Statistical Discrimination and Optimal Mismatch in College Major Selection

Mary Kate Batistich
University of Notre Dame

Timothy N. Bond
Purdue University and IZA

Sebastian Linde
Texas A&M University

Kevin J. Mumford
Purdue University

January 31, 2024

Abstract

We develop a model of college major selection in a labor market with statistical discrimination. Heterogeneous students who differ in their aptitude choose from a continuum of majors with differing human capital production functions. Employers do not observe productivity but do observe student major, and a signal of productivity which is more precise for white applicants. Our model predicts that black students will choose more difficult majors than similar white students, but receive lower equilibrium labor market returns to major difficulty. We find empirical support for our model using administrative data from several large universities and two nationally representative surveys.

JEL Codes: J71, J15, I26

Batistich: Department of Economics, University of Notre Dame, 3060-I Jenkins Nanovic Hall, Notre Dame, IN 46556, USA; email: mbatisti@nd.edu. Bond: Department of Economics, Daniels School of Business, Purdue University, 403 Mitch Daniels Blvd., West Lafayette, IN 47907, USA; email: tnbond@purdue.edu. Linde: School of Public Health, Texas A&M, 212 Adriance Lab Rd. 1266 TAMU, College Station, TX 77843, USA; email: slinde@tamu.edu. Mumford: Department of Economics, Daniels School of Business, Purdue University, 403 Mitch Daniels Blvd., West Lafayette, IN 47907, USA; email: mumford@purdue.edu. We are grateful to seminar and conference participants at Georgia State University, the Midwest Economics Association Annual Meeting, the Mid-Midwest Applied Microeconomics Workshop, the Royal Economic Society and Scottish Economic Society Annual Conference, the Society of Labor Economics Annual Meeting, the Southern Economic Association Annual Meetings, and the Western Economic Association International Annual Conference for their helpful comments. All remaining errors are our own. IRB approval for this project was obtained through the Purdue University Institutional Review Board.

1 Introduction

The wage difference between college graduates with high paying and low paying degrees is nearly as large as that between high school and college graduates (Altonji et al., 2012). It is natural then that researchers and policymakers are concerned about major choice, especially for underrepresented minority (URM) students. In fact, a central argument against affirmative action in university admissions is that it causes URM students to graduate in a lower paying major than they would have if they had instead attended a less prestigious institution (e.g., Arcidiacono et al., 2012, 2016). This "mismatch" hypothesis was recently cited by Supreme Court Justice Clarence Thomas in his concurring opinion on the *Students for Fair Admissions v. Harvard* decision which curtailed the use of racial admissions preferences in the United States.

In this paper, we focus on black students specifically and explore an alternative force which impacts black major choice: anticipated statistical discrimination. Using a model similar in spirit to Lang and Manove (2011), we show that statistical discrimination distorts the human capital investment choices of black students, causing them to "mismatch" by selecting majors which are more difficult than those chosen by whites with similar academic backgrounds. In equilibrium, black welfare is lower than if all black students were to choose their human capital maximizing investment. However, each individual choice is constrained optimal. Thus, in contrast to the "mismatch" hypothesis discussed above, moving individual black students to less difficult investments would make them worse off. Our model generates several unique predictions on racial differences in major choice and career outcomes, which we find strong empirical support for across three datasets.

In our model, students choose from a continuum of college majors which augment their initial level of aptitude to create human capital. Firms do not observe human capital, but do observe major choice and an unbiased signal of productivity. As is standard in the statistical discrimination literature, the signal is more precise for white workers than black workers. Thus, employers put more weight on observable information when evaluating black job candidates. Responding to these incentives, black students choose more difficult majors than similar whites, our first empirical prediction. While black students' major choice is optimal, they earn lower wages than white students within the same major because of the equilibrium mismatch, our second empirical prediction. Employers correctly anticipate that black candidates will be less productive than white candidates from the same major, because black candidates were initially less prepared and graduate with less human capital. Finally, the equilibrium of the signaling game results in a larger distortion in the investment of more academically prepared students. Thus our third prediction is that the the size of the black-white wage gap will grow in the difficulty of the major. This stands in contrast with conventional wisdom that the black-white wage gap is lowest among the highest skilled workers (e.g., Lang and Lehmann, 2012).

We test and confirm our model's predictions using administrative data from twelve large public universities and two nationally representative surveys on labor market outcomes. First, we find strong evidence that black students select and graduate in higher paying majors than white students with similar academic preparation. Consistent with the other predictions of our model, the black-white difference in major difficulty is increasing in academic preparation and the within-major racial wage gap is increasing in major difficulty. This holds for measures of difficulty based on labor market returns as well as course content, is true for both early career and prime age workers, and is robust to controls for institution quality.

Despite the dramatic differences in labor market returns across college major, racial differences in major choice have seen surprisingly little attention.[1] Arcidiacono et al. (2012) show that black students at Duke University are more likely to begin schooling in a science major than whites students, but have lower rates of finishing a major in science. Arcidiacono et al. (2016) similarly find substantial gaps in preparation between URM students who finish a STEM degree and those who do not within the University of California system. Sovero et al. (2021) show minority students at the University of California, Los Angeles actually have higher rates of STEM persistence after controlling for preparation. Bleemer and Mehta (2023) document a trend toward lower paying degrees for URM students since the 1990s, which they attribute to an increase in major enrollment restrictions.

Our study differs from these papers in several important ways. First, we show that racial

[1] See Altonji et al. (2016) for a recent review on the returns across major.

disparities in major choice are reversed after controlling for college preparation across a large set of universities of varying selectivity. Second, we document this finding across a fuller set of majors than typically studied in this literature. Third, we provide a theoretical foundation, grounded in the statistical discrimination literature, for understanding why URM students would enroll in higher paying majors. Fourth, our model generates additional predictions on labor market outcomes which we confirm using two nationally representative data sets.

Our paper contributes more broadly to the literature on race in higher education. It has been widely observed that black students attend higher quality universities than white students with similar academic backgrounds, possibly due to race conscious admissions policies (e.g., Sander, 2004; Arcidiacono et al., 2011b). Whether this leads to better or worse outcomes for black students is a question of much debate.[2] Loury and Garman (1995) find that black students have lower earnings and are less likely to graduate when they attend a university with a median SAT score substantially above their own. Hinrichs (2012) finds that statewide bans on affirmative action in admissions cause a shift of black students away from highly selective institutions, but do not decrease the share of black degree holders in the population.[3] Hinrichs (2014) finds similarly that such bans raise the graduation rate for black students at selective institutions while lowering the overall number of black graduates from these same selective institutions. Arcidiacano et al. (2014) find that a specific ban introduced in California led to improvements in black college graduation rates, in part by causing black students to attend less selective institutions. However, Bleemer (2022) finds this same ban led to decreased wages for URMs as adults, driven by Hispanics. Comparing students with identical application and admissions portfolios in Texas, Mountjoy and Hickman (2021) find no evidence that black students who attend more selective institutions perform worse on the labor market than those who attend less selective institutions.[4]

[2]The theoretical ambiguity is due to the "quality-fit tradeoff." Higher quality institutions have better resources, faculty, and peers, but less prepared students may struggle to learn if the teaching material is targeted towards a more advanced audience. For evidence that student quality and institution quality are complements see Light and Strayer (2000), Sallee et al. (2008) and Dillon and Smith (2020).

[3]Using a similar approach, Backes (2012) finds affirmative action bans led to modest decreases in the number of black graduates of public institutions.

[4]Mountjoy and Hickman (2021) note one exception: Black students who attend the two historically black universities (HBCUs) in Texas earn higher wages than those who attend more selective institutions.

We can apply the insight of our model to reconcile the seemingly disparate findings in the university admissions literature. If black students anticipate facing statistical discrimination in the labor market, those who are admitted (with or without affirmative action) will optimally choose to enroll in more selective universities than similarly prepared white students because they are disproportionately rewarded for observable information.[5] Affirmative action increases the choice set of colleges that black students can use to signal, which could potentially lower overall welfare for black students. However, all college decisions are correct in equilibrium. Moving a black student from a highly selective to a less selective institution will make that student worse off, due to a decrease in the market's beliefs about his aptitude. Thus, empirical strategies which compare marginal students across institutions will be unable to detect negative effects of mismatch. In contrast, affirmative action bans change all student investment choices, and thus market beliefs. If mismatch is detrimental, empirical strategies that use bans would be more likely to detect the negative effects of mismatch, and this could vary based on the amount of affirmative action occurring in the state before the ban was implemented.[6]

Our empirical results contribute to the growing body of evidence that student major selection responds to labor market incentives. Previous studies have found that students switched majors in response to cyclical fluctuations in energy prices, the dot-com bust, the fracking boom, and the 2007-2008 financial crisis (Ersoy, 2020; Han and Winters, 2020; Weinstein, 2022). Similarly, Aalto et al. (2022) find the COVID-19 pandemic caused a decrease in applications to hospitality vocational programs by high school students in Sweden, while Ganguli et al. (2022) find the pandemic increased the demand for online courses promoting telework skills in Saudi Arabia. Blom et al. (2021) show that students enroll in majors with better labor market prospects during recessions. Our paper's

[5]One theoretical issue in the mismatch literature is rationalizing why black students would enroll in selective institutions which make them worse off. Arcidiacono et al. (2011a) propose that universities have private information on student-university match quality as well as preferences for diversity. Marginal black students who are admitted to improve diversity mis-perceive their admissions as a positive signal of the university's private information. Our model in contrast can produce mismatch in an environment with full information about student quality and the human capital investment function.

[6]Studies based on bans of racial admissions preferences implicitly only identify whether the amount of affirmative action before the ban is preferable to no affirmative action at all. They cannot identify whether the current level of affirmative action is optimal. Thus the estimated effects of a ban will depend on the population studied if, for example, universities in some states give more preferences than others, or if some URM groups receive larger preferences than others.

empirical results add an additional labor market characteristic that affects student major choice: statistical discrimination.

We also add important additional evidence on the role anticipated discrimination has on pre-market factors. Lang and Manove (2011) use a model closely related to our own to show that statistical discrimination will cause black workers to overinvest in education, an observable measure of productivity. Consistent with this, they find that black students obtain nearly a year more of education on average than white students with the same AFQT test score.[7] Conversely, Lundberg and Startz (1983) and Coate and Loury (1993) show theoretically that statistical discrimination will cause black workers to underinvest in unobservable measures of productivity. Fryer and Loury (2005) use a tournament model to show that affirmative action can increase effort provision by a disadvantaged group. Akhtari et al. (forthcoming) find empirical evidence for this theory using data on student SAT scores before and after racial university admissions preference bans. Our paper shows that major selection also responds to anticipated labor market discrimination.

The rest of the paper is organized as follows. In section 2, we introduce our model where students select a college major taking into account the statistical discrimination behavior of future employers. In section 3, we describe our three data sources. In section 4, we empirically test our model's predictions on major selection and labor market outcomes. In Section 5, we discuss the implication of our results for empirical strategies designed to test the mismatch hypothesis. Section 6 concludes.

2 A Model Of Major Selection with Statistical Discrimination

We develop a two-period model similar in spirit to Lang and Manove (2011). There exists a large number of students who are either (b)lack or (w)hite. They differ in a_i which is bounded over $[a_L, a_H]$. We will interchangeably refer to a_i as "aptitude" or "college preparation" for ease of

[7] Building on work by Arcidiacono et al. (2010), Lang and Manove (2011) also allow for education to increase the precision of labor market signal received by employers. Thus their model predicts that racial educational attainment will converge moving up the AFQT distribution, which they confirm empirically. While we find some suggestive evidence for this convergence effect for college majors, we are unable to formally test this theory, as we lack sufficient samples of black students who overlap with the highest SAT score whites.

exposition, but it more accurately measures the stock of skills a student possesses when making postsecondary educational decisions and reflects both innate ability as well as the impact of early childhood investments, primary and secondary school quality, etc.

In period 1, students select from a continuum of investments m which differ in their human capital production function. In our empirical section m will represent college major choice, but our arguments would follow for any observable investment, including university quality. A student who selects m will produce p_i when they enter the labor market, where

$$p_i = f(a_i, m_i) + \zeta_i. \tag{1}$$

ζ_i is an idiosyncratic productivity shock with mean 0 and standard deviation σ_ζ. $f(a, m)$ is the investment-specific human capital production function. It is strictly increasing in a. Further, m is indexed by its complementary with a; $f(a, m)$ is single-peaked in m, with $\arg\max_m f(a, m)$ increasing in a. Finally, denoting $F(a) \equiv \max_m f(a, m)$, $\frac{\partial F}{\partial a} > 0$, so that we would expect the highest m to also be the highest paid m in the labor market. We will refer to higher levels of m as being "more difficult," again for ease of exposition.

