Which Match Matters? HBCUs, Same-Race Teacher Effects, and Black Student Achievement*

Lavar Edmonds
Stanford University

This Draft: July 2024

Abstract

In many settings, researchers study racial concordance effects on economic and social outcomes. This paper presents new evidence on potential channels underlying racial match effects by studying the impacts of teachers trained at Historically Black Colleges and Universities (HBCUs) on Black student outcomes. Using data from North Carolina elementary schools, I find Black students score higher on end-of-grade math exams, and are less likely to be suspended, when assigned to an HBCU-trained teacher. I show that both Black and White HBCU-trained teachers are more effective with Black students than their same-race, non-HBCU peers are. Results are consistent with HBCU effects stemming from different teacher education practices across colleges. Together, this work implies a role for training and learning in affecting outcomes beyond racial (mis)match. JEL Codes: I21, I23, I24, J15

*Contact: ledmonds@stanford.edu. A previous version of this manuscript was circulated under the title "Role Models Revisited: HBCUs, Same-Race Teacher Effects, and Black Student Achievement." Thanks to Kara Bonneau and the North Carolina Education Research Data Center for providing data and administrative assistance. 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, Brian Cadena, Janet Carlson, Tom Dee, Ellora Derenoncourt, Matthew Desmond, Michael Gottfried, Peter Hinrichs, Christopher Neilson, Isaac Sorkin, Karolyn Tyson, Jon Valant, Anne Kat Alexander, Joe Fish, Jennifer Kowalski, Lillian Leung, Jacob Light, Renee Louis, Brendan Moore, João Pugliese, and China Mae Stepter for insightful comments and suggestions on previous proposals and drafts of the paper. I also thank seminar and conference participants from the Princeton Education Research Section, the Eviction Lab, the Economics of Racism, the Urban Institute, the NBER Race and Stratification working group, and the W.E. Upjohn Institute for Employment Research for their helpful feedback. All errors – and there will be some – are all my own.

1 Introduction

Racial representation and concordance studies across the social sciences explore 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 (Abrams, Bertrand, and Mullainathan, 2012; Steffensmeier and Britt, 2001), police-civilian interactions and criminal justice reform (Ba et al., 2021; Headley and Wright, 2020; Hoekstra and Sloan, 2022), 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 behavior between individuals (e.g., Arnold, Dobbie, and Yang, 2018; Rose, 2021). Education research has similarly proposed that students benefit from exposure to demographically-congruent teachers (Bartanen and Grissom, 2021; Dee, 2004, 2005; Delhommer, 2022; Egalite, Kisida, and Winters, 2015; Gershenson et al., 2022; Holt and Gershenson, 2019; Klopfenstein, 2005; Penney, 2017), especially racial and ethnic minoritized students (Villegas and Lucas, 2004). However, much of this work on racial representation focuses on whether race matters for outcomes without demonstrating how or why race matters, leaving the racial match effect mechanisms largely unaddressed.

In this paper, I revisit same-race teacher effects and their mechanisms through an analysis of Historically Black Colleges and Universities (HBCUs). Specifically, I address two questions. First, what is the impact of having an HBCU-trained teacher on Black student outcomes? Second, how does HBCU attendance interact with same-race teacher effects for Black students? HBCUs are an appropriate context to study (racial) match effects 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). They also train a considerable number of White teachers – in this sample, nearly one-quarter of HBCU-trained teachers are White – providing an observable environment where trainees are exposed to Black students and professors, and potentially learned skills can be measured after their matriculation into K-12 classrooms.

I study HBCU attendance and teacher quality by estimating teacher effects on end-of-grade math and English Language Arts (ELA) exams using administrative panel data from elementary schools in North Carolina, producing four important results. First, I find HBCU-trained teachers have a positive effect on Black student achievement. Different specifications indicate that, on average, Black students score 0.022-0.042 standard deviations (SD) higher on standardized math exams in the year they are assigned to an HBCU-trained teacher. Effects are substantively large: Black student assignment to HBCU-trained teachers reduces their within-school conditional average Black-White test score gap in gains and in levels by 27% and 7%, and reduces their within-classroom conditional average Black-White test score gap in gains and in levels by 11% and 6%, respectively. Estimated magnitudes are at least comparable to those reported in previous papers on same-race teacher matches. Though effects are strongest for previously lower-scoring students, I also find positive effects of HBCU-trained teacher assignment for previously high-achieving Black students. Results are robust to alternative specifications that address biases that can arise in TWFE regressions with differential timing in teacher assignment. While the paper focuses on Black student outcomes, it is noteworthy that I do not observe significant negative effects for non-Black students, with even some evidence of positive effects for Hispanic students. I also present results from corresponding analyses for ELA test scores. Similar to several prior studies that show null teacher effects on reading scores, I find positive but small effects (0.01 SD) that are more sensitive to model specification.

Second, I find little evidence that Black students in classrooms with Black teachers generally score better on standardized exams, with the notable exception of the positive effect on test scores for Black students with Black teachers who attended an HBCU. Instead, I find HBCU-driven variation in teacher quality within teacher race: Black students score higher on math exams when paired with Black HBCU-trained teachers, compared against Black teachers from non-HBCUs. This result extends to Black students matched with White HBCU-trained teachers (compared against White teachers from other colleges), implying the existence of race-independent, training-based teacher characteristics underlying match effects that improve student outcomes.

Third, I show that the HBCU-trained teacher effect extends to behavioral measures (namely, student suspensions). Black students assigned to an HBCU-trained teacher have a lower probability of being suspended than their Black peers in other classrooms. Notably, the association is largely determined by impacts for Black boys. Additionally, whereas there are no significant differences in suspension probabilities conditional on having an HBCU-trained Black teacher compared to a non-HBCU-trained Black teacher, White teachers from HBCUs are appreciably less likely to suspend Black students than their White counterparts from non-HBCUs.

Finally, I show suggestive evidence that these results are consistent with observed teacher effectiveness reflecting pedagogical training and skills developed at HBCUs, rather than simply a downstream effect for the type of student who elects to attend a Historically Black College. Using college major and teacher certification data available for over one-third of the teachers in the sample, I highlight that HBCU-trained teachers with majors in education-related fields are more effective with Black students than both their peer education graduates from other institutions and HBCU-trained teachers with majors in other, non-education fields. In particular, I observe the largest impacts for White teachers who attend HBCUs and major in education-related fields.

In summary, this paper makes two key contributions to prior literature. Principally, this research has implications for interpreting effects from same-race pairings. The presented evidence proposes that positive outcomes for racial minorities matched to a same-race individual reflect a bundle of (learnable) skills, rather than inherent attributes determined by race. It also proposes a simple mechanism for how such skills may be acquired and developed by different-race individuals: higher education training. To the extent that developing these relevant, human capital-boosting skills can be replicated at other training institutions (e.g., medical school, law school, etc.), these findings pose a promising area for future practitioner training and development policies. Specific to education research and policy, this paper also advances literatures on teacher preparation in higher education (e.g., Goldhaber et al., 2013; Koedel et al., 2015; Kumar and Lauermann, 2018), teacher value-added (Bacher-Hicks and Koedel, 2023; Jackson et al., 2014), and teacher recruitment and retention (Bartanen and Grissom, 2021; Ingersoll et al., 2017). In doing so, it offers new insights for addressing persistent racial academic achievement gaps driven by resource inequities, particularly the unequal distribution of high-quality teachers for racial/ethnic minority and low-income students (Clotfelter et al., 2023; Goldhaber et al., 2018; Jackson, 2009; James and Wyckoff, 2022).

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 Black pedagogical training. Section 3 outlines the data used, and Section 4 details the econometric models 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 highlight and empirically test 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 and Mawhinney, 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).

Although they are important 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," institutions created to prepare teachers for the classroom (Akbar and Sims, 2008). Today, HBCUs carry the legacy of a period where Black teachers, principals, and colleges worked explicitly in concert to 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 comprises decades of intentional teaching practices 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.

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), while HBCUs produce 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] Further, 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).

Despite being historical and contemporary powerhouses for Black teacher training, HBCUs have largely been unexplored in economics. To date, extant economic research on HBCUs has focused generally on finance and analyses of college funding (Sav, 2010), accountability and effects on employment outcomes (Price, Spriggs, and Swinton, 2011), effects on wages (Mykerezi and Mills, 2008), and collegiate outcomes (e.g., likelihood of graduation) (Wilson, 2007). 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 minoritized 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, to the best of my knowledge, this is the first to explicitly study HBCUs and how preservice teacher education at these institutions may directly affect student academic outcomes.

2.1 Alternative explanation for same-race teacher effects: teacher training

Teacher training reveals a new lever for understanding same-race teacher effects beyond those previously discussed in economics research, such as role model effects, stereotype threat, and teacher expectations. In the first case, Black teacher success with Black students may stem from exposure to educated, successful Black adults in the classroom, updating their beliefs on returns to their own academic achievement.[2] In the second, same-race teachers may alleviate the student's anxiety of having their performance be an indicator of and benchmark for their racial group's (negatively) perceived intellect. Finally, different-race teachers with lower academic expectations

[1] Based on the author's calculations.

[2] For further discussion and critiques of the role model effects hypothesis, see Harper and Davis (2012), Irvine (1989, 1990), Maylor (2009), and Villegas and Lucas (2004).

of their students and implicit biases may adversely influence student effort and educational attainment.[3]

While I leave open the possibility that these theories suffice in some settings, ultimately they fall short in the effort to fully explain same-race teacher effects. Extant same-race teacher research, as with the broader same-race match literature, poses little evidence on underlying mechanisms, in part because it is often reliant on race alone as the defining feature. Being Black does not automatically enable teachers to positively affect Black student outcomes. Even if a teacher possesses cultural knowledge from which to draw for instruction, there is nothing innate about race that guarantees one has the capacity and language to do so.[4] Alternatively, there need be some impetus, like formal teacher education, for thinking about and structuring instruction in this way. As such, a more suitable framework for understanding same-race teacher effects would account for diversity in teacher training and perspective, the guiding principle of this research.

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 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 schools they attended (demographic counts). For teachers, data include demographic characteristics (race, gender, and age), preservice history (colleges attended, level of education, certification test scores), and years of experience teaching in North Carolina.

As in prior studies using North Carolina education data (e.g., Bates et al., 2023; Hill and Jones, 2021; Jackson, 2018), I use a course membership file to link students

[3] In unreported regressions, I adopt the test for racial evaluation bias used in Rangel and Shi (Forthcoming) and Nguyen et al. (2023) and find no systematic differences between HBCU-trained teachers and non-HBCU-trained teachers in how they subjectively predict student performance on standardized math exams.

