,365	864,887	262,490

All regressions control for a student's economic disadvantage status, disability status, and English language learner status (all in a given year), teacher race/ethnicity, gender, and age, years of experience and its corresponding square term, median score on NTE/Praxis teacher certification exams, the number of classes a teacher taught during the school year, and indicators for having a graduate degree, whether the teacher attended college out of state, whether their undergraduate college is located in the school district where they teach, class size, proportion variables for how many students were Black, White, Hispanic, female, economically disadvantaged, school-level proportions of student race/ethnicity, gender, students receiving free or reduced-price lunch, and whether a school was classified as Title 1 fund eligible. Heteroskedasticity-consistent standard errors clustered at the student-teacher level in parentheses.** $p < 0.05$.

Table A3. Alternative specifications to main TWFE regression

	(1)	(2)	(3)	(4)	(5)	(6)
Sample restricted to students treated in 5th grade	0.050*** (0.010)					
de Chaisemartin and D'Haultfoeuille (2020b) estimator		0.052*** (0.007)				
Sample restricted to students observed in all grades			0.044*** (0.008)			
TWFE with school-by-year fixed effect				0.026*** (0.006)		
Lagged student and classroom math test scores					0.022** (0.010)	
Instrument for HBCU assignment using school-year share of HBCU-trained teachers						0.039* (0.022)
Student fixed effects	Y	Y	Y	Y	N	Y
Year fixed effects	Y	Y	Y	N	N	Y
School-by-year fixed effects	N	N	N	Y	Y	N
Grade fixed effects	N	N	N	N	Y	N
Test score mean	-0.329		-0.344	-0.352	-0.358	-0.352
Test score SD	0.876		0.866	0.874	0.868	0.874
R^2	0.85		0.85	0.89	0.69	0.89
N	270,176	182,964	173,149	370,075	247,721	370,964

All regressions control for a student's economic disadvantage status, disability status, and English language learner status (all in a given year), teacher race/ethnicity, gender, and age, years of experience and its corresponding square term, median score on NTE/Praxis teacher certification exams, the number of classes a teacher taught during the school year, and indicators for having a graduate degree, whether the teacher attended college out of state, whether their undergraduate college is located in the school district where they teach, class size, proportion variables for how many students were Black, White, Hispanic, female, economically disadvantaged, school-level proportions of student race/ethnicity, gender, students receiving free or reduced-price lunch, and whether a school was classified as Title 1 fund eligible. Given the computation of the de Chaisemartin and D'Haultfoeuille (2020b) estimator, test score mean, test score SD, and R^2 are unknown and omitted from this table. This estimator's standard errors were computed using 500 bootstrap replications. he lagged test score specification includes controls for a given student's prior year end-of-grade math score and the average end-of-grade math score for all other students in class with that student. The coefficient for the first stage of the Bettinger and Long (2005) IV specification is 0.007 (SE 0.0002; F-statistic 1,194.32). Heteroskedasticity-consistent standard errors clustered at the student-teacher level in parentheses for column 1 and columns 3-7. Heteroskedasticity-consistent bootstrapped standard errors clustered at the student level in parentheses in column 2. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A4. Admissions and SAT statistics for HBCUs and non-HBCUs

	HBCU		Non-HBCU		Absolute Difference in Means
	Mean	SD	Mean	SD	
% Admitted, 2001-2013	54.9	14.7	67.8	15.6	12.9***
SAT Math 25th percentile, 2001-2013	388.5	41.9	485.4	62.3	96.9***
SAT Math 75th percentile, 2001-2013	490.5	53.6	594.9	58.8	104.4***
SAT Reading 25th percentile, 2001-2013	392.9	42.7	479.2	59.4	86.3***
SAT Reading 75th percentile, 2001-2013	494.0	57.4	589.4	59.0	95.4***
N	50		831		

*** denotes statistical difference at the 0.01 level.