In period 2, ζ and p are realized, and students enter the labor market. The market does not observe p, but it does observe m and an unbiased signal of a student's productivity s:

$$s_i = \log p_i + \epsilon_i, \tag{2}$$

where ϵ_i is normally distributed with mean zero and standard deviation σ_k^2, and $k \in \{b, w\}$ is a student's race. This reflects information this is learned, for example, from an interview. Following the tradition in the statistical discrimination literature, employers are better able to interpret this information for whites, so that $\sigma_w^2 < \sigma_b^2$.

In this model, an equilibrium requires that both students and employers are making optimal choices. Denote π_k as the race-specific employer belief function, w_k as the race-specific wage function, and M_k as the race-specific function which maps from aptitude to educational investment.

We then describe an equilibrium with the following definition:

Definition. *An equilibrium is a set of functions π_k, w_k, and M_k such that*

1. *w_k generates zero expected profit for firms given π.*

2. *M_k maximizes expected utility given w_k.*

3. *π_k is defined by Bayes' rule whenever possible.*

As in Lang and Manove (2011) we restrict attention to separating equilibria which are "well-behaved" as defined below. In what follows, we assume the existence of a well-behaved equilibrium and analyze its properties. We then prove the existence of a well-behaved separating equilibrium.

Definition. *A well-behaved equilibrium is an equilibrium with the following properties:*

1. *M_k is smooth, continuous, differentiable, and monotonically increasing in aptitude, a.*

2. *For any major, m, which is not utilized by any students of race k in equilibrium, $\pi_k = a_L$.*

2.1 Employer Beliefs and Wages

Note that in a well-behaved equilibrium, college major selection reveals a student's aptitude, a, to the market. Denote $A_k(m)$ as the inverse of $M_k(a)$. The distribution of productivity for students of race k with major m is normally distributed with mean $f(A_k(m), m)$ and standard deviation σ_ζ^2. As s is normally distributed, we can apply Bayes' rule to find the employer beliefs for all m that are used in equilibrium,

$$\pi_k(m, s) = \lambda_k f(A_k(m), m) + (1 - \lambda_k)s, \tag{3}$$

where $\lambda_k \equiv \frac{\sigma_k^2}{\sigma_\zeta^2 + \sigma_k^2}$ is the familiar expression from Bayesian updating with a normally distributed prior and signal. It then follows from the zero profit condition that wages are

$$w_k(\pi_k) = \pi_k(m, s). \tag{4}$$

2.2 Optimal Major Selection and Educational Outcomes

Now consider a student's optimal investment problem:

$$\max_{m} E_k(w|m,a), \tag{5}$$

where $E_k(w|m,a)$ is the expected wage for student of race k with aptitude a who attempts investment m,

$$E_k(w|m,a) = \lambda_k f(A_k(m), m) + (1-\lambda_k) f(a,m). \tag{6}$$

This follows from taking the expectation of (4), recognizing that s is equal to $f(a,m)$ in expectation. The expected wage is a weighted average of the market's beliefs about a student with major m and the student's actual productivity, with more weight being placed on the market's beliefs when the signal has higher variance. In other words, in choosing a more difficult major students gain benefits from a "sheepskin" effect $[f(A_k(m), m)]$, but beyond a certain point, students bear a cost of lower actual human capital from being in a major that is more difficult than optimal for their aptitude.

Proposition 1. *Denote $M^*(a)$ as $\arg\max_m f(a,m)$. In any well-behaved equilibrium, $M_k(a_L) = M^*(a_L)$, and $M_k(a') \geq M^*(a') \forall a' > a_L$*[8]

Proposition 1 follows from employer belief structures in well-behaved equilibria and says that students with the lowest level of aptitude will select the least difficult major. Students with a_L do not receive a benefit from choosing a higher m than the full-information optimum because they receive no sheepskin effect. In equilibrium, employers believe the least difficult major that is utilized must indicate workers of the lowest aptitude type, and therefore the lowest type workers will want to choose their full-information optimal major.

[8] Proofs of this and all other results can be found in Appendix A

Proposition 2. *In equilibrium, $M_k(a)$ can be characterized by the differential equation*

$$\frac{\partial M_k(a)}{\partial a} = -\lambda_k \frac{\partial f(a,m)}{\partial a}\left[\frac{\partial f(a,m)}{\partial m}\right]^{-1}.$$

Proposition 3. *In equilibrium, black students attempt more difficult majors than white students conditional on a for all $a > a_L$.*

Propositions 2 and 3 tell us that in equilibrium, black students select more difficult majors than white students with similar college preparation. All students choose a major, m, higher than the full-information optimum. However, black students are more "mismatched" than white students. This phenomenon is driven by statistical discrimination in the labor market. Black students have a higher marginal return to observable information than white students, which gives them larger incentives to increase their academic credentials by investing in more difficult majors.

2.3 Labor Market Outcomes

Our model predicts that black students will ceteris paribus choose relatively more difficult majors than white students due to labor market statistical discrimination. We now analyze the impact of this on labor market outcomes.

Proposition 4. *Black students earn lower wages than white students conditional on m for all $a > a_L$.*

Proposition 4 follows directly from Propositions 1 and 3. Black students graduate in majors that are more difficult than white students with the same a. Since in equilibrium human capital is decreasing at the margin in m, black graduates will be less productive than white graduates in the same m. This will be reflected in their wages.

Proposition 5. *The observed labor market return to major difficulty for black college graduates is lower than for white college graduates.*

Under statistical discrimination the observed return to an increase in m will be lower for black students than white students because the equilibrium of the signaling game is non-distortionary for the lowest abilities and majors, but induces a racial productivity gap for higher m. Thus, in the cross-section we expect to find a racial wage gap that increases in m.

3 Administrative and Survey Data

In this section we describe our three main data sources that we use in our analysis, as well as the construction of our major difficulty measures.

3.1 Major Difficulty Measures

In our model, we classify the "difficulty" of educational investments on an index, m, related to their complementary with aptitude, a, in the production of human capital. We construct three measures to translate this idea to our empirical analysis of college majors.

The first two are constructed from labor market outcomes in the American Community Survey (ACS) from 2011 to 2021 excluding the year 2020.[9] The ACS is an annual survey of people in the United States conducted by the U.S. Census Bureau. Importantly for our purposes, the ACS includes information on field of degree (the college major) aggregated to 173 categories for all individuals who hold a bachelor's degree or above. We adopt a similar approach to Bleemer and Mehta (2023), and compute the residuals from a regression of real log earnings on indicators for major as well as age and year fixed effects on a sample of white, prime age (25-54 year old), native-born, full-time, year-round, employed workers with at least a bachelor's degree. Our results are robust to instead using all prime age workers, or using only white men.[10] We compute two measures. The "Wage Return" is simply the value of these residuals, while the "Percentile Return" is the percentile ranking of each major on these residuals. This approach follows naturally from our model where, in equilibrium, the highest m educational investments are chosen by the highest a students, and produce the highest p workers, who receive the highest wages.

We also construct a separate measure of major difficulty that relies on course content rather than adult outcomes. We use administrative student transcript records from 12 large public universities,

[9] We exclude 2020 due to potential sampling difficulties related to the COVID-19 pandemic.

[10] See Appendix Tables B.2, B.3, B.4, and B.5.

which we refer to as the "state schools sample," and calculate the fraction of course credits in STEM for the average graduate of each major. We describe these data more fully in the next subsection. We report the values for all three of our measures for each major in Appendix Table B.1.

3.2 State Schools Sample

The state school sample is constructed from administrative student transcript records from 12 large public universities: Clemson, Colorado, Colorado State, Florida, Florida State, Georgia Tech, North Carolina State, North Carolina – Charlotte, Oklahoma, Purdue, Utah State, and Virginia Tech.[11] While these universities are not nationally representative, Denning et al. (2022) show that these students are quite similar to those from the nationally representative NELS:88 and ELS:2002 sample of top-50 public universities in race, gender, and the distribution of SAT scores. The data were obtained from school registrars through the MIDFIELD partnership. Institutions that participate in the MIDFIELD partnership share de-identified longitudinal student records for all degree-seeking undergraduate students. These records include demographic characteristics and admissions data as well as course grades, major, and degree earned. They cover the years 1987 through 2018, though not all universities are included in all years. The records contain no information on post-graduation outcomes.

Table 1 reports summary statistics for the state schools sample. We restrict attention to students who identify as black or white, and exclude students without a reported SAT or ACT test score. The primary advantage of these data is the large sample size, with over 900,000 student records. Black students have lower graduation rates, are disproportionately female, and have lower high school and college GPAs. Strikingly, despite having on average 111 point lower SAT scores, black students initially enroll in majors that have a higher wage return on both the return and percentile measures during their first year of college. Consistent with Arcidiacono et al. (2012)'s findings for Duke, these differences are reversed by graduation, with whites graduating in more difficult majors by all three of our measures. We also see black students are more likely to be enrolled in chemistry,

[11] The MIDFIELD partnership does not allow us to report any results separately by institution that would enable readers to identify the institution.

biology, or business while less likely to be enrolled in history, English, or agriculture.

3.3 American Community Survey Wages

To test our model's predictions on wages, we will return to the ACS data discussed previously. For these analyses, we restrict our sample to working age (16 to 64) native non-Hispanic white and black workers with at least a bachelor's degree who were full-time year round employed in the previous year, resulting in nearly 2.8 million observations. Table 2 displays summary statistics for this sample. We find that the black workforce has a higher fraction of female workers than the white workforce, in line with well-known racial differences in labor force participation (Neal, 2004). We also observe a substantial racial earnings gap of $16,250 annually (a 0.26 difference in log earnings). As we found in the state school sample, black workers have less difficult degrees than white workers.

3.4 The Baccalaureate and Beyond

The biggest weakness for our purposes in the ACS is a lack of information on college quality. One concern then is that any racial differences we find in the labor market returns to major choice would be due to differences in university enrollment patterns between black and white students. A central concern of the affirmative action and mismatch literature is that affirmative action in admissions leads black students to graduate in lower return majors than they would have had they attended a less selective college (e.g., Arcidiacono et al., 2016).[12] We therefore provide additional evidence from the Baccalaureate and Beyond 2008/18 (B&B).

The B&B is a nationally representative longitudinal study of 2007-2008 college graduates collected by the National Center for Education Statistics (NCES). It combines demographic characteristics, college admissions measures, and college academic records with follow-up surveys focused on employment, post-baccalaureate education, and other outcomes. Follow-up surveys were con-

[12] Note that our results would not be biased if black students attended worse colleges than white students as long as black major selection is uncorrelated with college quality, as the college quality effect would be accounted for by the black indicator. The mismatch literature proposes that for black students, college major quality is inversely correlated with university quality.

ducted one, four, and ten years after graduation (in 2009, 2012, and 2018). We restrict the sample to students who identify as either black or white and exclude those who are age 30 or older when they graduate from college.

We display descriptive statistics for our B&B sample in Table 3. Similar to what we observe in the state school data, black students are more likely to be female, and they graduate with a lower GPA than white students. There is a 150 point racial gap in average SAT scores. In raw terms, the racial wage gap in each year is much smaller than the unconditional racial gap in the United States. This reflects both the youth of the sample, as well as the fact that the racial gap is generally thought to be lower in more educated individuals (Lang and Lehmann, 2012). Black students graduate in majors that are more difficult according to our wage percentile measure, but less difficult in terms of STEM courses. The largest disadvantage of using the B&B data is the much smaller sample size with only about 11,500 individuals in the sample.[13]

4 Testing for Optimal Mismatch

We now test our model's main predictions in the environment of college major selection. The major itself is our measure of investment, while we use SAT scores, a measure of college preparation, as the stand-in for our model's aptitude parameter.

4.1 Academic Preparation and Major Selection

Our model makes a specific prediction on the relationship between major choice and academic preparation. There should be no racial differences in major choice for the least prepared students. However, as we move up the academic preparation distribution, black students will select more difficult majors than white students.