[4] Other scholars have made a similar argument. Cherry-McDaniel (2019) cautions against assuming "teachers of color are inherently capable of teaching in culturally responsive and sustaining ways" and posits that that assumption "has led to an overall uncritical approach to diversity training in teacher preparation programs" (pgs. 249-250).

to teachers who taught specific courses during a year. To provide greater reliability to estimates, I impose several sample restrictions, including dropping non-traditional schools (e.g., charters, magnets, alternative schools), restricting to students for whom I observe at least two years of test scores, removing classrooms with fewer than 10 or greater than 50 students, and removing classrooms where greater than 50% of students are designated recipients of some special accommodations (i.e., gifted, disabilities, or English language learners). Using this approach, I compile a sample of 247,288 unique Black students (and 1,091,721 unique students total) taught by 27,501 unique teachers over the 11-year period, covering 539,423 Black student-year observations in math and ELA.[5]

3.1 Descriptive statistics

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, a large gap exists 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.

I also compare Black students disaggregated by if they ever have an HBCU-trained

[5]See the online data appendix for additional details on sample construction and the matching procedure.

teacher (Panel B). Black students with at least one HBCU-trained teacher and those who never have one are similar on a few dimensions, though there is also some evidence of negative selection into HBCU-trained teacher classrooms (e.g., these students have considerably lower average test scores and are more likely to have received a suspension judgment).

4 Empirical Strategy

4.1 Modeling student achievement

To begin, I estimate teacher effects using two, linear and additively separable regression specifications akin to the one-step value-added model commonly used in education research (Bacher-Hicks and Koedel, 2023). First, I employ a two-way fixed effects estimator (TWFE) that relies on within-student variation in teacher assignment over time.[6] 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. The corresponding regression equation is presented in Equation (1).

$$A_{it} = \beta(\text{HBCU}_{it}) + X'_{ijst}\gamma + \delta_i + \eta_t + \lambda_s + \epsilon_{it}. \quad (1)$$

The dependent variable is the score on an end-of-grade exam (with separate equations estimated for math and ELA), standardized within grade and year to mean 0 and standard deviation 1, for a student i in year t. I regress this test score on a vector X of time-varying student, teacher (j), and school (s) characteristics, as well as student (δ_i), year (η_t), and school (λ_s) 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. Following guidance from Abadie et al. (2023), I cluster

[6] Standard two-way fixed effects regressions include fixed effects for unit and time levels. Given the schooling context and substantive variation among students and teachers across schools, I augment the traditional specification with school fixed effects.

standard errors at the classroom level, the level of treatment assignment (all students in the same classroom are assigned the same treatment status).[7]

Second, as an alternative to within-student comparisons, I also fit a regression using lagged achievement, comparing outcomes across students assigned to different teachers. Empirically, this takes the form

$$A_{it} = \beta(\text{HBCU}_{it}) + \Phi f(A_{it-1}) + \Omega f(A_{-i(t-1)}) + X'_{ijst}\gamma + \theta_g + \eta_{st} + \epsilon_{it}. \quad (2)$$

A few differences separate this model from the TWFE approach in Equation (1). Most importantly, instead of student fixed effects, I condition on a nonlinear (cubic polynomial) function of lagged test scores in math or ELA, as well as a nonlinear function of lagged leave-out classroom test scores (i.e., the average of all other students' prior test scores) — Φ and Ω are vectors of parameters that correspond with each term of the functions. I also include grade fixed effects and additional student-level controls (namely, student gender, age, lagged student absences, an indicator for whether the student received a suspension in the prior year, and an indicator for whether the student previously had their current year's teacher). The remaining variables and standard error clustering are the same as Equation (1). While lagged achievement regressions are relatively popular in the literature, a drawback in this setting is it necessarily requires student test score availability in the prior year, which is only feasible in 4th and 5th grade (end-of-grade testing begins in 3rd grade in North Carolina), thus removing a considerable amount of the students from the sample.

4.2 Alternative estimator to address variation in treatment timing

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 et al., Forthcoming; Callaway and Sant'Anna, 2020; de Chaisemartin and D'Haultfœuille, 2020; 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

[7] Results are robust to clustering at the teacher-year level and two-way clustering at the student and teacher levels.

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 across three years, 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) explicates, 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 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 (2020) and calculate the weights associated with TWFE regressions that would bias estimated coefficients. 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. The "treatment" group comprises Black students who are assigned an HBCU-trained teacher exactly once (in the 5th grade), and the "control" group 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 (2020) that, by construction, is robust to heterogeneous treatment effects, differential timing in treatment, and dynamic effects.

5 Results

In this section, I present empirical evidence of HBCU-trained teacher positive effects on Black student math test scores. I first argue the OLS results I present warrant a causal interpretation, following evidence from a quasi-experimental design adapted from Chetty et al. (2014) (Section 5.1). The main findings for math test scores, discussed in Section 5.2, are invariant to controls for observable teacher and classroom characteristics, other modeling strategies, and different sample restrictions (Section 5.3). I demonstrate these teachers help to mitigate the Black-White test score gap (Section 5.4) and have effects on students across the distribution of student achievement, rather than just improving the test scores of previously lower-scoring students (Section 5.5). HBCU-trained teacher effects are independent of racial match, as I observe within-teacher race differences in effectiveness conditional on whether a teacher attended an HBCU (Section 5.6). Lastly, I extend the analysis to non-testing outcomes by showing HBCU-trained teachers are also associated with a lower probability of suspension for Black students (Section 5.7). Suspension results, too, imply race-independent effects on test scores for teachers trained at HBCUs.

5.1 Identifying assumption and causal interpretation

Causal identification in this context assumes conditional random assignment of students to teachers. In practice, students are not randomly assigned to teachers, which raises the concern of higher achieving students sorting to certain teachers and the econometrician misinterpreting improved test scores as higher value-added teachers. However, conditional random assignment holds if, given some observable set of characteristics, there is quasi-random variation in student allocation across different teachers. Following Jackson (2018), formally this means

$$\mathbb{E}[\phi_{it-1} \mid \text{HBCU}_{it}, \chi_{ist}] = \mathbb{E}[\phi_{it-1} \mid \chi_{ist}]. \tag{3}$$

where ϕ_{it-1} is a student i's incoming academic ability (i.e., their stock of ability at the end of the year preceding assignment to an HBCU-trained teacher) and χ_{ist} is a vector that holds observable student, classroom, and school characteristics, including student, year, and school fixed effects. In expectation, conditional on χ_{ist}, knowing a student's assignment to an HBCU-trained teacher provides no additional information about a student's incoming academic ability. I argue the assumption holds here, for a couple of reasons. First, Kane and Staiger (2008), Kane et al. (2013), and Bacher-Hicks et al. (2019) compare teacher value-added estimates between observational and experimental data and show the former provides unbiased predictors of teacher effects on student achievement when controlling for prior student and peer achievement. This is echoed in Chetty et al. (2014), who, using observational data and quasi-experimental methods, posit "controlling for prior student-level test scores is the key to obtaining unbiased value-added estimates" (pg. 2629). I satisfy this requirement with estimates in columns 4, 5, and 6 of Table 2, which present evidence of HBCU-trained teacher effects using regressions that condition on cubic polynomial functions of prior student and peer (classroom leave-student-out) achievement.

Second, as an additional validity check against concerns about nonrandom student sorting on unobservables, I assess a simple modification to the quasi-experimental approach developed by Chetty et al. (2014) that evaluates forecast bias in teacher value-added by exploiting quasi-random variation in teacher mobility. Applied to this setting, the logic is as follows: different school-grade-year groups of Black students vary in their exposure to HBCU-trained teachers, variation induced by plausibly exogenous mobility of teachers across grades and schools. In this 11-year sample, 20.9% of HBCU-trained teachers change schools, and, even among teachers who stay in the same school, 19.5% switch to a different grade. Assuming teacher moves are independent of changes in student or school quality, if HBCU-trained teachers have real predictive effects on Black students, then this would be picked up in the test scores for cohorts that have more or fewer HBCU-trained teachers due to teacher entry or exit into the school-grade over time. Put differently: if HBCU-trained teacher effects merely reflect the sorting of higher achieving students, this would presumably impact only the distribution of scores among Black students, not the average score across all Black students.

Following this, I fit a regression that projects the yearly change in mean Black student test score $\left(\Delta A_{sgt}^{\text{Black}}\right)$ across school-grade cells (i.e., the school-grade cohort's average Black student test score in year t minus the average Black student test score for the prior school-grade cohort in year t-1) onto the change in the cohort mean share of HBCU-trained teachers ($\Delta(\text{HBCU Share}_{sgt})$) and school-by-year fixed effects, weighted by the number of Black students in a given school-grade-year. The baseline model takes the form

$$\Delta A_{sgt}^{\text{Black}} = b\Delta(\text{HBCU Share}_{sgt}) + \eta_{st} + \epsilon_{sgt}. \tag{4}$$

Results are presented in Online Appendix Table A3. Specifications vary with the inclusion of controls for other changes in average teacher characteristics at the school-grade-year level (namely, average experience, average teacher certification exam score, share of teachers with a graduate degree, share of Black teachers, share of White, Hispanic, Asian, and female teachers). I also include a cubic polynomial in the lagged average Black student ELA test score. Ultimately, across different specifications, I reach the same result: over time, across different cohorts, when the share of HBCU-trained teachers in a school-grade-year increases, so too does the change in the cohort's average Black student test score. Since this is essentially a test score gains specification, in columns 3 and 4, I consider more familiar (and easily interpretable) regressions in levels where I amend Equation (7) such that the outcome is simply average Black student test scores for the cohort (with cubic polynomials of lagged scores in math and ELA as controls). Results hold the same interpretation: cohorts with an increased share of HBCU-trained teachers have higher average scores on math exams. Overall, this exercise provides a credible check that the estimated HBCU-trained teacher effects I duscuss further in subsequent sections capture causal impacts on Black student test scores.

5.2 Baseline estimates: HBCU-trained teacher effects on Black student test scores

To begin, I recover baseline results by estimating Equation (1) for all Black students using ordinary least squares estimation. Regression results appear in Table 2, suppressed to the HBCU attendance variable. Estimates vary from column to column based on different model specifications (student fixed effects or lagged test scores, and the included controls). Column 1 shows the relationship between having an HBCU-trained teacher and math/ELA scores in the baseline model. Test scores are in standard deviation units, meaning this coefficient reflects that a Black student assigned to an HBCU-trained teacher, on average, scores approximately 0.030 standard deviations (SD) higher on their end-of-grade math exam and 0.012 SD higher on their end-of-grade ELA exam than they score in a different year when assigned to a non-HBCU-trained teacher. The math result in particular is comparable to estimates in prior literature estimating same-race teacher match effects on test scores in North Carolina (0.020 SD; Clotfelter, Ladd, and Vigdor, 2007), and Black teacher-Black student match effects reported in Missouri and Tennessee (0.021 SD; Bartanen and Grissom, 2021), Florida (0.030 SD; Egalite et al., 2015), and North Carolina (0.035 SD; Goldhaber and Hansen, 2010).