We first analyze this in the raw data for our state schools sample by plotting the relationship between SAT scores (in 25 equal sized bins for each race) and the major percentile return in Figure 1. In Panel A, we show first-year major selection among all students. Consistent with our model,

[13] The NCES requires that the number of observations in each table is rounded to the nearest 10.

black and white students at the bottom of the SAT distribution initially select similar majors. However, as we move up the SAT distribution black students rapidly overtake white students in percentile return. There is possibly some convergence at the top of the SAT distribution but we caution that the upper SAT bins are much wider for black students.[14] Many students change majors during college, and less than half in our sample ultimately graduate. In Panel B, we instead consider the graduation major of those students who graduate. We earlier saw that white students graduate in more difficult majors than black students (Table 1). However, once we account for student SAT scores, the racial gap reverses, consistent with our model.

The state schools sample is nearly ideal for testing our model's predictions regarding major choice as it is a large administrative data set spanning several different universities across a long time period. One shortcoming is its lack of national representation. We therefore also test our predictions on the B&B sample, which is a national representative survey of a single graduating cohort. Figure 2 reports the raw relationship between SAT score (in 20 equal sized bins) and the major percentile return in the B&B data for graduates.[15] The figure is much more noisy than Figure 1 because of the smaller sample size, but the relationship has a similar pattern. Students with the lowest SAT scores appear to select similar majors regardless of race. As we move up the SAT distribution, black students tend to graduate in higher return majors than white students.

To formally test our prediction, we estimate:

$$\text{Major}_{ijt} = \beta_1 \text{Black}_i + \beta_2 \left(\text{Black}_i \times \text{SAT}_i\right) + \boldsymbol{\theta X_i} + \gamma_{jt} + \epsilon_{ijt} \tag{7}$$

where the subscript i indicates the individual student, j the educational institution, and t is the year of enrollment. Black_i is an indicator for the student having identified as black. X_i is a set of individual characteristics. γ_{jt} is a vector of institution-by-enrollment year fixed effects. Major_{ijt} is one of our three measures of major difficulty for the primary major selected by student i in their

[14]One explanation for this convergence is provided by Lang and Manove (2011), who show in their model of statistical discrimination with education choice that if higher levels of education also make the productivity signal more precise, as proposed by Arcidiacono et al. (2010), then racial differences in education choice should converge for high levels of aptitude.

[15]The B&B does not include first-year major.

first year (first-year major) or for the student's primary major at graduation (graduation major).

We report estimates of Equation 7 using both the state schools and B&B data in Table 4. Due to differences in data availability, the two samples offer slightly different control variables. For the state school sample, student characteristics include high school GPA, a female indicator, and a transfer student indicator. For the B&B sample, student characteristics include the student's age at graduation and a female indicator. All specifications across both samples include student SAT score fixed effects. The state schools sample controls for institution-by-start-year fixed effects. As the B&B sample is drawn from a single cohort, and has few observations within any single institution, it only includes fixed effects for the Carnegie classification for the institutions.[16]

Panel A of Table 4 uses our wage return measure as the outcome variable. Consistent with Table 1, we find in column (1) that black students enroll in first-year majors with a 3.2 log point higher residual wage than white students with the same SAT score. In column (2) we add to the SAT fixed effects an interaction between SAT and race. As predicted by our model, the racial gap in major selection is increasing in SAT score. We find similar results in columns (3) and (4) when we look at graduation major for the sample of college graduates. In columns (5) and (6) we turn to graduates from B&B sample and find similar results.

In Panel B of Table 4 we repeat this analysis using instead the major percentile return as the outcome variable. We again see strong evidence in support of our model. Column (1) reports that black students enroll in a first-year major 4.2 percentiles higher on average than white students with the same SAT score, and, as shown in column (2), this difference is increasing in SAT score. These results hold for graduation major selection in both the state schools sample and the B&B as shown in columns (3) - (6).

In Panel C of Table 4 we instead use our STEM courses major difficulty measure. Similar to our findings in Panels A and B, we find that black students select first-year majors with 2.9 percent more STEM course credits on average than white students with the same SAT score, and that this gap is increasing in SAT score. These results also hold for graduation major selection across both

[16]We find similar point estimates when controlling for institution fixed effects rather than institution Carnegie classification fixed effects in specifications using the B&B sample, but with much larger standard errors due to the small number of student observations per institution. See Appendix Table B.6.

samples. In sum, we find strong evidence in Table 4 for our model's predictions on major choice.

4.2 Major Selection and Career Outcomes

Our model makes two further predictions on racial differences in career outcomes. Black workers will earn less than white workers who graduated with the same college major, and this racial wage gap will grow in major difficulty. To test this, we first use the ACS to estimate:

$$Y_{irst} = \alpha_1 \text{Black}_i + \alpha_2 \text{Major}_i + \alpha_3 \left(\text{Black}_i \times \text{Major}_i\right) + \boldsymbol{\theta X_i} + \gamma_{rs} + \delta_t + \epsilon_{irst} \tag{8}$$

where subscript i is for the individual, r indicates race, s indicates state of residence, and t indicates time. X_i is a set of individual controls. γ_{rs} is a set of possibly race-specific state fixed effects. δ_t is a set of time fixed effects. Black_i is an indicator for the student having identified as black. Major_{ijt} is one of our three measures of major difficulty for the individual's primary major at graduation. Our model predicts $\alpha_1 < 0$ and $\alpha_3 < 0$. Black college graduates should have lower wages than white graduates in the same major, and the measured return to major difficulty should be higher for whites.

In column (1) - (3) of Table 5, we estimate equation (8) using our ACS sample and cluster the standard errors by graduation major. In Panel A we use the wage return as the measure of major difficulty. With only our a basic set of controls (gender, age, and age-squared) we find strong evidence for both predictions our model. Black graduates earn 22% lower wages than white graduates in the same major, and have an observed return to major difficulty that is 32.5% lower than whites. This result is unchanged with the addition of state and year fixed effects in column (2) and race-specific state fixed effects in column (3). Panels B and C repeat this analysis using the percentile return and STEM courses measures of major difficulty, respectively. Our results are similar.

While the ACS offers a large and nationally representative sample, it does not contain information on the university the worker attended. To ensure our estimates are not driven by differences in the quality of the degree-granting institution, we turn to the B&B data in columns (4) and (5) of Table 5. We control for the Carnegie classification of institution type in both specifications.[17] The cost of using the B&B data is a much smaller sample which is limited to early career outcomes for a specific cohort. We use the up to three earnings observations (at year 1, year 4, and year 10) for each individual and again cluster the standard errors by graduation major. In column (5), we find similar though somewhat smaller point estimates, which perhaps reflects the youth of the sample, and we lose statistical significance when using the STEM courses. But our estimates remain consistent with our model.[18]

Our model's predictions on wages are driven by the interaction between student major choice and market beliefs. Black students choose more difficult majors than similar prepared white students, for which they are a worse match. This leads black students to graduate with less human capital than white students on average in the same major. As the market cannot observe human capital, but can observe major and race, firms pay lower wages to all black workers. One concern then is that our empirical results are entirely driven by the with-in major differences in preparation rather than the equilibrium effects of incomplete information. We test this in column (5) of Table 5 by including SAT fixed effects. If anything, our results are stronger, consistent with the importance of statistical discrimination.

4.3 Testing Alternative Explanations: Race or SES?

One reason a student may pursue a more difficult, higher paying major is financial need. For example, low socioeconomic status (SES) students may be less likely to select "risky" majors with lower expected payoffs, or that require graduate school (Monaghan and Jang, 2017). As black students come from on average lower SES backgrounds than whites, this provides a potential alternative mechanism for our empirical results. It is however difficult to argue that low SES whites

[17] We do not include institution-by-race fixed effects as institution-specific features that improve black student outcomes could influence major selection. We find similar point estimates when controlling for institution fixed effects rather than institution Carnegie classification fixed effects in specifications using the B&B sample, but with much larger standard errors due to the small number of student observations per institution. These results are included in Appendix Table B.6.

[18] In Appendix Tables B.4 and B.5 we shows that the results in Table 5 are robust to using alternative measures of major difficulty.

face statistical discrimination, at least to the extent to that faced by blacks.[19] Thus, we can test our theory both by analyzing whether our racial effects hold after accounting for measures of SES status (and thus whether they hold for both high and low SES blacks), as well as by comparing the outcomes of low SES whites to that of blacks.

Unfortunately none of our data have direct measures of childhood conditions or SES background. However both the state schools sample and the B&B include students' home zip codes. We therefore test this alternative hypothesis by including controls for three zip code SES characteristics: median household income, median education, and income mobility. Median income and education measures are taken from the 2017-2021 ACS, while the income mobility measure is from Opportunity Insights (Chetty et al., 2018).[20]

We begin with graduation majors in the state schools sample. As not all students have zip code data we have fewer observations for this exercise. In column (1) of Table 6 we reproduce column (4) of Table 4 for this sample, using wage return as our measure of difficulty. Reassuringly, the change in sample has little impact on our results. In column (2) we include the median household income control as well as its interaction with that student's SAT score. In contrast to the alternative hypothesis, we in fact we find that students from wealthier zip codes graduate in *more* difficult majors than those from less wealthy zip codes, though this difference is decreasing in SAT scores. The inclusion of these controls has no impact on our point estimates for the black indicator or its interaction with SAT. We find similar results when we instead measure SES status through median education (column 3) or income mobility (column 4).

We provide results from several related exercises in Appendix. In Table B.7 we find evidence

[19] The most common argument for why blacks would face stronger statistical discrimination is rooted in differences in language usage, which may be difficult to interpret by white employers (e.g., Lang, 1986). While there may be differences in English dialects between low and high SES whites, it seems unlikely they would be as severe as differences between Standard American English and African-American English. Bond and Salisbury (2018) argue that those outside of a region are unable to ascertain the information content of within region variation in white dialects.

[20] We connect home zip code to ZCTA using the UDS mapper (https://udsmapper.org/zip-code-to-zcta-crosswalk/), and then merge it with IPUMS NHGIS Data (Manson et al., 2023) from 2017-2021 at the ZCTA level. We also connect home zip code to county using the HUD crosswalk from first quarter 2010 (https://www.huduser.gov/apps/public/uspscrosswalk/home), and then merge it with county-level data from Opportunity Insights (https://www.opportunityatlas.org/). Income mobility is measured by county and is the share of individuals whose parents' incomes were at the 25th percentile that are in the top 20% of household incomes at age 35.

that low SES students choose more difficult first-year majors in the state school sample. However, our main race effects remain robust. We reproduce our results with the SES controls using the Percentile Return measure in Table B.8 and the STEM courses measure in Table B.9. Our results are again robust.

In Table 7, we produce analogous results to Table 6 for the B&B data. While we find some evidence that students from low SES zip codes are more likely to graduate in more difficult majors, our results on race are unaffacted. We repeat this exercise using the percentile return snd STEM courses measures in Appendix Table B.10 and find similar results.

In Table 8 we compare the race and SES effects on the earnings of B&B graduates. In column (1) we use the wage return measure and replicate column (4) of Table 5 on our sample of graduates with home zip code information. Our results are similar. In column (2) we include the median income measure. Students from high SES zip codes earn higher wages than students from low SES zip codes, but we see no evidence that this effect differs across SAT scores. We find similar effects when we instead use median education (column 3) or income mobility (column 4). Our results are further unchanged when we include SAT fixed effects in column (5). We repeat this exercise for our percentile return and STEM courses difficulty measures in Appendix Table B.11. We find little qualitative differences across these difficulty measures.

In summary, the results of Section 4.3 strongly support statistical discrimination as the mechanism for our empirical findings. The predicted racial effects hold across all specifications that include SES controls. In one dataset, we find that low SES students graduate in *less* difficult majors than high SES students with similar academic preparation. Finally, we find no evidence that that observed labor market return to major difficulty varies with SES status.

4.4 Heterogeneous Effects

Table 9 reports wage outcomes by gender and age group by estimating equation (8) using the ACS sample. First, in columns (1) and (2) we find some evidence that the black-white within major earnings gap is larger for men than women. However, the differences in the observed return to major difficulty are similar by sex. Columns (3)-(5) instead compare workers of different age groups. Our results suggest that the within major racial earnings gap is smallest among workers under the age of 30. This provides a potential explanation for why we generally find smaller effect sizes on the recent college graduates of the B&B. We find little evidence of racial differences in the observed return to major difficulty across age.

5 Empirical Strategies for Affirmative Action and Mismatch

Thus far we have derived a theoretical model of educational investment choice where individuals anticipate statistical discrimination in the labor market. Motivated by the increased value of signaling, black students optimally mismatch. They choose more difficult investments than white students, and investments which are more difficult than that which would maximize their human capital. Choosing a less difficult investment would raise their human capital but lower their wage, due to the impact that investment has on employer beliefs about worker productivity. Our empirical results are consistent with our theory in an environment, college major choice, that is likely to be less confounded by other factors that may lead to mismatch, such as affirmative action in college admissions decisions.

We now consider the implications for empirical strategies evaluating affirmative action and mismatch. Consider a modified version of our model. First, suppose that black students face an investment cost $c(m_i)$, with $\frac{\partial c}{\partial m_i} > 0$ and $\frac{\partial^2 c}{\partial m_i} > 0$. This could represent structural barriers caused by discrimination in higher educational institutions, or a mitigable (with cost) preparation disadvantage due to inequalities in secondary and primary education.

Proposition 6. *When black students face an additional investment cost, they may choose investments that are more difficult, less difficult, or equal to those of white students. They may choose investments that or more difficult, less difficult, or equal to $M^*(a)$*

The proposition shows that adding costs may cause black students to choose less difficult investments than white students. However, this need not be the case, and depends on the nature of the cost function. The incentives created by statistical discrimination still motivate black students to take on more difficult investments that partially counteract their higher costs.

Perversely, some amount of barriers may actually *improve* black labor market outcomes. When these costs are relatively low, they push black student investments closer to $M^*(a)$, which will raise their accumulated human capital. It will also raise black wages conditional on a, as firm beliefs are correct in equilibrium. However as the costs continue to increase, black students are pushed to investments that are *below* $M^*(a)$, creating a human capital gap with white students due to *underinvestment*, and a larger resulting wage gap as well.

Policymakers who are concerned with this loss of equity may seek to remedy this through an "affirmative action" subsidy, $b(m)$. What should this subsidy be? We can imagine several different aims. A policymaker may choose to set $b(m) = c(m)$, so that black and white students face an identical investment choice problem. Under this regime, the model reduces to that analyzed in Section 2. Black students will choose more difficult investments than white students, and receive lower wages in the market. Alternatively, a policymaker may choose a $b(m)$ so that $M_b(a) = M^*(a)$. That is, choose an affirmative action policy which incentivizes black students to select their human capital maximizing investment. Following our analysis in Section 2, under this regime, white students will overmatch, choosing more difficult investments than black students. Black students will then outearn white students with the same a, as black students' equilibrium investment choices generate higher levels of human capital than whites'. Finally, a policymaker may choose a $b(m)$ so that $M_b(a) = M_w(a)$. Under this regime, any aggregate racial wage gap will be due only to differences in a.

While being agnostic about the goals of policymakers, in our framework, we can view the mismatch hypothesis as stating that, at current $b(m)$, $M_b(a) > M^*(a)$, and therefore a reduction in $b(m)$ will improve black labor market outcomes.[21] Two different approaches have been proposed to test this theory. The first relies on the some students being as-good-as-randomly assigned to universities, while the second relies on state-level policy changes to affirmative action.

[21] Note that even in our signaling framework, any policy which induced $M_b(a) = M^*(a)$ would maximize black wages, because it would maximize the average black worker's human capital.

The as-good-as-random assignment strategy uses a natural experiment which shifts a small number of blacks students from a high quality institution to a low quality institution and then compares the wages of these students. In principle, this could come through a regression discontinuity design at an admissions test threshold, as is common in the returns to school quality literature (e.g., Hoekstra, 2009; Zimmerman, 2019). Mountjoy and Hickman (2021) instead compare students who have applied and were admitted to the same set of universities, but made different matriculation choices. In their data from Texas, they find large disparities in preparation between black and white students at top public universities. Yet, once excluding historically black colleges and universities (HBCUs), black students who attend better universities perform better on the labor market. They thus conclude that mismatch does not harm black students. The following proposition shows that the as-good-as-random assignment approach is essentially uninformative.

Proposition 7. *Assigning a black student with ability a to investment $m' < M_b(a)$ will lower this student's wages for any $M_b(a)$.*

While this approach identifies the causal return to university selectivity on labor market outcomes, it is not effective at evaluating whether $M_b(a) > M^*(a)$ when investments act as signals.[22] To see this, assume that $M_b(a) > M^*(a)$, and consider taking a small number of students and instead assigning them to to $M^*(a)$. From equation (6), such a student will receive expected wages

$$E_b(w|M^*(a), a) = \lambda_b f\left(A\left(M^*(a)\right), M^*(a)\right) + (1-\lambda_b) f(a, M^*(a)). \tag{9}$$

The first term of equation (9) is the signaling value of investment while the second is the human capital component. While this change in assignment raises accumulated human capital, it lowers the signaling value of investment, as the market now believes the worker has the lower level of ability typical associated with $M^*(a)$. It must be the case that equation (9) is lower than $E_b(w|M_b(a), a)$, since by definition of equilibrium $M_b(a)$ solves the optimal investment problem, not $M^*(a)$.

Now consider the second identification strategy. Several states have banned racial preferences

[22] Hoekstra (2009) also notes that such estimators will include both the causal effect of education quality on human capital, and the signaling value of institution quality.

in admissions, which allows researchers to compare the outcomes of cohorts who differed in their exposure to affirmative action (e.g., Hinrichs, 2012, 2014; Arcidiacono et al., 2016). Most recently Bleemer (2022) finds that a ban on racial admissions preferences in California led to reduced wages as young adults for URMs, with effects concentrated on Hispanics. Because these bans change the investment decisions of all black students, they also change employer beliefs. This offers two advantages over the as-good-as-random assignment approach. First, it can successfully evaluate the welfare effects of the particularly policy studied. Second, it will identify if mismatch is present whenever URM students perform better after the policy change. However, the next proposition shows that it is cannot *rule out* the mismatch hypothesis.

Proposition 8. *Consider two policy regimes $b_1(m) \geq 0$ and $b_2(m) = 0$. Policy b_2 may lead to higher, lower, or equal wages to policy b_1 if under b_1, $M_b(a) > M^*(a)$. Policy b_2 will lead to higher wages than b_1 only if under b_1, $M_b(a) > M^*(a)$*

The intuition is as follows. Under large barriers (high $c(m)$) without affirmative action ($b(m) = 0$), black investments will be too low relative to the human capital maximizing optimum. Thus, *even if* affirmative action induces overmatch, black students may see higher wages because this overmatch is less severe than the *undermatch* they experience in its absence. If we observe that black wages decrease after a ban in racial admissions preferences, it only tells us that black students were undermatched without these preferences. It does not tell us whether some other policy which lessened, but did not eliminate, racial admissions preferences would lead to higher black wages by reducing the amount of overmatch in black human capital investments. In contrast, since a ban on racial admissions preferences can only reduce the difficulty of black human capital investments, an increase in black wages would provide strong evidence that these preferences induced mismatch. Systematically reducing $M_b(a)$ will raise wages only if $M_b(a) > M^*(a)$.

While the above analysis shows that neither of the two methods can fully evaluate whether affirmative action induces mismatch when there is incomplete information about productivity in the labor market, we emphasize that the literature using affirmative action bans is able to inform on the benefits of that specific policy. However, we also note that our concerns about the as-good-as-random assignment method would be eliminated in a full information environment. If employers learn about worker productivity with labor market experience, in the long run the signaling value of investments approaches zero. In practice, the market appears to learn about worker productivity relatively quickly (Lange, 2007; Aryal et al., 2022). A regression discontinuity or similar approach using the wages of mid-career or older workers may be able to fully capture the human capital effects of institution quality on black students, and thus evaluate whether affirmative action leads to suboptimal mismatch. We caution however that taking this approach implicitly assumes that career trajectories and on-the-job human capital accumulation are not affected by initial labor market beliefs. Altonji and Pierret (2001) find that, holding cognitive skills fixed, the black-white wage gap widens with labor market experience, which is at least consistent with early labor market statistical discrimination influencing the lifecycle trajectory of skill acquisition.

6 Conclusion

In this paper we integrated a canonical statistical discrimination framework into a model of major choice. Doing so revealed a new tension common with that in education choice models: Black students are incentivized to overinvest in observable human capital measures. The equilibrium outcome of the signaling game leads black students to attempt majors with a higher return to aptitude than similarly prepared whites. Yet they receive in equilibrium a lower wage return to these majors because the market correctly incorporates the incentives to overcredentialize that black students face. We find broad support for our predictions using administrative data from 12 large public universities, the ACS, and the B&B.

Our paper provides a novel contribution to the literature on academic mismatch and affirmative action. In equilibrium, black students are "overmatched" in their major choices, but not due to information asymmetries or deficiencies, and not due to affirmative action. Instead, it is the rational response to anticipated statistical discrimination on the labor market. This suggests a potential policy role for universities seeking to improve the outcomes of their URM students. They should take action to enhance the ability of URM students to convey information about their skills and competencies to employers. For example, universities could provide additional interview training for black students, teaching them methods to overcome the information disadvantage at the core of statistical discrimination. Universities could also work to provide better opportunities for black students in lower return majors to reveal their aptitude to employers. This could include research opportunities that produce tangible results, or academic competitions.

Our work also provides a valuable lesson on the interpretation of regression discontinuity approaches when the measured outcome is determined by a market with incomplete information. In fact, we should expect a discontinuity in wage outcomes between individuals just below and just above a university admissions cutoff, independent of any human capital effect of that university itself, because there is a sharp change in employer beliefs at this cutoff. Particularly in the context of the affirmative action literature, large policy changes that change equilibrium beliefs may provide a more useful way of testing for mismatch than narrow policies that incrementally change student university choices.

References

Aalto, Aino-Maija, Dagmar Müller, and J. Lucas Tilley (2022) "From epidemic to pandemic: Effects of the COVID-19 outbreak ou high school program choices in Sweden," IZA Discussion Paper No. 15107.

Akhtari, Mitra, Natalie Bau, and Jean-William P. Laliberté (forthcoming) "Affirmative action and pre-college human capital," *American Economic Journal: Applied Economics*.

Altonji, Joseph G., Peter Arcidiacono, and Arnaud Maurel (2016) "The analysis of field choice in college and graduate school: Determinants and wage effects," in Eric A. Hanushek, Stephen Manchin, and Ludger Wößmann eds. *Handbook of the Economics of Education*, Vol. 5: North Holland, pp. 305–396.

Altonji, Joseph G., Erica Blom, and Costas Meghir (2012) "Heterogeneity in human capital investments: High school curriculum, college major, and careers," *Annual Review of Economics*, Vol. 4, pp. 185–223.

Altonji, Joseph G. and Charles R. Pierret (2001) "Employer learning and statistical discrimination," *Quarterly Journal of Economics*, Vol. 116, pp. 313–350.

Arcidiacano, Peter, Esteban Aucejo, Patrick Coate, and V. Joseph Hotz (2014) "Affirmative action and university fit: Evidence from Proposition 209," *IZA Journal of Labor Economics*, Vol. 3.

Arcidiacono, Peter, Esteban M. Acuejo, and Ken Spenner (2012) "What happens after enrollment? An analysis of the time path of racial differences in GPA and major choice," *IZA Journal of Labor Economics*, Vol. 1.

Arcidiacono, Peter, Esteban M. Aucejo, Hanmin Fang, and Kenneth I. Spenner (2011a) "Does affirmative action lead to mismatch? A new test and evidence?" *Quantitative Economics*, Vol. 2, pp. 303–333.

Arcidiacono, Peter, Esteban M. Aucejo, and V. Joseph Hotz (2016) "University differences in the graduation of minorities in STEM fields: Evidence from California," *American Economic Review*, Vol. 106, pp. 525–562.

Arcidiacono, Peter, Patrick Bayer, and Aurel Hizmo (2010) "Beyond signaling and human capital: Education and the revelation of ability," *American Economic Journal: Applied Economics*, Vol. 2, pp. 76–104.

Arcidiacono, Peter, Shakeeb Khan, and Jacob L. Vigdor (2011b) "Representation versus assimilation: How do preferences in college admissions affect social interactions?" *Journal of Public Economics*, Vol. 95, pp. 1–15.

Aryal, Gaurab, Manudeep Bhuller, and Fabian Lange (2022) "Signaling and employer learning with instruments," *American Economic Review*, Vol. 112, pp. 1669–1702.