Notably, HBCU-trained teacher effectiveness holds even after accounting for other observable measures of teacher background. Column 2 extends the baseline model to include controls for teacher and classroom characteristics. For teachers, these comprise their level of experience[8], graduate degree receipt, receipt of a graduate degree from an HBCU, teacher licensure test score[9], the number of classes the teacher taught in the 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. For classrooms, these include classroom-level aggregates of student race/ethnicity, gender, economic disadvantage, disability services eligibility, limited English proficiency designation, and class size.[10] The estimated math effect is invariant to these controls (in fact increasing in magnitude), indicating the estimated HBCU teacher effect on math scores is not driven by HBCU-trained teachers' relatively higher average teaching experience, 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, any general comparative advantage from many HBCUs' histories specializing in teacher education, or the composition of a student's peers. In the ELA equation, the magnitude declines by one-third (0.012 to 0.009), though it remains statistically significant.

In column 3, I extend the regression models with additional student-level controls that may impact test scores, such as whether the student was suspended during the school year, their number of days absent from school, student gifted status, teacher National Board Certification status, and leave-student-out classroom average for suspensions. These variables do not substantially change the estimates presented in previous columns. They do, however, introduce additional endogeneity concerns, as these variables are post-treatment (i.e., teacher assignment) outcomes and may also be influenced by a student's teacher. Thus, they are omitted from subsequent regressions.

Finally, in column 4, I consider the lagged achievement models outlined in Equation (2), as well as extensions where I also control for cubic polynomials of lagged student and peer scores in the other subject (e.g., controlling for ELA scores in the math regression) (in column 5) and control for twice-lagged scores (in column 6). The estimated math effect attenuates, but remains large and statistically significant

[8]Regression includes both experience and a quadratic term, to capture possible nonlinearities in the returns to teacher experience.

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

[10]To account for possible nonlinearities in the relationship between test scores and classroom composition, I include indicators for different classroom-level demographic thresholds (i.e., the classroom contains at least 25%, 50%, or 75% of each student demographic group, with less than 25% as the omitted reference category.

at the 1% and 5% levels. The ELA effect, conversely, shrinks further and is imprecisely estimated.

Taking stock of differing results for math and ELA: that I find inconsistent evidence of teacher effects on ELA test scores is relatively common across research in education. Proposed explanations include that a disproportionate share of learning in reading takes place at home rather than in school, so school/teacher effects are muted (Jackson et al., 2014), and that state accountability plans and test designs incentivize different levels of test preparation in the classroom, such that teachers put in varying levels of effort to math compared to ELA (Riehl and Welch, 2022). All the same, for parsimony, I focus on math teachers and math test score effects throughout the remainder of the paper using variations on the specification in column 2.[11]

5.3 Robustness checks

To probe sensitivity of the estimated HBCU effects, I explore alternative specifications in Online Appendix Table A1. Following the discussion in Section 4.2., 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 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, I implement the estimator derived in de Chaisemartin and D'Haultfœuille (2020b) that is robust to the differential treatment timing concern.[12] 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.

I also consider a variety of specifications based on sample restrictions using my preferred regression specification (column 2 from Table 2). First, I drop student observations from atypical schooling occurrences: students who, within a school year, appear in multiple classrooms, transfer schools, and or are assigned to multiple teach-

[11] Subsequent analyses for ELA test scores are available upon request to the author.

[12] 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 13% of the weights were negative, summing to approximately -0.041, further indicating negative weights are unlikely to be an empirical issue in this setting.

ers (column 3 of Online Appendix Table A1).[13] Second, to examine whether post-undergraduate experiences confound the "true" HBCU attendance signal (e.g., if any relevant teaching skills are developed post-HBCU undergraduate attendance), I focus on teachers who only hold a bachelor's degree (column 4). Third, I check whether effects are driven by a few "superstar" HBCUs in North Carolina by only comparing out-of-state teachers (column 5). Fourth, I estimate effects over the "full information" sample of students for whom I observe all three grades, probing any effects of differential student attrition from the sample and changing student composition, which may bias results if I systematically observe students in atypical years of their schooling (column 6). Fifth, in column 7, I reduce the sample based on the number of Black students in a classroom. A large share of teachers in any given year only teaches one Black student (e.g., across all teacher-year observations, the 25th percentile in the distribution of Black students taught is one, and the median is four). Teaching so few Black students may confound the reliability of these estimates as a signal of teacher effectiveness with Black students (Kim, 2022; Staiger and Rockoff, 2010). To address this, I limit the sample to teachers who teach at least three Black students in a given school year. Finally, column 8 shows estimates from a regression that only includes teachers observed in the sample for three or more years, following caution raised in Koedel and Betts (2011) about using the first couple of years of teacher experience to make inferences about teacher quality. This specification has the added benefit that sorting bias can be mitigated by restricting to teachers with multiple classrooms of students over time. None of these restrictions substantially alters the estimated effects. In short, across a battery of alternative specifications, I find consistent evidence of positive HBCU-trained teacher effects on Black student math test scores.

5.4 HBCU-trained teachers and the Black-White test score gap

Another way of understanding the size of HBCU-trained teacher effects is to place them in terms of the persistent Black-White test score gap (Anderson, 2007; Ladson-Billings, 2006). To do so, I estimate a regression equation across the sample of all students that amends Equation (2) to include interactions between indicators for student race/ethnicity (with White students coded as the reference category) and assignment to an HBCU-trained teacher. This equation takes the form

[13] These were previously kept for precision, and prior regressions include controls for each of these less common occurrences (10% of observations in the analytic sample).

$$A_{it} = \beta_1(\text{HBCU}_{it}) + \sum_{k=2}^{5} \beta_k(\text{Student Race}_{im}) + \sum_{l=6}^{9} \beta_l(\text{HBCU}_{it} * \text{Student Race}_{im})$$
$$+ \Phi f(A_{it-1}) + \Omega f(A_{-i(t-1)}) + X'_{ijst}\gamma + \theta_g + \eta_{st} + \epsilon_{it}. \quad (5)$$

For students i belonging to a race/ethnicity group m $\in$ {Black, Hispanic, Asian, Other}, there are race/ethnicity indicator coefficients (β_k) that report the conditional average test score difference between students belonging to that group and White students.[14] Each interaction term's coefficient (β_l) expresses the change in that gap when students are assigned to an HBCU-trained teacher. With school-by-year fixed effects, the estimated changes in the gaps are identified by students of the same race in the same school in the same year who vary in their assignment to teachers from HBCUs. I display estimates from Equation (3) in Table 3. Note that I define the gap in two ways. First, in column 1, with the inclusion of lagged test scores for students, I capture differences in the average test score gap in test score gains, similar to Hill and Jones (2021). Second, by omitting prior test scores (and prior average peer test scores), I show differences in the raw test score gaps between students, or in levels. Columns 3 and 4 duplicate these regressions, but replace school-by-year and grade fixed effects with classroom fixed effects, showing the change in the minority/majority test score gap among students in the same classroom. The coefficients capture different effects on the test score gap between classes taught by HBCU-trained teachers and classes with other teachers.

Various specifications produce a consistent result: HBCU-trained teachers have substantial impacts on the Black-White test score gap. Comparing the main effects for the Black student term to its interaction with HBCU teachers, estimates show the average within-school Black-White test score gap (-0.103 SD after controlling for prior achievement and -0.47 SD in raw levels) is reduced by 27% (in gains) and 7% (in levels) when Black students are assigned to HBCU-trained teachers. The effects on within-classroom Black-White test score gap are smaller, but still considerable: 11% (in gains) and 6% (in levels). These results also suggest positive effects for Hispanic students, contributing to a narrowing of their (albeit smaller) test score gap with White students. Further, given the coefficient on the HBCU teacher assignment variable, which in this setting denotes the effects on White students, is nearly zero and

[14] Because of a limited sample size and how race/ethnicity categories are coded in the raw data, all non-White/Black/Hispanic/Asian students are coded as "other race/ethnicity."

statistically insignificant (as is similarly the case for Asian and other race students), these results suggest no negative test score effects on non-Black and Hispanic students assigned to HBCU-trained teachers.[15]

5.5 Distributional effects on Black student test scores

In Online Appendix Table A2, I split the sample into students based on prior academic achievement. I construct quartiles of prior year math test scores within grade and year and run separate regressions for students who are lower scoring (i.e., in the bottom two quartiles of the prior year's performance distribution across all same grade-year test takers in the state) and higher scoring (in the upper two quartiles). Unsurprisingly, the largest effects in magnitude are for students in the bottom half of the prior achievement distribution — these are students with very low test scores who mechanically had the largest gains to be made.[16] I do, however, still see positive, significant effects for previously high achieving students, indicating HBCU-trained teachers can benefit *all* Black students, regardless of incoming academic preparedness.

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

Building on the same-race teacher effects literature, in this section I show that a given HBCU-trained teacher is not only more effective than their non-HBCU-trained counterpart, but also that this dynamic holds even when the two teachers are the same race. Notably, I also find little evidence of a same-race teacher effect more generally for Black students, except in the case of Black teachers who attended an HBCU.

A straightforward explanation for HBCU effects 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. Given all prior regressions control for teacher race, empirically this seems unlikely from the start. Nevertheless, I more explicitly draw further distinction between these channels and strengthen the evidence of HBCU teacher effects in two ways. First, I disaggregate the analysis of HBCU-teacher effectiveness for Black students by teacher race. Second, as in previous same-race teacher effects

[15] In related work, Delgado (2023) likewise finds that teachers with a comparative advantage in raising Black students' test scores have no discernible effect on non-Black-students' test scores.

[16] Prior work has noted teacher-student match effects can be especially important for lower-scoring students e.g., Ahn, Aucejo, and James (2023)

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

$$
\begin{aligned}
A_{it} = {} & \beta_1(\text{HBCU}, \text{Black}_{it}) + \beta_2(\text{HBCU}, \text{nonBlack}_{it}) + \beta_3(\text{nonHBCU}, \text{nonBlack}_{it}) \\
& + X'_{ijst}\gamma + \delta_i + \eta_t + \lambda_s + \epsilon_{it}.
\end{aligned} \tag{6}
$$

A vector of student-, teacher-, classroom-, and school-level controls (as in column 2 of Table 2) is represented by X. 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.

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 0.027 standard deviations higher on math exams, compared with their performance in years with a non-HBCU-trained Black teacher.

Building on this point, I extend Equation (4) to White teachers in the sample, focusing on comparing HBCU-trained White teachers to White teachers who attended other institutions (column 2). Here, I find that Black students in classrooms with White teachers who attended an HBCU score on average 0.045 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.