Backes, Ben (2012) "Do affirmative action bans lower minority college enrollment and attainment? Evidence from statewide bans," *Journal of Human Resources*, Vol. 47, pp. 435–455.

Bleemer, Zachary (2022) "Affirmative action, mimsatch, and economic mobility after California's Proposition 209," *Quarterly Journal of Economics*, Vol. 137, pp. 115–160.

Bleemer, Zachary and Aashish Mehta (2023) "College major restrictions and student stratification," mimeo.

Blom, Erica, Brian C. Cadena, and Benjamin K. Keys (2021) "Investment over the business cycle: Insights from college major choice," *Journal of Labor Economics*, Vol. 39, pp. 1043–1082.

Bond, Timothy N. and Laura Salisbury (2018) "Local information, income segregation, and geographic mobility," *B.E. Journal of Economic Analysis and Policy*, Vol. 18.

Chetty, Raj, John N Friedman, Nathaniel Hendren, Maggie R Jones, and Sonya R Porter (2018) "The opportunity atlas: Mapping the childhood roots of social mobility,"Technical report, National Bureau of Economic Research.

Coate, Stephen and Glenn C. Loury (1993) "Will affirmative-action policies eliminate negative stereotypes," *American Economic Review*, Vol. 83, pp. 1220–1240.

Denning, Jeffrey T., Eric R. Eide, Kevin J. Mumford, Richard W. Patterson, and Merril Warnick (2022) "Why have college completion rates increaseed?" *American Economic Journal: Applied Economics*, Vol. 14, pp. 1–29.

Dillon, Eleanor Wiske and Jeffrey Andrew Smith (2020) "The consequences of academic match between students and colleges," *Journal of Human Resources*, Vol. 55, pp. 767–808.

Ersoy, Fulya Y. (2020) "The effects of the Great Recession on college majors," *Economics of Education Review*, Vol. 77.

Fryer, Roland G. Jr. and Glenn C. Loury (2005) "Affirmative action in winner-take-all markets," *Journal of Economic Inequality*, Vol. 3, pp. 263–280.

Ganguli, Ina, Jamal I. Haidar, Asim Ijaz Khwaja, Samuel W. Stemper, and Basit Zafar (2022) "Economic shocks and skill acquisition: Evidence from a a national online learning program at the onset of COVID-19," NBER Working Paper No. 29921.

Han, Luyi and John V. Winters (2020) "Industry fluctations and college major choices: Evidence from an energy boom and bust," *Economics of Education Review*, Vol. 77.

Hinrichs, Peter (2012) "The effects of affirmative action bans on college enrollment, educational attianment, and the demographic composition of universities," *Review of Economics and Statistics*, Vol. 94, pp. 712–722.

——— (2014) "Affirmative action bans and college graduation rates," *Economics of Education Review*, Vol. 42, pp. 43–52.

Hoekstra, Mark (2009) "The effect of attending the flagship state university on earnings: A discontinuity-based approach," *Review of Economics and Statistics*, Vol. 91, pp. 717–724.

Lang, Kevin (1986) "A language theory of discrimination," *Quarterly Journal of Economics*, Vol. 101, pp. 363–382.

Lang, Kevin and Jee-Yeon K. Lehmann (2012) "Racial discrimination in the labor market: Theory and empirics," *Journal of Economic Litarature*, Vol. 50, pp. 959–1006.

Lang, Kevin and Michael Manove (2011) "Education and labor market discrimination," *American Economic Review*, Vol. 101, pp. 1467–1496.

Lange, Fabian (2007) "The speed of employer learning," *Journal of Labor Economics*, Vol. 25, pp. 1–35.

Light, Audrey and Wayne Strayer (2000) "Determinants of college completion: School quality or student ability?" *Journal of Human Resources*, Vol. 35, pp. 299–332.

Loury, Linda Datcher and David Garman (1995) "College selectivity and earnings," *Journal of Labor Economics*, Vol. 13, pp. 289–308.

Lundberg, Shelley J. and Richard Startz (1983) "Private discrimination and social intervention in competitive labor markets," *American Economic Review*, Vol. 73, pp. 340–347.

Manson, Steven, Jonathan Schroeder, David Van Riper, Katherine Knowles, Tracy Kugler, Finn Roberts, and Steven Ruggles (2023) "IPUMS National Historical Geographic Information System: Version 18.0 [dataset]," University of Minnesota.

Monaghan, David and Sou Hyun Jang (2017) "Major payoffs: Postcollege income, graduate school, and the choice of "risky" undergraduate majors," *Sociological Perspectives*, Vol. 60, pp. 722–746.

Mountjoy, Jack and Brent R. Hickman (2021) "The returns to college(s): Value-added and match effects in higher education," NBER Working Paper No. 29276.

Neal, Derek (2004) "The measured black-white wage gap among women is too small," *Journal of Political Economy*, Vol. 112, pp. S1–S28.

Sallee, James M., Alexandra M. Resch, and Paul N. Courant (2008) "On the optimal allocation of students and resources in a system of higher education," *B.E. Journal of Economic Analysis & Policy*, Vol. 8.

Sander, Richard H. (2004) "A systematic analysis of affirmative action in American law schools," *Stanford Law Review*, Vol. 57, pp. 367–483.

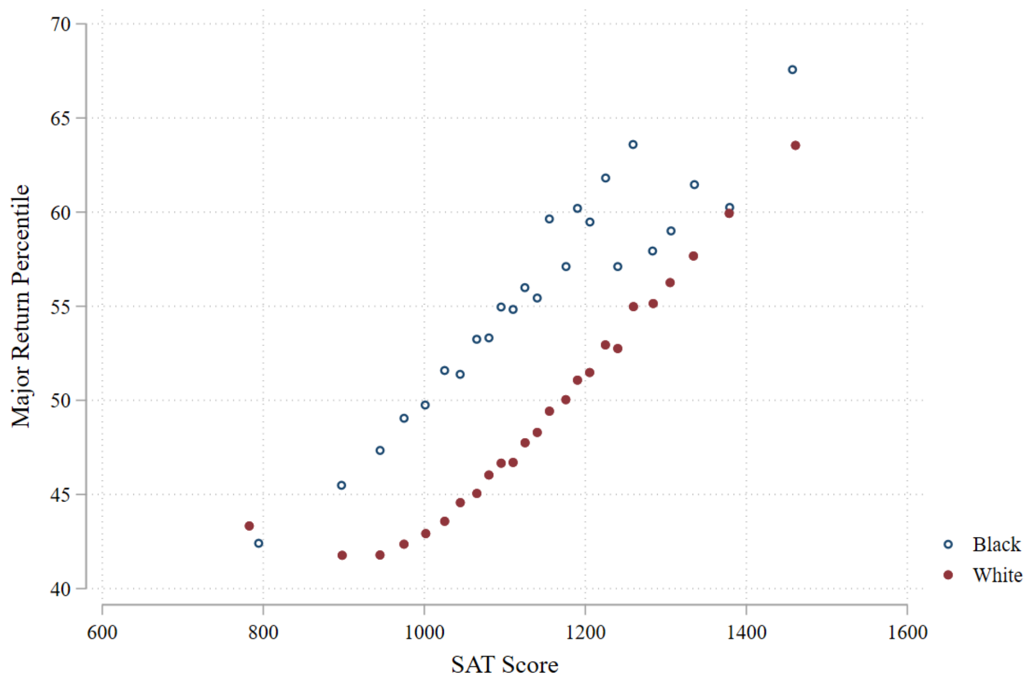
Sovero, Veronica, Moshe Buchinsky, and Matthew D. Baird (2021) "Playing catch up: A term-level investigation of teh racial gap in STEM retention," *Economics of Education Review*, Vol. 83.

Weinstein, Russell (2022) "Local labor markets and human capital investments," *Journal of Human Resources*, Vol. 57, pp. 1498–1525.

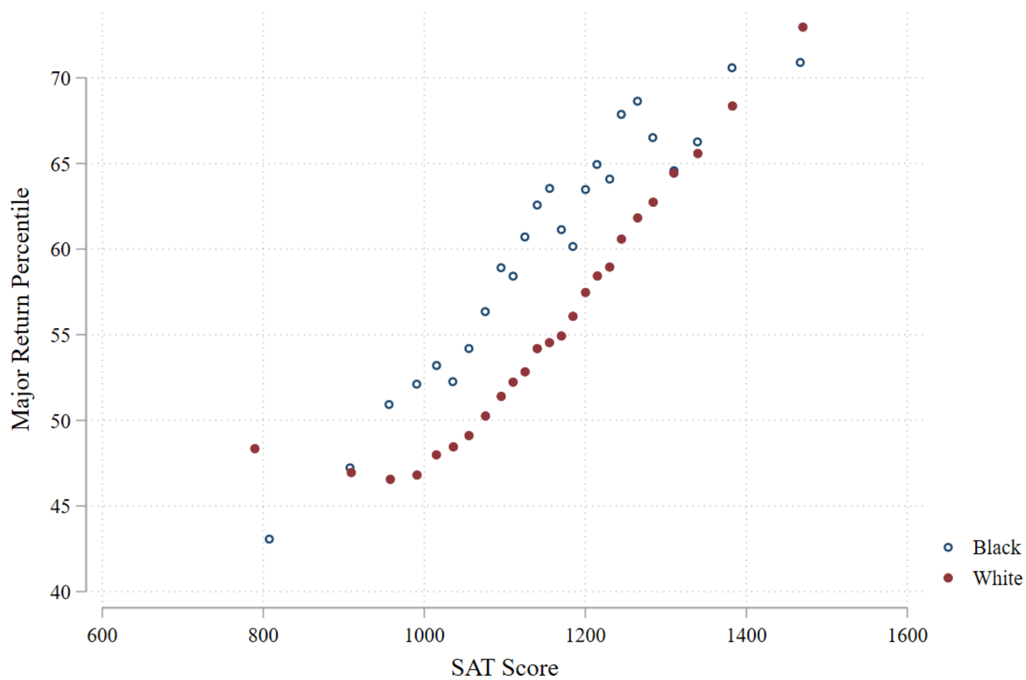
Zimmerman, Seth D. (2019) "Elite colleges and upward mobility to top jobs and top incomes," *American Economic Review*, Vol. 109, pp. 1–47.

Figure 1: SAT Scores and Major Percentile Return by Race: State School Sample

Panel A. First Year Major

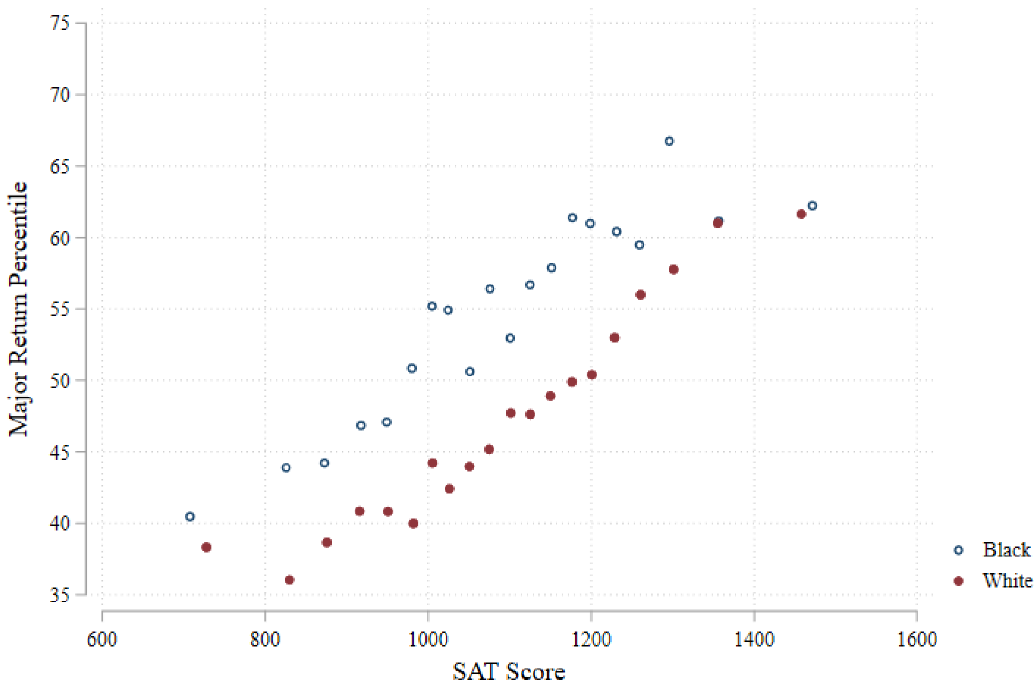


Panel B. Graduation Major



Source - Multiple-Institution Database for Investigating Engineering Longitudinal Development (MIDFIELD)
Notes - The state schools sample includes all black and white students with observed SAT scores at Clemson, Colorado, Colorado State, Florida, Florida State, Georgia Tech, North Carolina State, North Carolina – Charlotte, Oklahoma, Purdue, Utah State, and Virginia Tech. The sample includes students who entered college between 1987 and 2018, with incomplete time coverage for some institutions. Students over age 30 and those not identified as either Black or White are excluded from the sample.

Figure 2: SAT Scores and Major Percentile Return by Race: B&B Sample



Source - U.S. Department of Education, National Center for Education Statistics, 2008/18 Baccalaureate and Beyond Longitudinal Study (B&B:08/18)

Source - The B&B sample is a nationally representative survey of 2007-2008 college graduates. Students over age 30 and those not identified as either Black or White are excluded from the sample.

Table 1: Summary Statistics, State Schools Sample

	White (1)	Black (2)	T-Test P-Value (3)
Female	0.475	0.562	0.000
Transfer Student	0.151	0.123	0.000
Year Entered College	2001.7	2000.2	0.000
High School GPA	3.003	2.934	0.000
SAT Score	1144.6	1033.4	0.000
First-Year College GPA	2.848	2.459	0.000
College GPA at Graduation	2.899	2.441	0.000
First Major Wage Return	-0.0029	0.0117	0.000
First Major Percentile Return	0.497	0.516	0.000
First Major STEM Courses	0.354	0.352	0.006
Graduation Major Wage Return	0.0363	0.0251	0.000
Graduation Major Percentile Return	0.551	0.535	0.000
Graduation Major STEM Courses	0.333	0.302	0.000
Graduated College	0.489	0.394	0.000
Chemistry Major	0.010	0.011	0.001
Biology Major	0.073	0.092	0.000
Social Science Major	0.042	0.053	0.000
Communications Major	0.040	0.042	0.010
Business Major	0.126	0.130	0.010
Liberal Arts Major	0.168	0.171	0.043
Engineering Major	0.178	0.163	0.000
History Major	0.012	0.007	0.000
English Major	0.019	0.015	0.000
Education Major	0.041	0.037	0.000
Agriculture Major	0.031	0.012	0.000
Observations	873,662	60,786	

Source - Multiple-Institution Database for Investigating Engineering Longitudinal Development (MIDFIELD)
Notes - The state schools sample includes all black and white students with observed SAT scores at Clemson, Colorado, Colorado State, Florida, Florida State, Georgia Tech, North Carolina State, North Carolina – Charlotte, Oklahoma, Purdue, Utah State, and Virginia Tech. The sample includes students who entered college between 1987 and 2018, with incomplete time coverage for some institutions. Students over age 30 and those not identified as either Black or White are excluded from the sample.

Table 2: Descriptive Statistics, American Community Survey Sample

	White	Black	T-Test P-Value
	(1)	(2)	(3)
Female	0.465	0.628	0.000
Age	43.43	42.99	0.000
Log Earnings	11.17	10.91	0.000
Major Wage Return	-0.0022	-0.0308	0.000
Major Percentile Return	0.504	0.470	0.000
Major STEM Courses	0.278	0.246	0.000
Observations	2,585,094	200,428	

Source - U.S. Census Bureau, 2011-2021 American Community Survey, Public Use Microdata

Notes - The ACS sample includes working age (16 to 64) native non-Hispanic black and white college graduates who were employed full time in the previous year. Survey years 2011 through 2021 are included with the year 2020 excluded. Log earnings is the log of the sum of wage income and salary income in 2020 dollars.

Table 3: Summary Statistics, Baccalaureate and Beyond Sample

	White	Black	T-Test P-Value
	(1)	(2)	(3)
Female	0.578	0.674	0.000
SAT Score	1099.2	949.0	0.000
GPA at graduation	3.347	3.089	0.000
Age at graduation	22.80	22.99	0.000
First Generation Student	0.412	0.604	0.000
Major Wage Return	-0.0194	-0.0117	0.210
Major Return Percentile	0.476	0.491	0.066
Major STEM Courses	0.340	0.326	0.051
Log Earnings 2009	10.16	10.10	0.025
Log Earnings 2012	10.58	10.50	0.000
Log Earnings 2018	11.08	10.94	0.000
Observations	10,420	1,200	

Source - U.S. Department of Education, National Center for Education Statistics, 2008/18 Baccalaureate and Beyond Longitudinal Study (B&B:08/18)

Notes - The B&B sample includes 2007-2008 college graduates with follow-up surveys at 1, 4, and 10 years after graduation. Students over age 30 and those not identified as either Black or White are excluded from the sample.

Table 4: Major Selection by Race and SAT Score

	State Schools				B&B	
	1st-Yr. Major		Grad. Major		Grad. Major	
	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A: Major Wage Return					
Black	0.032***	0.037***	0.030***	0.037***	0.053***	0.067***
	(0.002)	(0.002)	(0.003)	(0.003)	(0.006)	(0.009)
Black × SAT		0.005***		0.008***		0.008**
		(0.001)		(0.001)		(0.003)
	Panel B: Major Percentile Return					
Black	0.042***	0.049***	0.037***	0.047***	0.076***	0.094***
	(0.003)	(0.003)	(0.004)	(0.004)	(0.009)	(0.012)
Black × SAT		0.007***		0.010***		0.011**
		(0.001)		(0.001)		(0.004)
	Panel C: Major STEM Courses					
Black	0.029***	0.034***	0.018***	0.027***	0.042***	0.060***
	(0.002)	(0.002)	(0.003)	(0.003)	(0.008)	(0.013)
Black × SAT		0.005***		0.009***		0.010**
		(0.001)		(0.001)		(0.004)
Student Characteristics	X	X	X	X	X	X
SAT Fixed Effects	X	X	X	X	X	X
Institution x Start Year FE	X	X	X	X		
Carnegie Classificiation FE					X	X
Observations	934,448	934,448	450,987	450,987	11,530	11,530

Source - U.S. Department of Education, National Center for Education Statistics, 2008/18 Baccalaureate and Beyond Longitudinal Study (B&B:08/18) and Multiple-Institution Database for Investigating Engineering Longitudinal Development (MIDFIELD)

Notes - The outcomes variables are measures of major difficulty. The Panel A outcome variable is the average wage return from the ACS for white graduates by major. The Panel B outcome variable is the percentile ranking of the average wage return from the ACS for white graduates by major. The Panel C outcome variable is the fraction of course credits in STEM courses by major. Student characteristics in the state school sample include high school GPA, a female indicator, and transfer student indicator. Student characteristics in the B&B sample include an indicator for gender and student age at graduation. Each institution in the B&B sample is categorized into one of 17 Carnegie classifications. Students not identified as either black or white are excluded from the analysis. For the state school sample, standard errors clustered by institution and year of college entry are reported in parenthesis. For the B&B sample, standard errors are clustered by the institution: $*p < .1, **p < .05, ***p < .01$

Table 5: Adult Log Earnings by Graduation Major Selection and Race

	ACS			B&B	
	(1)	(2)	(3)	(4)	(5)
	Panel A: Major = Wage Return				
Black	-0.220***	-0.229***		-0.083***	-0.074***
	(0.016)	(0.016)		(0.016)	(0.014)
Major	0.866***	0.832***	0.833***	0.594***	0.586***
	(0.024)	(0.033)	(0.033)	(0.123)	(0.124)
Major × Black	-0.325***	-0.321***	-0.325***	-0.142**	-0.152**
	(0.052)	(0.051)	(0.052)	(0.071)	(0.072)
	Panel B. Major = Percentile Return				
Black	-0.099***	-0.109***		-0.033	-0.019
	(0.013)	(0.013)		(0.029)	(0.026)
Major	0.649***	0.625***	0.625***	0.424***	0.417***
	(0.023)	(0.029)	(0.029)	(0.097)	(0.098)
Major × Black	-0.246***	-0.242***	-0.245***	-0.102*	-0.110**
	(0.038)	(0.037)	(0.038)	(0.052)	(0.053)
	Panel C: Major = STEM Courses				
Black	-0.182***	-0.193***		-0.044	-0.033
	(0.040)	(0.041)		(0.031)	(0.030)
Major	0.460***	0.452***	0.452***	0.380***	0.367***
	(0.080)	(0.084)	(0.084)	(0.111)	(0.113)
Major × Black	-0.121**	-0.118**	-0.118**	-0.080	-0.079
	(0.057)	(0.057)	(0.058)	(0.079)	(0.079)
Student Characteristics	X	X	X	X	X
Year FE	X	X	X	X	X
State FE		X		X	X
State × Race FE			X		
Carnegie Classification FE				X	X
SAT FE					X
Observations	2,650,399	2,650,399	2,650,399	26,350	26,350

Source - U.S. Department of Education, National Center for Education Statistics, 2008/18 Baccalaureate and Beyond Longitudinal Study (B&B:08/18) and U.S. Census Bureau, 2011-2021 American Community Survey, Public Use Microdata

Notes - The outcome variable is log earnings (wage and salary income). Major is defined as the graduation major average wage return from the ACS for white graduates in Panel A, as the percentile wage return from the ACS for white graduates in Panel B, and as the fraction of course credits in STEM courses by major in Panel C. Only students who identified as either black or white are included in the analysis. Student characteristics in the ACS sample include a female indicator and age indicators. Student characteristic controls for the B&B sample include a female indicator and age at graduation. In the B&B sample, log earnings are measured at 1, 4 and 10 years after graduation. Each institution in the B&B sample is categorized into one of 17 Carnegie classifications. Standard errors clustered by the graduation major are reported in parenthesis: $*p < .1, **p < .05, ***p < .01$

Table 6: Graduation Major Selection by Race, SAT Score, and Neighborhood Characteristics, State School Sample

	State Schools			
	(1)	(2)	(3)	(4)
Black	0.042***	0.043***	0.043***	0.042***
	(0.004)	(0.004)	(0.004)	(0.004)
Black × SAT	0.009***	0.008***	0.008***	0.007***
	(0.002)	(0.002)	(0.002)	(0.002)
Median Income (10,000s)		0.001***		
		(0.000)		
Median Income × SAT		-0.001***		
		(0.000)		
Median Education			0.002***	
			(0.001)	
Median Education × SAT			-0.001***	
			(0.000)	
Income Mobility				0.086***
				(0.018)
Income Mobility × SAT				-0.090***
				(0.013)
Student Characteristics	X	X	X	X
SAT FE	X	X	X	X
Institution x Start Year FE	X	X	X	X
Observations	311,520	311,520	311,520	311,520

Source - Multiple-Institution Database for Investigating Engineering Longitudinal Development (MIDFIELD)
Notes - The outcome is graduation major average wage return from the ACS for white graduates by major. Only students who identified as either black or white are included in the analysis. Student characteristics include high school GPA, a female indicator, and transfer student indicator. Median household income and median education are measured at the ZCTA level and income mobility is measured at the county level. Standard errors clustered by institution and year of college entry are reported in parenthesis: $*p < .1, **p < .05, ***p < .01$

Table 7: Graduation Major Selection by Race, SAT Score, and Neighborhood Characteristics, Baccalaureate and Beyond Sample

	B&B			
	(1)	(2)	(3)	(4)
Black	0.069***	0.068***	0.068***	0.069***
	(0.009)	(0.009)	(0.009)	(0.009)
Black × SAT	0.007*	0.007*	0.007*	0.007*
	(0.004)	(0.004)	(0.004)	(0.004)
Median Income (10,000s)		-0.001		
		(0.001)		
Median Income × SAT		-0.000		
		(0.000)		
Median Education			-0.003*	
			(0.001)	
Median Education × SAT			-0.000	
			(0.000)	
Income Mobility				-0.033
				(0.056)
Income Mobility × SAT				-0.005
				(0.026)
Student Characteristics	X	X	X	X
SAT FE	X	X	X	X
Carnegie Classification FE	X	X	X	X
Observations	8,500	8,500	8,500	8,500

Source - U.S. Department of Education, National Center for Education Statistics, 2008/18 Baccalaureate and Beyond Longitudinal Study (B&B:08/18)