I also fit regressions that model the relationship between all Black teachers and Black student test scores.

$$
A_{it} = \beta(\text{Black Teacher}_{it}) + X'_{ijst}\gamma + \delta_i + \eta_t + \lambda_s + \epsilon_{it}. \tag{7}
$$

Control vector X is the same as column 2 in Table 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 (though with a smaller magnitude), while I observe a larger, negative impact from assignment to a non-HBCU-trained Black teacher. Thus, 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.

5.7 Beyond test scores: the relationship between HBCU-trained teachers and student suspensions

Positive HBCU-trained teacher effects on Black student achievement may arise from teacher behavior and their classroom management practices. For instance, recent survey data focused on thousands of teachers of color show that, even among Black teachers, HBCU graduates reported the greatest amount of engagement with students, with more hours spent on mentoring and counseling and a greater reported number of students allowed to socialize in their classrooms outside regular class time (Bristol and El-Mekki, 2022). One avenue for investigating heterogeneous teacher effects on Black student achievement may lie with understanding differences in disciplinary action. Researchers have consistently showed the detrimental impacts suspensions can have on suspended students, including lower test scores for those suspended (Pope and Zuo, 2019), increased rates of school dropouts (Lee et al., 2011), and lower graduation rates and higher rates of being arrested and incarcerated (Bacher-Hicks et al., Forthcoming). Empirical evidence also 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 my focus on Black students, and especially so with over one-fourth of Black students in the sample being suspended at some point. To investigate the relationship between HBCU-trained teachers and Black student suspensions, I estimate Equation (6) shown below.

$$\text{Suspension}_{it} = \rho(\text{HBCU}_{it}) + \tau(\text{Suspension}_{it-1}) + W'_{ijct}\phi + \psi_g + \omega_{st} + \zeta_{it}. \quad (8)$$

With a relatively short panel — students are typically observed for no more than three years — and thousands of fixed effects, a nonlinear model such as a logit or probit would produce inconsistent estimates for the average marginal effect of interest (Angrist and Pischke, 2008; Greene, 2004). Combined with its ease of interpretation and computational feasibility, I instead use a linear probability model (LPM) to predict the probability that a Black student i is suspended in a given school-year t, specified in Equation (6). Whereas previous specifications use within-student variation in teacher assignment, I compare suspension probabilities across students, in a regression analogous to the lagged achievement model in Equation (2). Specifically, I regress whether the student has a suspension judgment in a given year on an indicator for teacher assignment (HBCU-trained or not), an indicator for receiving a suspension in the previous school year, time-varying student-, teacher, and classroom-level covariates, captured in vector W, and grade and school-by-year fixed effects.[17] The school-by-year fixed effect in particular accounts for any changes in school administration and leadership over time that may influence school suspension rates.

Figure 1 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. As a baseline, estimates for Black and White teachers align with prior work suggesting suspensions are less (more) likely with teacher racial (mis)match (Holt and Gershenson, 2015; Lindsay and Hart, 2017), possibly owing to racialized differences in teacher perceptions of Black student behavior (Downey and Pribesh, 2004). Moving to the 3rd row, I observe the change in suspension likelihood is negative for those assigned to HBCU-trained teachers, even after controlling for the teacher's race.[18]

Notably, this result on its own masks important, gender-specific variation in suspension rates. The association between HBCU-trained teacher assignment and sus-

[17]Each of the student-, teacher-, and classroom-level covariates is dichotomized to zero or one, replacing continuous variables with indicators for different quartile thresholds.

[18]Online Appendix Figure A1 demonstrates a negligible relationship between HBCU teacher assignment and suspension likelihood for Hispanic and White students.

pension probability for Black female students attenuates from the estimate for all students, while the association for Black male students is somewhat larger (-0.011, SE 0.005). As a point of reference, the mean predicted probability of suspension for the control group (Black male students with non-HBCU-trained teachers) is approximately 0.25, with a standard deviation of 0.43. Thus, assignment to an HBCU-trained teacher is associated with roughly a 0.026 SD decline 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. As with the test score analysis in Section 5.4., the likelihood of suspension for Black students declines with assignment to an HBCU-trained teacher, even among teachers of the same race (though estimates for Black teachers are smaller and imprecise). This is particularly stark when comparing White teachers to HBCU-trained White teachers: the probability of suspension increases when a Black student has a White teacher, but White teachers who went to an HBCU are less likely to suspend their Black students, compared with other White teachers (-0.011, SE 0.006 for Black girls, -0.020, SE 0.008 for Black boys).

Suspension results here are consistent with teachers from these universities having a particular approach to teaching and classroom management that uniquely correlates with reduced disciplinary actions against Black students. Irvine (1989) outlines the predicament Black students can often find themselves in at school 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). Townsend (2000) echoes this sentiment in her study of the disproportionate disciplinary actions faced by Black students, posing teacher training that prioritizes cultural competency as a remedy to address cultural misalignments. Through this lens, one interpretation of the White teacher result is HBCU-trained White teachers are less prone to referring Black students for suspensions than their White peers due to their background and training at HBCUs that mitigate cultural discontinuities.

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 possess the requisite cultural competency, with the capacity and willingness to guide Black students in bridging these gaps, may be reflected in differences in suspension rates.

6 Discussion of mechanisms

6.1 Motivating HBCUs and effects on Black student achievement: the role of teacher preparation

Why are HBCU-trained teachers more effective with Black students than their non-HBCU attending peers? Following the within-teacher race effects discussed in Section 5.6, especially as they pertain to White teachers, one can reason HBCU-trained teacher effectiveness is not due to a simple teacher-student racial match. A different explanation, then, focuses on a teacher's active behavior in the classroom. In particular, teachers may draw from *asset-based pedagogy* (ABP), an umbrella term for pedagogical practices that have emerged in the past three decades of teacher education research, including culturally relevant, responsive, sustaining, and revitalizing pedagogies (Ladson-Billings, 1995, 2014; Gay, 2002; Paris and Alim, 2014; McCarty and Lee, 2014; Kwok et al., 2023). ABP as a paradigm "views students' culture as a strength, countering the more widespread view that achievement disparities stem from deficiencies in the child and/or the child's culture" (López, 2017, pg. 193). In practice, it proposes that teachers possess key skills – e.g., critical awareness, cultural knowledge, and cultural content integration – necessary to affirm students' backgrounds and experiences as part of their instruction.

Consequently, the case for HBCU-trained teacher success with Black students may stem from the pedagogical approaches employed during their own preparation, which influence their classroom practices as teachers. Building on decades of qualitative work, quantitative studies in recent years that focus on implementation of asset-based pedagogy, culturally-fluent pedagogical practices and curriculum, and related themes have consistently shown these paradigms can positively impact academic outcomes for racial/ethnic minorities (e.g., Dee and Penner, 2017, 2021; Bonilla, Dee, and Penner, 2021; López, 2017). While culturally-fluent teaching includes general (e.g., high expectations for students) and specific (e.g., socially and culturally competent curriculum and instruction) practices, this approach is largely pedagogical in nature. How teachers are prepared for the classroom is a nontrivial input in determining their effectiveness. For example, while the instructors are all Black men in Dee and Penner (2021), 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). Asset-based pedagogy as a device in a teacher's instructional toolkit is applicable in the context of HBCUs given their historical approach to teacher education. 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. Dilworth (2012), in comparing departments of education mission statements, proposes that HBCUs, in addition to providing the requisite skills for high-quality teaching, aim to produce students who serve the community. Toldston and Pearson (2019) find a similar theme in interviews with deans of education at four HBCUs. Interviewees all expressed their HBCU's (1) commitment to fostering relationships with the community and local school districts; (2) continued professional development for teacher candidates; and (3) emphasis on "an expectation that their teacher candidates excel in areas related to social justice and socio-emotional learning" (pg. 22). Taken together, I propose that the sociocultural component of teaching is more likely a direct feature of teacher education for HBCU graduates than it is for those from other institutions and a more plausible explanation for HBCU-trained teacher success with Black students than racial match alone.

6.2 Empirically testing the teacher training hypothesis: evidence from teacher undergraduate majors

Nevertheless, while teacher training theoretically explains the observed HBCU effects, one outstanding question that may confound results considers selection into these colleges: who goes to an HBCU? For example, despite the relatively large share in this sample, it is fairly uncommon for White students to attend HBCUs; their motivations for attending these colleges and attitudes towards working with Black people 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. 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.

Alternatively, HBCU teacher *training* may matter less than *attendance* at these institutions. The predictive feature of HBCU-trained teacher success with Black students could be orthogonal to the program they are trained in, instead owing to exposure to an environment with a substantial share of Black students and professors with whom to interact and from whom to learn, akin to intergroup contact theory (Allport, 1954; Pettigrew and Tropp, 2006).

Available data render an exercise exploring pre-college matriculation characteristics infeasible in this paper. However, I can glean insights using information about teacher college major background. For over one-third of the teachers in the analytic sample (37%), I observe the specific fields undertaken for their undergraduate studies and can use this variation to explore not only HBCU attendance, but also HBCU teacher training. I define a teacher to hold an education undergraduate major if the listed first or second major for the teacher is in a traditionally educated-related field (e.g., education, early childhood education, school psychology or counseling, etc.). Because I focus on math exam scores, I follow a similar strategy and flag STEM graduates, marking any teachers who have majors or graduate degrees in Science, Technology, Engineering, or Math fields. I supplement the college major data with information about certifications a teacher may have obtained. That is, I include controls for whether a teacher received a license in mathematics, elementary education, and/or a temporary teaching license. I then use these data in regressions that include indicators from three of four constructed categories: HBCU graduates who majored in an education field, HBCU graduates who majored in a non-education field, and the two analogous groupings for non-HBCU graduates. For both regressions, the coefficient of interest corresponds to the indicator for HBCU education graduates, compared against some omitted category. All regressions control for a teacher's STEM degree receipt and any certifications obtained.

Online Appendix Table A4 provides three descriptive comparisons between teachers in the college major sample: HBCU-trained teachers versus non-HBCU (Panel A), HBCU versus non-HBCU among education majors (Panel B), and HBCU-trained education majors versus HBCU graduates in other fields (Panel C). Overall, relative to the full set of teachers in the analytic sample (Table 1), this subset preserves the direction of the HBCU to non-HBCU comparisons, although some of the magnitudes differ (e.g., larger share of White HBCU-trained teachers, higher NTE/Praxis exam scores among HBCU-trained teachers). Reflecting on the college major and teacher certification data in this sample, HBCU-trained teachers in this sample are both less likely to hold an undergraduate degree in an education and more likely to have obtained a temporary, emergency, or provisional certification to teach.[19]

Using these data, I reveal three key findings, with regression results displayed in Table 6 and Figures 2 and 3. First, HBCU-attending teachers who specifically majored in an education-related program are more effective at raising Black student math test scores than education majors from non-HBCUs. Second, within the group of HBCU attendees, education field graduates are more effective than teachers who attended HBCUs but majored in other fields. This speaks to selection into HBCUs and how to differentiate attendance from training received at the colleges, as I am essentially conditioning away the selection portion by comparing teachers who made the same college choice (and thus are plausibly comparable on key unobservables). Estimates by teacher race on both these dimensions are noisier, though note there is a strikingly large difference between White HBCU-trained education majors and White education majors from non-HBCUs (0.072 SD).