Notes - The outcome is graduation major average wage return from the ACS for white graduates by major. Only students who identified as either black or white are included in the analysis. Student characteristics include an indicator for gender and student age at graduation. Each institution in the B&B sample is categorized into one of 17 Carnegie classifications. Median household income and median education, are measured at the ZCTA level and income mobility is measured at the county level. Standard errors clustered by the institution are reported in parenthesis: $*p < .1, **p < .05, ***p < .01$

Table 8: Log Earnings by Graduation Major Selection, Race, and Neighborhood Characteristics

	B&B				
	(1)	(2)	(3)	(4)	(5)
Black	-0.088***	-0.069***	-0.079***	-0.086***	-0.062***
	(0.022)	(0.022)	(0.022)	(0.021)	(0.019)
Major	0.626***	0.541***	0.552***	0.628***	0.562***
	(0.117)	(0.139)	(0.131)	(0.175)	(0.149)
Black × Major	-0.297***	-0.304***	-0.296***	-0.291***	-0.273***
	(0.084)	(0.080)	(0.082)	(0.079)	(0.084)
Median Income (10,000s)		0.017***			0.017***
		(0.002)			(0.002)
Median Income × Major		0.010			0.006
		(0.008)			(0.008)
Median Education			0.020***		
			(0.004)		
Median Education × Major			0.009		
			(0.010)		
Income Mobility				0.974***	
				(0.134)	
Income Mobility × Major				-0.039	
				(0.722)	
Student Characteristics	X	X	X	X	X
Year FE	X	X	X	X	X
Carnegie Classification FE	X	X	X	X	X
SAT FE					X
Observations	21,920	21,920	21,920	21,920	21,920

Source - U.S. Department of Education, National Center for Education Statistics, 2008/18 Baccalaureate and Beyond Longitudinal Study (B&B:08/18)

Notes - The outcome variable is log earnings measured at either 1, 4 or 10 years after graduation. Major is defined as the graduation major average wage return from the ACS for white graduates. Median household income and median education, are measured at the ZCTA level and income mobility is measured at the county level. Each institution in the B&B sample is categorized into one of 17 Carnegie classifications. Standard errors clustered by the institution are reported in parenthesis: $*p < .1, **p < .05, ***p < .01$

Table 9: Effect of Major Choice on Log Earnings, Heterogeneous Effects, American Community Survey Sample

	Gender		Age		
	Male	Female	< 30 Yrs.	31-50 Yrs.	51+ Yrs.
	(1)	(2)	(3)	(4)	(5)
	Panel A: Major Wage Return				
Major	0.873***	0.760***	0.727***	0.864***	0.840***
	(0.037)	(0.053)	(0.083)	(0.030)	(0.050)
Major × Black	-0.280***	-0.232***	-0.285***	-0.337***	-0.332***
	(0.063)	(0.051)	(0.047)	(0.070)	(0.058)
	Panel B: Major Percentile Return				
Major	0.668***	0.560***	0.547***	0.648***	0.630***
	(0.024)	(0.043)	(0.068)	(0.023)	(0.042)
Major × Black	-0.222***	-0.170***	-0.211***	-0.253***	-0.253***
	(0.042)	(0.041)	(0.036)	(0.050)	(0.042)
	Panel C: Major STEM Courses				
Major	0.452***	0.419***	0.360**	0.461***	0.491***
	(0.084)	(0.116)	(0.150)	(0.083)	(0.088)
Major × Black	-0.044	-0.056	-0.132**	-0.095	-0.150**
	(0.052)	(0.052)	(0.050)	(0.066)	(0.059)
Female Indicator			X	X	X
Year Fixed Effect	X	X	X	X	X
Age Fixed Effects	X	X	X	X	X
State X Black Fixed Effects	X	X	X	X	X
Observations	1,371,887	1,278,512	465,175	1,352,932	832,292

Source - U.S. Census Bureau, 2011-2021 American Community Survey, Public Use Microdata

Notes - Data from the American Community Survey. Outcome is log wage and salary income in 2020 real dollars. Robust standard errors clustered at the major level reported in parentheses. Column (1) includes only male workers. Column (2) includes only female workers. Column (3) includes only workers less than thirty years old. Column (4) includes only workers 31 to 50 years old. Column (5) includes only workers over fifty-one years old. Standard errors clustered by the graduation major are reported in parenthesis: $*p < .1, **p < .05, ***p < .01$

Online Appendix

A Proofs of Main Results

A.1 Proposition 1

Proof. A similar proof is provided in Lang and Manove (2011). For the first part of the proposition, suppose $M_k(a_L) < M^*(a_L)$. As $A_k(m)$ is monotonic in m, an increase in m will raise $f(A_k(m), m)$ and $f(a, m)$. Thus increasing m is strictly preferred.

Now, suppose that $M_k(a_L) > M^*(a_L)$. By definition, $M_k(a_L)$ can only provide higher expected utility than $M^*(a_L)$ if $f(A_k(M_k(a_L)), M_k(a_L)) > f(A_k(M^*(a_L)), M^*(a_L))$. But in equilibrium, beliefs must be correct, so $f(A_k(M_k(a_L), M_k(a_L)) = f(a_L, M_k(a_L))$. Since $\frac{\partial f(a,m)}{\partial m} < 0$ when $m \geq M^*(a)$ and in any well-behaved equilibrium employers believe that all students who choose $m < M_k(a_L)$ have aptitude a_L, students could deviate to $M^*(a_L)$ and increase their expected utility.

For the second part of the proposition, suppose $M_k(a') < M^*(a')$. As $A_k(m)$ is monotonic in m, an increase in m will raise $f(A_k(m), m)$ and $f(a, m)$. Thus increasing m is strictly preferred. $\square$

A.2 Proposition 2

Proof. By applying the chain rule to (6) and recognizing that in equilibrium $f(A_k(m), m) = f(a, m)$, we arrive at a first order condition of

$$\frac{\partial f(a,m)}{\partial m} + \lambda_k \frac{\partial f(a,m)}{\partial a} \frac{\partial A_k(m)}{\partial m} = 0.$$

Note that since $A_k(m) = M_k^{-1}(a)$, $\frac{\partial A_k(m)}{\partial m} = \frac{\partial M_k(a)}{\partial a}^{-1}$. Rearranging terms then proves the proposition. $\square$

A.3 Proposition 3

Proof. At a_L, we know from $M_b(a_L) = M_w(a_L)$ from Proposition 1. As $\lambda_b > lambda_w$, we know from Propositions 2 that $\frac{\partial M_b(a_L)}{\partial a} > \frac{\partial M_w(a_L)}{\partial a}$ and an ε increase in a will lead to $M_b(a+\varepsilon) > M_w(a+\varepsilon)$.

Now suppose that there was some $a' > a_L$ for which $M_w(a') \geq M_b(a')$. Since M is continuous, it then must be the case that in some ball around a' there is an $a'' < a'$ for which $\frac{\partial M_w(a,w_w)}{\partial a} > \frac{\partial M_b(a,w_b)}{\partial a}$. But as the major choices approach equality, $\frac{\partial M_b(a,w_b)}{\partial a} > \frac{\partial M_w(a,w_w)}{\partial a}$ which is a contradiction. $\square$

A.4 Proposition 4

Proof. First note that from Proposition 3, black students have lower aptitude, a, within m. They will thus have lower productivity, p, when entering the labor market. From equation (6) we can see that wages are simply employer's beliefs about student's human capital. $\square$

A.5 Proposition 5

Proof. A similar proof is provided in Lang and Manove (2011). First note from Proposition 1, $M_b(a_L) = M_w(a_L) \equiv M^*(a_L)$. Now consider the equilibrium observed return to human capital from major $m' > M^*(a_L)$,

$$\frac{f\left(A_k(m'), m'\right) - f\left(a_L, M^*(a_L)\right)}{m' - M^*(a_L)}.$$

It follows from Proposition 5 that $f(A_w(m'), m') > f(A_b(m'), m')$. This holds for any $m' > M^*(a_L)$. $\square$

A.6 Proposition 6

Proof. The proposition follows because $c(m_i)$ directly influences black student investment choice. Thus, for sufficiently low levels of $c(m_i)$, we can obtain qualitatively identical results to Proposition 3, and for sufficiently high levels of $c(m_i)$ we can obtain the opposite results.

For black students with aptitude a_L, not that if $c(M^*(a_L)) = 0$, $M_b(a_L) = M^*(a_L)$ as in Proposition 1. If $c(M^*(a_L)) > 0$ it then follows by the same arguments in the proof of Proposition 1 that $M_b(a_L) = \arg\max E_k(w|m, a_L) - c(m)$ which is lower than $M^*(a_L)$ as $c(m) > 0$.

Now, consider the modified first order condition for black students in this environment:

$$\frac{\partial f(a,m)}{\partial m} + \lambda_b \frac{\partial f(a,m)}{\partial a} \frac{\partial A_k(m)}{\partial m} = \frac{\partial c}{\partial m}, \tag{10}$$

where we again recognize that $f(A_k(m), m) = f(a, m)$ in equilibrium. Rearranging terms,

$$\frac{\partial M_b(a)}{\partial a} = \lambda_b \left[\frac{\partial c}{\partial m} - \frac{\partial f(a,m)}{\partial m}\right]^{-1} \frac{\partial f(a,m)}{\partial a}, \tag{11}$$

which follows as $\frac{\partial A_k(m)}{\partial m} = \frac{\partial M_b(a)}{\partial a}^{-1}$. Compared to the differential equation in Proposition 2, it is clear that an increase in $\frac{\partial c}{\partial m}$ reduces $\frac{\partial M_b(a)}{\partial a}$. Thus, for a sufficiently large $\frac{\partial c}{\partial m}$, black students may choose the same investments as whites, or choose less difficult investments. $\square$

A.7 Proposition 7

Proof. In equilibrium, $M_b(a)$ solves

$$\max_m E_b(w|M_b(a), a) + b(m) - c(m). \tag{12}$$

Thus, for m' to yield higher wages than $M_b(a)$ it must be the case that at m', $b(m') - c(m') < b(m) - c(m)$. $c(m)$. That is $M_b(a)$ provides higher utility than m' for workers with ability a because they receive a sufficiently large affirmative action benefit to offset their lower wages.

When assigned to $M_b(a)$, it follows from equation (6) that his expected wages are

$$E_b(w|M_b(a), a) = f(a, M_b(a)). \tag{13}$$

If instead assigned to $m' < M_b(a)$ he earns

$$E_b(w|m', a) = \lambda_b f(A_b(m'), m') + (1 - \lambda_b) f(a, m') \tag{14}$$

$\square$

A.8 Propostion 8

Proof. The first part of the proposition follows from Proposition 6. When $b(m) = 0$, $M_b(a)$ may be greater than, less than, or equal to $M^*(b)$. Since increasing $b(m)$ will unambiguously increase $M_b(a)$, $b_1(m)$ will lead to unambiguously lower black wages relative to $b_2(m)$ by reducing black human capital accumulation when under $b_2(m)$, $M_b(a) \geq M^*(a)$. However, when under $b_2(m)$, $M_b(a) < M^*(a)$ some positive values of $b(m)$ will raise black wages (by sending $M_b(a)$ closer to $M^*(a)$, while higher values of $b(m)$ may lower black wages (by sending $M_b(a)$ too far above $M^*(a)$.