Third, and arguably most illuminating: in Figures 2 and 3, echoing results from Figure 1, I find a large, negative effect on the likelihood of suspension for Black students assigned to HBCU-trained teachers who majored in an education field (compared against non-education HBCU graduates). However, the association is largely driven by HBCU-trained White teachers. Beyond selection into HBCUs, which may reflect a greater proclivity for engaging with Black community, results suggest there may be specific skill-learning, such as classroom management practices or cultural competency, that HBCU-trained White teachers are acquiring in their education departments. Juxtaposed with a similar result when comparing HBCU education majors to education majors from other types of colleges, these skills are conceivably HBCU-specific, as opposed to simply a function of having a background in education. In sum, while not necessarily conclusive evidence (especially given the relatively smaller sample sizes), estimates using teacher college major data are consistent with the proposition that teacher preparation at HBCUs, rather than selection into or attendance at these colleges, is a key force behind HBCU-trained teacher efficacy, particularly for White teachers.

[19]Given the historical focus on education fields at HBCUs, it is somewhat surprising that there are fewer HBCU education majors relative to other fields. As this paper is unable to consider selection into education degrees at HBCUs, future research may also delve into the college major choice specifically at these universities.

7 Conclusion

This is, to my knowledge, the first quantitative paper studying the effect of HBCU-trained teachers on Black student outcomes. Using various econometric approaches, I show not only do Black elementary school students in North Carolina assigned to HBCU-trained teachers score better on end-of-grade exams and reduce their likelihood of disciplinary action, but also this relationship extends beyond same-race teacher match: both Black and White HBCU-trained teachers are more effective with Black students than their same-race peer teachers from non-HBCUs. Additionally, I offer evidence that the effects likely stem from HBCU-specific training, particularly for White teachers, rather than selection into or mere attendance at HBCUs.

Specific to education policy, at least three major implications are derived from this work. First, that HBCU-trained teacher effectiveness extends beyond racial match should inform policy interventions regarding teacher-student matches. Focus beyond a strict racial match can alleviate both the potential for overburdening Black teachers towards burnout and attrition (Kohli, 2018; Mustafaa, 2023) and concerns among non-Black teachers who may undervalue their capacity for teaching Black students due to their racial mismatch. Second, this paper demonstrates the need for further study of HBCUs, to better understand the tenets of their teacher preparation programs and how other education schools can integrate relevant practices into their own curricula. Finally, this research highlights the potential benefits of hiring and retaining more teachers trained at Historically Black Colleges, despite the challenges affecting many, including declining enrollment shares (Hinrichs, 2024) and financial distress (Smith-Barrow, 2019). HBCUs are uniquely producing teachers capable of raising Black student outcomes and, consequently, policymakers should grapple with what gains may be lost in the absence of these institutions.

I focus on an educational setting, but these results apply more broadly to other social institutions. The interacting effects of race, training, and disposition on outcomes in this context should contribute to our intuition of the gains from same-race pairings. Racial concordance research often focuses heavily on the racial match and implicitly (or explicitly) defers to explanations for observed effects that rely on innateness. I propose evidence that the efficacy of racial congruence is less inherent to the paired individual's race, but due to a bundle of skills — in the teacher context, skills like asset-based pedagogy, classroom management practices, etc. — that can be learned and developed, regardless of race. It highlights the need for researchers to be clearer about implicit assumptions and underlying mechanisms when writing and employing models of match effects; otherwise, we can derive incorrect policy solutions. Future work may find value in extending this approach to other settings with an eye towards not *just* racial representation, but also matching on gender and socioeconomic background. Likewise, scholars may push further towards understanding which features are most salient to improving the outcomes of individuals outside of teacher education (e.g., medical patients, legal defendants, etc.), how important skills or traits are acquired, and how that knowledge can best be replicated and disseminated. For instance, Dahl et al. (2021) finds male recruits in a Norwegian military boot camp adopted more egalitarian attitudes about gender after being randomly exposed to female recruits in their squads, but the effects did not persist after treatment. This suggests a role for passive treatment affecting behavior in demographically incongruent settings, while leaving open the possibility that a more active treatment like structured training, as observed in the HBCU context, could impact longer-run outcomes.

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.

Ahn, T., E. Aucejo, and J. James (2021). The importance of matching effects for labor productivity: Evidence from teacher-student interactions. Working Paper 2106, California Polytechnic State University, Department of Economics.

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.

Allport, G. (1954). *The Nature of Prejudice.* Addison-Wesley.

Anderson, J. (2007). The historical context for understanding the test score gap. *National Journal of Urban Education and Practice 1*, 1–21.

Angrist, J. and J.-S. Pischke (2008). *Mostly Harmless Econometrics: An Empiricist's Companion.* Princeton University Press.

Aronson, J. (2002). Chapter 14 – stereotype threat: Contending and coping with unnerving expectations. In J. Aaronson (Ed.), *Improving Academic Achievement: Impact of Psychological Factors on Education.*

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 (Forthcoming). The school to prison pipeline: Long-run impacts of school suspensions on adult crime. *American Economic Journal: Economic Policy*.

Bacher-Hicks, A., M. J. Chin, T. J. Kane, and D. O. Staiger (2019). An experimental evaluation of three teacher quality measures: Value-added, classroom observations, and student surveys. *Economics of Education Review 73*, 1–15.

Bacher-Hicks, A. and C. Koedel (2023). Estimation and interpretation of teacher value-added in research applications. In *The Handbook of the Economics of Education*, Volume 6.

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.

Bates, M., M. Dinerstein, A. C. Johnston, and I. Sorkin (2023). Teacher labor market policy and the theory of the second best. Working paper.

Bertrand, M., E. Duflo, and S. Mullainathan (2004). How much should we trust differences-in-differences estimates? *Quarterly Journal of Economics 119*, 249–275.

Borusyak, K., X. Jaravel, and J. Spiess (Forthcoming). Revisiting event study designs: Robust and efficient estimation. *Review of Economic Studies*.

Bristol, T. and S. El-Mekki (2022). Unique impact, unique burdens: Insights into the black male educator experience. Retrieved from `https://www.donorschoose.org/black-male-educator-insights`.

Cherry-McDaniel, M. (2019). Skinfolk ain't always kinfolk: The dangers of assuming and assigning inherent cultural responsiveness to teachers of color. *Educational Studies 55*(2), 241–251.

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 C. R. Clifton (2023). Racial differences in student access to high-quality teachers. *Education Finance and Policy*, 1–29.

Clotfelter, C. T., H. F. Ladd, and J. L. Vigdor (2007). How and why do teacher credentials matter for student achievement? Working Paper w12828, NBER.

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.

Dahl, G. B., A. Kotsadam, and D.-O. Rooth (2021). Does integration change gender attitudes? the effect of randomly assigning women to traditionally male teams. *The Quarterly Journal of Economics 136*(2), 987–1030.

de Chaisemartin, C. and X. D'Haultfœuille (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review 110*, 2964–2996.

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), 158–165.

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.

Delgado, W. (2023). Disparate teacher effects, comparative advantage, and match quality. EdWorkingPaper 23-848.

Delhommer, S. (2022). High school role models and minority college achievement. *Economics of Education Review 87*.

Dilworth, M. E. (2012). Historically black colleges and universities in teacher education reform. *The Journal of Negro Education 81*(2), 121–135.

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.

Ferguson, R. F. (2003). Teachers' perceptions and expectations and the black-white test score gap. *Urban Education 38*(4), 460–507.

Freeman, K. (1999). Hbcus or pwis? african american high school students' consideration of higher education institution types. *The Review of Higher Education 23*(1), 91–106.

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. D. Hart, J. Hyman, N. W. Papageorge, and C. A. Lindsay (2022). The long-run impacts of same race teachers. *American Economic Journal: Economic Policy 14*(4), 300–342.

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.

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.

Goldhaber, D., V. Quince, and R. Theobald (2017). Does the match matter? exploring whether student teaching experiences affect teacher effectiveness. *American Educational Research Journal 54*(2), 325–359.

Goldhaber, D., V. Quince, and R. Theobald (2018). Has it always been this way? tracing the evolution of teacher quality gaps in us public schools. *American Educational Research Journal 55*(1), 171–201.

Greene, W. (2004). Fixed effects and bias due to the incidental parameters problem in the tobit model. *Econometric Reviews 23*(2), 125–147.

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.

Harper, B. and L. Mawhinney (2017). Teacher preparation at historically black colleges and universities: Remaining relevant in a climate of accountability. In L. Mawhinney and E. Petchauer (Eds.), *Teacher Education across Minority-Serving Institutions: Programs, Policies, and Social Justice*, pp. 161–176. New Brunswick, Camden, Newark, New Jersey; London: Rutgers University Press.

Harper, S. R. and I. Davis, Charles H. F. (2012). They (don't) care about education: A counternarrative on black male students' responses to inequitable schooling. *Educational Foundations 26*(1-2), 103–120.

Headley, A. M. and J. E. Wright (2020). Is representation enough? racial disparities in levels of force and arrests by police. *Public Administration Review 80*(6), 1051–1062.

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. (2024). An empirical analysis of racial segregation in higher education. *Educational Evaluation and Policy Analysis 19*(2), 218–251.

Hoekstra, M. and C. Sloan (2022). Does race matter for police use of force? evidence from 911 calls. *American Economic Review 112*(3), 827–860.

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. 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. Unpublished manuscript.

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

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.

James, J. and J. H. Wyckoff (2022). School segregation, teacher sorting, and the distribution of teachers. Annenberg Institute at Brown University EdWorkingPaper no. 22-659.

Kim, E. J. (2022). Signal weighted teacher value-added models. *Statistics and Public Policy 9*(1), 149–162.

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. and J. R. Betts (2011). Does student sorting invalidate value-added models of teacher effectiveness? an extended analysis of the rothstein critique. *Education Finance and Policy 6*(1), 18–42.

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.

Kohli, R. (2018). Behind school doors: The impact of hostile racial climates on urban teachers of color. *Urban Education 53*(3), 307–333.

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.

Kwok, A., J. Waddington, J. Davis, S. Halabi, D. Huston, and R. Hemsley (2023). Beginning teachers and strategies for asset-based pedagogy. Technical report. Annenberg Institute at Brown University EdWorkingPaper no. 23-732.

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.

Lee, T., D. Cornell, A. 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. D. 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.

López, F. A. (2017). Altering the trajectory of the self-fulfilling prophecy: Asset-based pedagogy and classroom dynamics. *Journal of Teacher Education 68*(2), 193–212.

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 J. R. Lockwood (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.

Mustafaa, F. N. (2023). Black educators' racial identity attitudes and culturally relevant pedagogy: A psychological framework and survey of within-race diversity. *American Educational Research Journal 60*(5), 1–35.

Mykerezi, E. and B. F. Mills (2008). The wage earnings impact of historically black colleges and universities. *Southern Economic Journal 75*(1), 173–187.

Nguyen, N., B. Ost, and J. A. Qureshi (2023). Ok boomer: Generational differences in teacher quality. Technical report. Annenberg Institute at Brown University EdWorkingPaper No. 23-831.

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.

Pettigrew, T. F. and L. R. Tropp (2006). A meta-analytic test of intergroup contact theory. *Journal of Personality and Social Psychology 90*(5), 751–783.

Pope, N. and G. W. Zuo (2020). Suspending suspensions: The education production consequences of school suspension policies.

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.

Rangel, M. A. and Y. Shi (Forthcoming). First impressions matter: Evidence from elementary school teachers. *Journal of Human Resources*.

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.

Sav, G. T. (2010). Funding historically black colleges and universities: Progress toward equality? *Journal of Education Finance 35*(3), 295–307.

Smith-Barrow, D. (2019). Many hbcus are teetering between surviving and thriving. The Hechinger Report, October 23, 2019. Accessed December 30, 2020, `https://hechingerreport.org/many-hbcus-are-teetering-between-surviving-and-thriving/`.

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.

Toldson, I. A. and D. Pearson (2019). Strategic priorities for historically black colleges and universities with teacher preparation programs. State Higher Education Executive Officers Association, 2019: 1-31.

Townsend, B. L. (2000). The disproportionate discipline of african american learners: Reducing school suspensions and expulsions. *Exceptional Children 66*(3), 381–391.

U.S. Department of Education, Office of Planning, E., P. Policy Development, and P. S. Service (2016). The state of racial diversity in the educator workforce.

U.S. Department of Education, O. f. C. R. (2016). 2013-14 civil rights data collection: A first look. Retrieved from `https://www2.ed.gov/about/offices/list/ocr/docs/2013-14-first-look.pdf`.

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.

Watkins, A. F. (2005). *Cultivating the Education of African American College Students: A Learning Styles Approach*, pp. 122–127. Boston: Pearson Education.

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.

Table 1. Descriptive statistics for teachers and students, 2008-2018

	Panel A: Teachers		
	HBCU-trained	**Non-HBCU-trained**	**Absolute Difference**
Black	73.9	6.33	67.6***
White	23.9	90.2	66.3***
Hispanic	0.87	1.02	0.15
Asian	0.24	0.59	0.35**
Female teacher	92.7	93.2	0.5
Mean age	43.7	40.0	3.7***
Mean years of experience	14.4	12.6	1.8***
Graduate degree	44.0	39.0	5.5***
Graduate degree from an HBCU	-0.52	0.23	0.75***
Mean NTE/Praxis z score	10.7	29.8	19.1***
Attended college out of state	7.3	12.8	5.5***
National Board Certified	1.50	1.39	0.11***
Mean classes taught in year	13.7	6.81	6.89***
Mean number of Black students taught in year	45.0	22.7	22.3***
Mean share of Black students in school			
N	2,423	25,078	

	Panel B: Black Students		
	Ever assigned HBCU-trained teacher	**Never assigned an HBCU-trained teacher**	**Absolute Difference**
Female student	49.6	49.6	0
Economically disadvantaged	86.0	82.1	3.9***
Disability services-eligible	15.2	16.8	1.6***
Limited English Proficiency	0.87	0.70	0.7***
Received suspension judgment	27.8	25.3	2.5***
Mean number of days absent	5.82	5.61	0.21***
Mean math exam score	-0.54	-0.47	0.07***
Mean ELA exam score	-0.47	-0.42	0.05***
N	68,773	178,515	

All values, except those associated with experience, age, NTE/Praxis score, classes taught, Black students taught, number of days absent, median math/ELA test scores are expressed as percent. NTE/Praxis scores are standardized by exam year. Both number of days absent and math/ELA exam scores are medians and represent the medians across all students after taking the median within students over time. Math/ELA exam scores are standardized by grade and year. 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% of Black students in the sample are ever categorized as economically disadvantaged during their academic tenure. *** $p < 0.01$, ** $p < 0.05$

Table 2. HBCU-trained teacher effects on math and ELA test scores

	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
HBCU-trained teacher effects, Math	0.030***	0.032***	0.030***	0.027***	0.024***	0.022**
	(0.005)	(0.005)	(0.006)	(0.006)	(0.006)	(0.009)
HBCU-trained teacher effects, ELA	0.012***	0.009***	0.009***	0.005	0.005	0.003
	(0.004)	(0.004)	(0.005)	(0.005)	(0.005)	(0.007)
Student, year, and school fixed effects	X	X	X			
School-by-year and grade fixed effects				X	X	X
Teacher and classroom controls	X	X	X	X	X	X
Endogenous controls				X		
Lagged own and peer test scores					X	X
Lagged own and peer test scores in both subjects						X
Twice lagged test scores						X
Math						
Control group test score mean	-0.482	-0.482	-0.482	-0.480	-0.470	-0.480
R^2	0.86	0.86	0.87	0.67	0.68	0.72
N	457,538	457,538	457,538	292,870	262,529	143,361
ELA						
Control group test score mean	-0.426	-0.426	-0.426	-0.426	-0.427	-0.444
R^2	0.87	0.87	0.87	0.64	0.65	0.69
N	474,464	474,464	474,464	305,958	277,391	145,244

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, age, and school-level proportions of student race/ethnicity, gender, students receiving free or reduced-price lunch. Teacher-level and classroom-level controls include a quadratic function of years of experience, 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, having a graduate degree from an HBCU, whether the teacher attended college out of state, and whether their undergraduate college is located in the school district where they teach, whether their college was historically a normal school, class year, and proportion variables for how many students were Black, White, Hispanic, female, economically disadvantaged, disability services eligible, and English language learner status. Endogenous student-level controls in Column 5 include an indicator for receipt of a suspension judgment during the year, number of days absent during the school year, whether the student is designated gifted, the share of students in the classroom who receive a suspension judgment or are designated gifted, and whether the teacher is National Board certified. Lagged achievement regressions (columns 4-6) use the controls from column 2, in addition to controls for student gender, age, whether the student had the teacher in a previous year, and lags for suspensions and days absent. Heteroskedasticity-consistent standard errors clustered at the classroom level in parentheses. *** $p < 0.01$, ** $p < 0.05$

Table 3. Effect of HBCU-trained teacher assignment on racial/ethnic minority gaps in math test scores

	(1)	(2)	(3)	(4)
HBCU-trained teacher	-0.003	0.006		
	(0.009)	(0.008)		
Black student	-0.102***	-0.470***	-0.094***	-0.445***
	(0.002)	(0.002)	(0.002)	(0.003)
Black*HBCU	0.028***	0.034***	0.010*	0.028***
	(0.005)	(0.008)	(0.005)	(0.010)
Hispanic student	0.009***	-0.070***	0.012***	-0.062***
	(0.002)	(0.003)	(0.002)	(0.003)
Hispanic*HBCU	0.015***	0.031***	0.008	0.035***
	(0.008)	(0.009)	(0.006)	(0.009)
Asian student	0.120***	0.256***	0.120***	0.252***
	(0.003)	(0.005)	(0.003)	(0.007)
Asian*HBCU	-0.009	-0.036	-0.004	-0.018
	(0.013)	(0.022)	(0.013)	(0.022)
Other racial/ethnic minority student	-0.036***	-0.174***	-0.033***	-0.162***
	(0.002)	(0.004)	(0.002)	(0.004)
Other*HBCU	0.025***	0.017	0.013	0.006
	(0.009)	(0.013)	(0.008)	(0.013)
Lagged own and peer test scores	X			
School-by-year fixed effects	X	X		
Grade fixed effects	X	X	X	X
Classroom fixed effects	X	X	X	X
Control group test score mean	0.23	0.22	0.23	0.22
R^2	0.72	0.33	0.76	0.41
N	1,245,398	1,274,645	1,244,342	1,267,475

All regressions include controls for a student's economic disadvantage status, disability status, and English language learner status (all in a given year), teacher race/ethnicity, gender, age, school-level proportions of student race/ethnicity, gender, students receiving free or reduced-price lunch, a quadratic function of the teacher's years of experience, 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, having a graduate degree from an HBCU, whether the teacher attended college out of state, whether their undergraduate college is located in the school district where they teach, whether their college was historically a normal school, class size, and classroom proportion variables for how many students were Black, White, Hispanic, female, economically disadvantaged, disability services eligible, and English language learner status, student gender, age, whether the student had the teacher in a previous year, and lags for suspensions and days absent. White students are the omitted reference group. Heteroskedasticity-consistent standard errors clustered at the classroom level in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table 4. Heterogeneous effects of HBCU-trained teachers on math test scores, by teacher race

Reference category:	**(1)** **Non-HBCU-trained Black teacher**	**(2)** **Non-HBCU-trained White teacher**
HBCU-trained Black teacher	0.027***	
	(0.006)	
HBCU-trained White teacher		0.045***
		(0.011)
Student fixed effects	X	X
Year fixed effects	X	X
School fixed effects	X	X
Control group test score mean	-0.557	-0.471
R^2	0.86	0.86
N	457,358	457,358

All regressions include controls for a student's economic disadvantage status, disability status, and English language learner status (all in a given year), teacher gender, age, school-level proportions of student race/ethnicity, gender, students receiving free or reduced-price lunch, a quadratic function of the teacher's years of experience, 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, having a graduate degree from an HBCU, whether the teacher attended college out of state, whether their undergraduate college is located in the school district where they teach, whether their college was historically a normal school, class size, and classroom proportion variables for how many students were Black, White, Hispanic, female, economically disadvantaged, disability services eligible, and English language learner status. Heteroskedasticity-consistent standard errors clustered at the classroom level in parentheses. *** $p < 0.01$.

Table 5. Heterogeneous effects of Black teachers on math test scores, by HBCU attendance

	(1)	(2)	(3)
All Black teachers	-0.005		
	(0.004)		
HBCU-trained Black teacher		0.013***	
		(0.004)	
Non-HBCU-trained Black teacher			-0.020***
			(0.005)
Student fixed effects	X	X	X
Year fixed effects	X	X	X
School fixed effects	X	X	X
Control group test score mean	-0.471	-0.481	-0.481
R^2	0.86	0.86	0.86
N	457,538	457,538	457,538

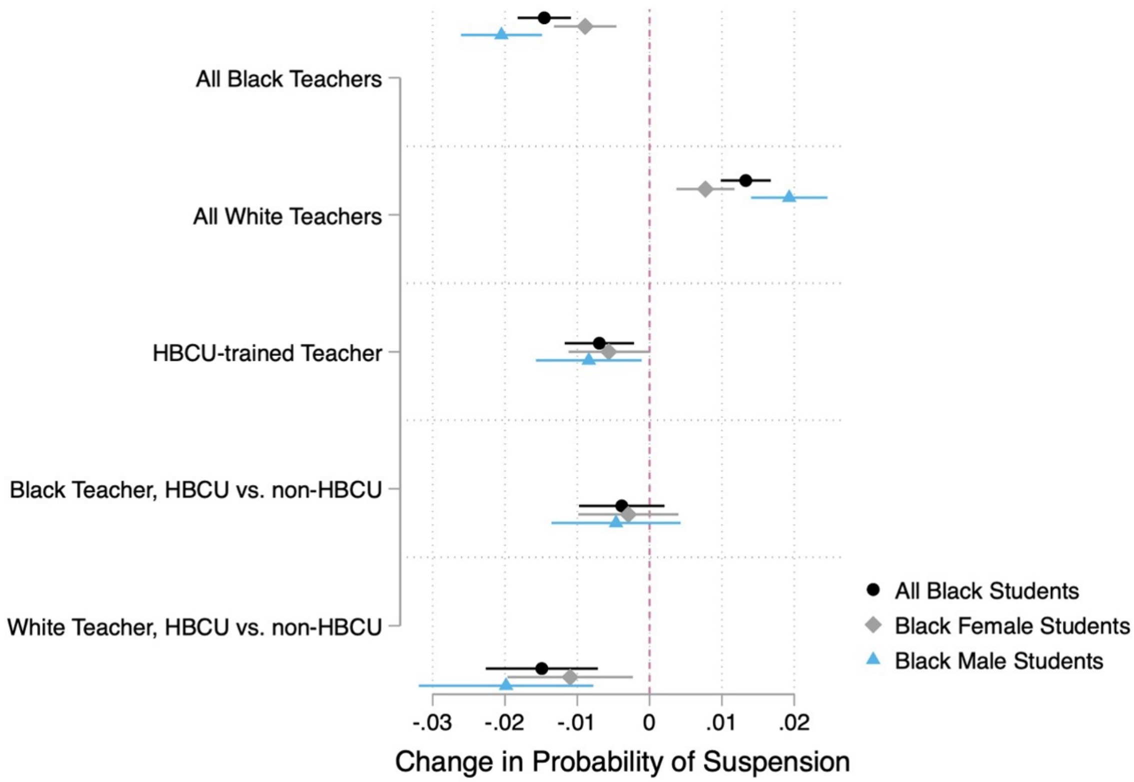
All regressions include controls for a student's economic disadvantage status, disability status, and English language learner status (all in a given year), teacher gender, age, school-level proportions of student race/ethnicity, gender, students receiving free or reduced-price lunch, a quadratic function of the teacher's years of experience, 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, having a graduate degree from an HBCU, whether the teacher attended college out of state, whether their undergraduate college is located in the school district where they teach, whether their college was historically a normal school, class size, and classroom proportion variables for how many students were Black, White, Hispanic, female, economically disadvantaged, disability services eligible, and English language learner status, student gender, age, whether the student had the teacher in a previous year, and lags for suspensions and days absent. Heteroskedasticity-consistent standard errors clustered at the classroom level in parentheses. *** $p < 0.01$.

Table 6. HBCU-trained teacher effects on math test scores, by teacher undergraduate major

Reference category:	**HBCU, Non-Education**			**Non-HBCU, Education**		
HBCU-trained teacher, Education field major	0.040***			0.037*		
	(0.012)			(0.019)		
HBCU-trained Black teacher, Education field major		0.028			0.020	
		(0.022)			(0.016)	
HBCU-trained White teacher, Education field major			-0.004			0.072***
			(0.049)			0.020
Student fixed effects	X	X	X	X	X	X
Year fixed effects	X	X	X	X	X	X
School fixed effects	X	X	X	X	X	X
Control group test score mean	-0.545	-0.539	-0.536	-0.509	-0.563	-0.503
R^2	0.89	0.89	0.89	0.89	0.89	0.89
N	83,658	83,658	83,658	83,658	83,658	83,658

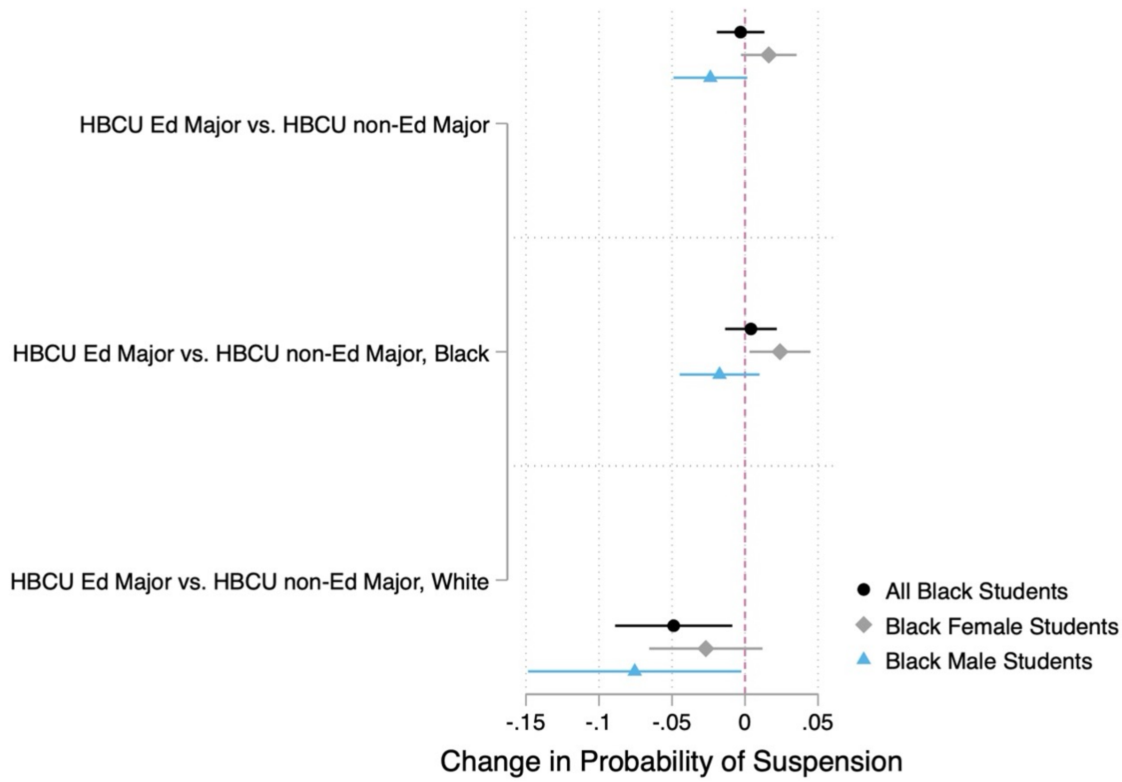
All regressions include controls for a student's economic disadvantage status, disability status, and English language learner status (all in a given year), teacher gender, age, school-level proportions of student race/ethnicity, gender, students receiving free or reduced-price lunch, a quadratic function of the teacher's years of experience, 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, having a graduate degree from an HBCU, whether the teacher attended college out of state, whether their undergraduate college is located in the school district where they teach, whether their college was historically a normal school, class size, and classroom proportion variables for how many students were Black, White, Hispanic, female, economically disadvantaged, disability services eligible, and English language learner status. Added controls for these specifications include indicators for having a STEM degree of any level and having a teacher license in math, elementary education, or a temporary/provisional license. Heteroskedasticity-consistent standard errors clustered at the classroom level in parentheses. *** $p < 0.01$, * $p < 0.10$.

Figure 1: Linear probability model estimates predicting suspension for Black students



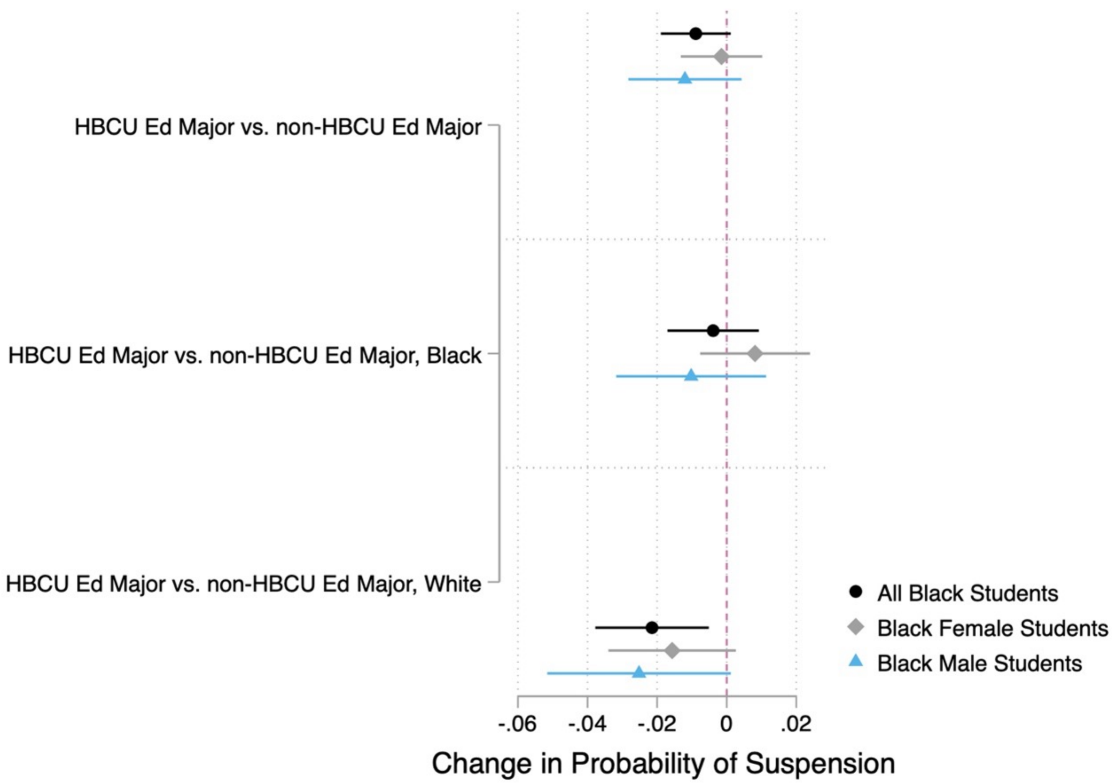
Regressions include controls for student gender, economic disadvantage status, disability status, English language learner status, prior year suspension judgment receipt, indicators for a student's quartile ranks of prior year's test score (based on quartile in grade distribution for the entire state in the prior school year) and for a teacher's quartile rank in the number of Black students taught in the prior year, whether a student had the teacher in the previous year, chronic absence in the prior school year (18 or more absences), teacher race, teacher gender, indicators for teacher age, novice teacher (three or fewer years of experience), attending an HBCU for graduate school, attending college out of state, teaching multiple classes in a year, large class size (greater than the 90th percentile of class size), classroom demographics (race, gender, economically disadvantaged, disability status eligible, English language learner status), and grade and school-by-year fixed effects. Zero line corresponds to 90% confidence interval. Heteroskedasticity-consistent standard errors clustered at the classroom level.

Figure 2: Predicting suspension for Black students, amongst HBCU graduates, by teacher race, and undergraduate major



Regressions include controls for student gender, economic disadvantage status, disability status, English language learner status, prior year suspension judgment receipt, indicators for a student's quartile ranks of prior year's test score (based on quartile in grade distribution for the entire state in the prior school year) and for a teacher's quartile rank in the number of Black students taught in the prior year, whether a student had the teacher in the previous year, chronic absence in the prior school year (18 or more absences), teacher race, teacher gender, indicators for teacher age, novice teacher (three or fewer years of experience), attending an HBCU for graduate school, attending college out of state, teaching multiple classes in a year, STEM degree, certification in math and/or elementary education, temporary teaching license, large class size (greater than the 90th percentile of class size), classroom demographics (race, gender, economically disadvantaged, disability status eligible, English language learner status), and grade and school-by-year fixed effects. Zero line corresponds to 90% confidence interval. Heteroskedasticity-consistent standard errors clustered at the classroom level.

Figure 3: Predicting suspension for Black students, by HBCU attendance, teacher race, and undergraduate major



Regressions include controls for student gender, economic disadvantage status, disability status, English language learner status, prior year suspension judgment receipt, indicators for a student's quartile ranks of prior year's test score (based on quartile in grade distribution for the entire state in the prior school year) and for a teacher's quartile rank in the number of Black students taught in the prior year, whether a student had the teacher in the previous year, chronic absence in the prior school year (18 or more absences), teacher race, teacher gender, indicators for teacher age, novice teacher (three or fewer years of experience), attending an HBCU for graduate school, attending college out of state, teaching multiple classes in a year, large class size (greater than the 90th percentile of class size), classroom demographics (race, gender, economically disadvantaged, disability status eligible, English language learner status), and grade and school-by-year fixed effects. Zero line corresponds to 90% confidence interval. Heteroskedasticity-consistent standard errors clustered at the classroom level.

ONLINE APPENDIX

Table A1. Alternative specifications to main TWFE regression

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
de Chaisemartin and D'Haultfoeuille (2020b) estimator	0.042*** (0.010)							
Only students with HBCU-trained teacher in 5th grade		0.036*** (0.008)						
Students with one teacher, one classroom, and one school in a given year			0.031*** (0.005)					
Teachers with only a bachelor's degree				0.024*** (0.007)				
Only teachers who attended college out-of-state					0.051*** (0.018)			
Students observed in all grades						0.037*** (0.006)		
Teachers who teach at least three Black students in a year							0.033*** (0.005)	
Teachers observed at least three years								0.034*** (0.005)
Student fixed effects	X	X	X	X	X	X	X	X
Year fixed effects	X	X	X	X	X	X	X	X
School fixed effects	X	X	X	X	X	X	X	
Control group test score mean		-0.469	-0.461	-0.500	-0.437	-0.471	-0.495	-0.450
R^2		0.87	0.86	0.88	0.89	0.84	0.89	0.87
N	227,819	351,731	398,135	210,630	60,035	238,568	430,793	333,417

All regressions include controls for a student's economic disadvantage status, disability status, and English language learner status (all in a given year), teacher gender, age, school-level proportions of student race/ethnicity, gender, students receiving free or reduced-price lunch, a quadratic function of the teacher's years of experience, 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, having a graduate degree from an HBCU, whether the teacher attended college out of state, whether their undergraduate college is located in the school district where they teach, whether their college was historically a normal school, class size, and classroom proportion variables for how many students were Black, White, Hispanic, female, economically disadvantaged, disability services eligible, and English language learner status. Given the computation of the de Chaisemartin and D'Haultfoeuille (2020b) estimator, test score mean and R^2 are unknown and omitted from this table. This estimator's standard errors were computed using 500 bootstrap replications. Heteroskedasticity-consistent standard errors clustered at the classroom level in parentheses. *** $p < 0.01$.

Table A2. Distributional effects of HBCU-trained teachers on math test scores

	(1) Previously low-scoring Black student	(2) Previously high-scoring Black student
HBCU-trained teacher	0.045***	0.020*
	(0.007)	(0.011)
Student fixed effects	X	X
Year fixed effects	X	X
School fixed effects	X	X
Control group test score mean	-0.895	0.691
R^2	0.81	0.82
N	150,398	41,073

"Previously low-scoring" and "previously high-scoring" refer to a student being in either the bottom two or top two quartiles of the test score distribution across the entire state for their grade in the prior school year. All regressions include controls for a student's economic disadvantage status, disability status, and English language learner status (all in a given year), teacher gender, age, school-level proportions of student race/ethnicity, gender, students receiving free or reduced-price lunch, a quadratic function of the teacher's years of experience, 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, having a graduate degree from an HBCU, whether the teacher attended college out of state, whether their undergraduate college is located in the school district where they teach, whether their college was historically a normal school, class size, and classroom proportion variables for how many students were Black, White, Hispanic, female, economically disadvantaged, disability services eligible, and English language learner status. Heteroskedasticity-consistent standard errors clustered at the classroom level in parentheses. *** $p < 0.01$, * $p < 0.10$.

Table A3. HBCU-trained teacher school-grade-year share effects on Black student test scores

	(1)	(2)	(3)	(4)
	Change in average cohort Black student test scores		Average cohort Black student test score	
Change in 1 SD of share HBCU-trained teachers in school-grade cohort	0.010** (0.0002)	0.010* (0.0002)	0.010** (0.0001)	0.010** (0.0001)
School-by-year fixed effects	X	X	X	X
Outcome mean	0.014	0.013	-0.421	-0.423
Outcome SD	0.515	0.511	0.478	0.460
R^2	0.35	0.54	0.68	0.70
N	23,825	22,502	24,953	22,502

Each observation in these regressions is at the school-grade-year level. Columns 1 and 3 regress the outcome variable only on the change in the share of HBCU-trained teachers from year t-1 to year t. Columns 2 and 4 and additional controls to capture other differences between cohorts, including a cubic polynomial of lagged ELA test scores (for the previous cohort) and changes in the share of Black teachers, White teachers, Hispanic teachers, Asian teachers, economically disadvantaged students, disability status, English language learner status, female teachers, changes in average teacher experience, NTE/Praxis test score, and teacher age. For interpretation, coefficients have been rescaled by multiplying by the standard deviation of the independent variable (18.4) to obtain estimates for the change in the outcome given a 1 SD change in the share of HBCU-trained teachers in the school-grade cohort. All regressions are weighted by the number of Black students in a school-grade-year cell. Heteroskedasticity-consistent standard errors clustered at the school-grade level in parentheses. ** $p < 0.05$, * $p < 0.10$.

Table A4. Descriptive statistics for teachers in college major data sample

Panel A: All Teachers	HBCU-trained	Non-HBCU-trained	Absolute Difference
Race/ethnicity			
Black	65.9	7.09	58.81***
White	29.6	90.2	60.6***
Hispanic	0.87	1.02	0.15
Asian	0.44	0.73	0.29
Female teacher	92.7	93.2	0.5*
Mean age	36.8	33.2	3.6***
Mean years of experience	6.75	6.28	0.47***
Graduate degree	34.1	31.8	2.3
Graduate degree from an HBCU	8.89	1.16	7.73***
Mean NTE/Praxis z score	-0.27	0.25	0.52***
Attended college out of state	9.48	25.3	15.82***
National Board Certified	3.26	5.10	1.84**
Mean classes taught in year	1.24	1.20	0.04
Mean number of Black students taught in year	12.4	6.43	5.97
Mean share of Black students in school	45.6	24.3	21.3
Major in education-related field	76.6	81.9	5.3***
Degree in STEM field	3.26	4.77	1.51
Certification in math	3.56	4.77	1.21
Certification in elementary education	98.7	99.5	0.8
Temporary/provisional/emergency certification	7.85	4.69	3.16***
N	675	8,620	
Panel B: Education majors	**HBCU-trained**	**Non-HBCU-trained**	**Absolute Difference**
Race/ethnicity			
Black	59.4	5.91	53.49***
White	35.4	90.1	54.7***
Hispanic	2.70	1.44	1.26**
Asian	0.58	0.62	0.04
Female teacher	90.1	93.0	2.9*
Mean age	36.1	32.6	3.5**
Mean years of experience	6.35	6.10	0.25**
Graduate degree	32.1	27.4	4.7***
Graduate degree from an HBCU	7.93	1.06	6.87***
Mean NTE/Praxis z score	-0.14	0.25	0.39***
Attended college out of state	6.0	20.4	14.4***
National Board Certified	3.48	5.10	1.62
Mean classes taught in year	1.23	1.20	0.03
Mean number of Black students taught in year	12.0	6.32	5.68***
Mean share of Black students in school	44.5	23.9	20.6***
Degree in STEM field	1.35	0.48	0.87***
Certification in math	3.09	4.32	1.23
Certification in elementary education	99.6	99.9	0.3
Temporary/provisional/emergency certification	1.93	1.59	0.34
N	517	7,059	
Panel C: HBCU-trained teachers	**Education majors**	**Non-Education Majors**	**Absolute Difference**
Race/ethnicity			
Black	59.4	87.3	27.9***
White	35.4	10.8	24.6***
Hispanic	2.70	0.63	2.07
Asian	0.58	0	0.58

Female teacher	90.1	88.0	2.1
Mean age	36.1	39.5	3.4***
Mean years of experience	6.35	8.24	1.9***
Graduate degree	32.1	40.5	8.4*
Graduate degree from an HBCU	7.93	12.0	4.07***
Mean NTE/Praxis z score	-0.14	-0.71	0.85***
Attended college out of state	6.0	20.9	14.9***
National Board Certified	3.48	2.53	0.95
Mean classes taught in year	1.23	1.28	0.05
Mean number of Black students taught in year	12.0	14.0	2.0***
Mean share of Black students in school	44.5	49.8	5.3***
Degree in STEM field	1.35	9.49	8.14***
Certification in math	3.09	5.06	1.97
Certification in elementary education	99.6	95.6	4***
Temporary/provisional/emergency certification	1.93	27.2	25.27***
N	517	158	

All values, except those associated with experience, age, NTE/Praxis score, classes taught, and Black students taught are expressed as percent. NTE/Praxis scores are standardized by exam year. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Figure A1. Linear probability model estimates predicting suspension for all, Hispanic, and White students