For the second part of the proposition, suppose that black wages are higher under $b_2(m)$ than $b_1(m)$. Since $b_2(m)$ has a lower $M_b(a)$ function than $b_1(m)$ this an only be the case when lowering $M_b(a)$ raises black wages, which can only be the case if $M_b(a) > M^*(a)$ under $b_1(m)$. $\square$

B Supplementary Tables

Table B.1: Wage Return, Wage Percentile, and STEM Courses by Major

Major Code	Name	Wage Return	Wage Percentile	STEM
1100	General Agriculture	-0.192	20	0.261
1101	Agriculture Production and Management	-0.072	42	0.260
1102	Agricultural Economics	0.054	59	0.233
1103	Animal Sciences	-0.148	27	0.322
1104	Food Science	0.119	66	0.378
1105	Plant Science and Agronomy	-0.190	21	0.369
1106	Soil Science	-0.162	25	0.373
1199	Miscellaneous Agriculture	-0.218	18	0.184
1301	Environmental Science	-0.098	37	0.410
1302	Forestry	-0.147	28	0.322
1303	Natural Resources Management	-0.180	24	0.355
1401	Architecture	0.032	57	0.147
1501	Area, Ethnic, and Civilization Studies	0.000	55	0.132
1901	Communications	-0.043	47	0.122
1902	Journalism	-0.065	44	0.108
1903	Mass Media	-0.130	32	0.141
1904	Advertising and Public Relations	-0.003	54	0.126
2001	Communication Technologies	-0.151	27	0.297
2100	Computer and Information Systems	0.068	64	0.647
2101	Computer Programming and Data Processing	0.029	57	0.201
2102	Computer Science	0.260	84	0.714
2105	Information Sciences	0.173	73	0.294
2106	Computer Information Management and Security	0.039	58	-
2107	Computer Networking and Telecommunications	-0.032	49	-
2201	Cosmetology Services and Culinary Arts	-0.324	6	-
2300	General Education	-0.308	7	0.205
2301	Educational Administration and Supervision	-0.130	33	-
2303	School Student Counseling	-0.284	9	-
2304	Elementary Education	-0.373	3	0.164
2305	Mathematics Teacher Education	-0.190	21	0.551
2306	Physical and Health Education Teaching	-0.233	15	0.157
2307	Early Childhood Education	-0.418	1	0.145
2308	Science and Computer Teacher Education	-0.237	14	0.530
2309	Secondary Teacher Education	-0.254	11	0.219
2310	Special Needs Education	-0.291	8	0.113
2311	Social Science or History Teacher Education	-0.221	18	0.155
2312	Teacher Education: Multiple Levels	-0.358	4	0.196
2313	Language and Drama Education	-0.271	10	0.104
2314	Art and Music Education	-0.312	6	0.065
2399	Miscellaneous Education	-0.232	16	0.183
2400	General Engineering	0.233	82	0.713
2401	Aerospace Engineering	0.393	98	0.738
2402	Biological Engineering	0.211	78	0.676
2403	Architectural Engineering	0.242	83	0.763
2404	Biomedical Engineering	0.375	96	0.771
2405	Chemical Engineering	0.408	99	0.797
2406	Civil Engineering	0.259	83	0.794
2407	Computer Engineering	0.385	96	0.508
2408	Electrical Engineering	0.355	94	0.530
2409	Engineering Mechanics, Physics, and Science	0.301	90	0.821
2410	Environmental Engineering	0.185	75	0.650
2411	Geological and Geophysical Engineering	0.271	85	-
2412	Industrial and Manufacturing Engineering	0.315	91	0.742
2413	Materials Engineering and Materials Science	0.271	85	0.513
2414	Mechanical Engineering	0.338	92	0.786
2415	Metallurgical Engineering	0.346	93	0.309
2416	Mining and Mineral Engineering	0.339	92	0.837
2417	Naval Architecture and Marine Engineering	0.350	94	-
2418	Nuclear Engineering	0.361	95	0.693
2419	Petroleum Engineering	0.696	100	0.848
2499	Miscellaneous Engineering	0.221	79	0.712
2500	Engineering Technologies	-0.013	51	0.233
2501	Engineering and Industrial Management	0.206	77	0.420
2502	Electrical Engineering Technology	0.054	60	0.198
2503	Industrial Production Technologies	0.063	62	0.293
2504	Mechanical Engineering Related Technologies	0.076	65	0.448
2599	Miscellaneous Engineering Technologies	0.065	64	0.195
2601	Linguistics and Comparative Language and Literature	-0.064	44	0.139
2602	French, German, and Other Common Languages	-0.078	42	0.169
2603	Other Foreign Languages	-0.013	51	0.199
2901	Family and Consumer Sciences	-0.293	7	0.192
3201	Court Reporting	-0.240	14	-
3202	Pre-Law and Legal Studies	-0.097	39	0.123
3301	English Language and Literature	-0.096	40	0.168
3302	Composition and Speech	-0.226	17	0.150
3401	Liberal Arts	-0.161	26	0.166
3402	Humanities	-0.183	23	0.119
3501	Library Science	-0.385	2	0.182
3600	Biology	0.153	72	0.675
3601	Biochemical Sciences	0.261	84	0.669
3602	Botany	-0.120	36	0.612
3603	Molecular Biology	0.230	81	0.673
3604	Ecology	-0.203	20	0.615
3605	Genetics	0.135	68	0.661
3606	Microbiology	0.182	74	0.630
3607	Pharmacology	0.275	87	-
3608	Physiology	0.123	68	0.574
3609	Zoology	0.147	70	0.604
3611	Neuroscience	0.197	77	0.602
3699	Miscellaneous Biology	-0.081	41	0.488
3700	Mathematics	0.143	69	0.590
3701	Applied Mathematics	0.291	88	0.756
3702	Statistics and Decision Science	0.233	81	0.677
3801	Military Technologies	0.180	74	-
4000	Interdisciplinary Studies (General)	-0.244	13	0.212
4001	Intercultural and International Studies	-0.005	53	0.182
4002	Nutrition Sciences	-0.086	40	0.438
4005	Mathematics and Computer Science	0.276	87	0.649
4006	Cognitive Science and Biopsychology	0.273	86	0.182
4007	Interdisciplinary Social Sciences	-0.147	29	0.331
4101	Physical Fitness, Parks, Recreation, and Leisure	-0.139	32	0.188
4801	Philosophy and Religious Studies	-0.034	48	0.156
4901	Theology and Religious Vocations	-0.384	2	-
5000	Physical Sciences	0.046	58	0.750
5001	Astronomy and Astrophysics	0.154	72	0.693
5002	Atmospheric Sciences and Meteorology	0.064	63	0.726
5003	Chemistry	0.229	80	0.675
5004	Geology and Earth Science	-0.011	52	0.697
5005	Geosciences	0.096	66	0.771
5006	Oceanography	-0.060	45	0.746
5007	Physics	0.223	80	0.704
5008	Materials Science	0.332	91	-
5098	Multi-disciplinary or General Science	-0.016	50	0.370
5102	Nuclear, Industrial, and Biological Technologies	-0.033	48	0.363
5200	Psychology	-0.142	30	0.214
5201	Educational Psychology	-0.244	12	-
5202	Clinical Psychology	-0.180	24	-
5203	Counseling Psychology	-0.345	5	-
5205	Industrial and Organizational Psychology	0.047	59	0.403
5206	Social Psychology	-0.231	17	0.116
5299	Miscellaneous Psychology	-0.123	35	0.252
5301	Criminal Justice and Fire Protection	-0.127	34	0.099
5401	Public Administration	0.006	56	0.130
5402	Public Policy	0.299	90	0.299
5403	Human Services and Community Organization	-0.414	1	-
5404	Social Work	-0.348	5	0.113
5500	General Social Sciences	-0.184	22	0.199
5501	Economics	0.369	95	0.488
5502	Anthropology and Archeology	-0.145	29	0.157
5503	Criminology	-0.114	36	0.183
5504	Geography	-0.096	39	0.310
5505	International Relations	0.223	79	0.211
5506	Political Science and Government	0.189	76	0.163
5507	Sociology	-0.141	31	0.145
5599	Miscellaneous Social Sciences	0.064	62	-
5601	Construction Services	0.152	71	0.202
5701	Electrical and Mechanic Repairs and Technologies	-0.147	28	0.153
5901	Transportation Sciences and Technologies	0.122	67	0.165
6000	Fine Arts	-0.271	10	0.095
6001	Drama and Theater Arts	-0.269	11	0.088
6002	Music	-0.231	16	0.086
6003	Visual and Performing Arts	-0.284	9	0.092
6004	Commercial Art and Graphic Design	-0.184	22	0.129
6005	Film, Video and Photographic Arts	-0.165	25	0.105
6006	Art History and Criticism	-0.127	33	0.091
6007	Studio Arts	-0.365	3	0.107
6099	Miscellaneous Fine Arts	-0.216	19	0.186
6100	General Medical and Health Services	-0.047	47	0.264
6102	Communication Disorders Sciences and Services	-0.097	38	0.092
6103	Health and Medical Administrative Services	-0.126	35	0.000
6104	Medical Assisting Services	-0.057	46	-
6105	Medical Technologies Technicians	-0.066	43	0.218
6106	Health and Medical Preparatory Programs	0.389	97	0.366
6107	Nursing	-0.001	54	0.124
6108	Pharmacy Sciences, and Administration	0.400	98	0.112
6109	Treatment Therapy Professions	-0.059	46	0.119
6110	Community and Public Health	-0.098	37	0.281
6199	Miscellaneous Health Medical Professions	-0.243	13	0.296
6200	General Business	0.059	61	0.193
6201	Accounting	0.148	70	0.174
6202	Actuarial Science	0.465	99	0.616
6203	Business Management and Administration	0.005	55	0.208
6204	Operations, Logistics and E-Commerce	0.161	73	0.355
6205	Business Economics	0.283	88	0.449
6206	Marketing and Marketing Research	0.056	61	0.191
6207	Finance	0.294	89	0.219
6209	Human Resources and Personnel Management	-0.068	43	0.171
6210	International Business	0.140	69	0.190
6211	Hospitality Management	-0.141	31	0.158
6212	Management Information Systems and Statistics	0.193	76	0.273
6299	Misc Business and Medical Administration	0.071	65	0.199
6402	History	-0.006	53	0.144
6403	United States History	-0.030	50	-

Notes - The wage return and the percentile wage return are calculated from the ACS working age (16 to 64) native non-Hispanic black and white college graduates who were employed full time in the previous year. Survey years 2011 through 2021 are included with the year 2020 excluded. The major STEM content is calculated from the state school sample and is the fraction of course credits in STEM courses by major.

Table B.2: Major Selection by Race and SAT Score: Alternative Measures of Major Difficulty, State School Sample

	All		White male	
	Wage	Pctl	Wage	Pctl
	(1)	(2)	(3)	(4)
Black	0.035***	0.044***	0.034***	0.049***
	(0.003)	(0.004)	(0.003)	(0.004)
Black × SAT	0.008***	0.011***	0.005***	0.007***
	(0.001)	(0.001)	(0.001)	(0.001)
Student Characteristics	X	X	X	X
SAT Fixed Effects	X	X	X	X
Institution x Start Year FE	X	X	X	X
Observations	450,987	450,987	450,987	450,987

Source - Multiple-Institution Database for Investigating Engineering Longitudinal Development (MIDFIELD)

Notes - The outcomes variables are alternative measures of major difficulty. Column (1) uses the Wage Return difficulty measure computed using all prime age workers. Column (2) uses the Percentile Return difficulty measure computed using all prime age workers. Column (3) uses the Wage Return difficulty measure computed using white male prime age workers. Column (4) uses the Percentile Return difficulty measure computed using white male prime age workers. Student characteristics include high school GPA, a female indicator, and transfer student indicator. Students not identified as either black or white are excluded from the analysis. Standard errors clustered by institution and year of college entry are reported in parenthesis: $*p < .1, **p < .05, ***p < .01$

Table B.3: Major Selection by Race and SAT Score: Alternative Measures of Major Difficulty, Baccalaureate and Beyond Sample

	All		White male	
	Wage	Pctl	Wage	Pctl
	(1)	(2)	(3)	(4)
Black	0.064***	0.090***	0.068***	0.100***
	(0.009)	(0.012)	(0.008)	(0.012)
Black × SAT	0.008**	0.011**	0.006*	0.008*
	(0.003)	(0.004)	(0.003)	(0.005)
Student Characteristics	X	X	X	X
SAT Fixed Effects	X	X	X	X
Carnegie Classificiation FE	X	X	X	X
Observations	11,600	11,600	11,600	11,600

Source - U.S. Department of Education, National Center for Education Statistics, 2008/18 Baccalaureate and Beyond Longitudinal Study (B&B:08/18)

Notes - The outcomes variables are alternative measures of major difficulty. Column (1) uses the Wage Return difficulty measure computed using all prime age workers. Column (2) uses the Percentile Return difficulty measure computed using all prime age workers. Column (3) uses the Wage Return difficulty measure computed using white male prime age workers. Column (4) uses the Percentile Return difficulty measure computed using white male prime age workers. Student characteristics include an indicator for gender and student age at graduation. Each institution is categorized into one of 17 Carnegie classifications. Students not identified as either black or white are excluded from the analysis. Standard errors clustered by institution are reported in parenthesis: $*p < .1, **p < .05, ***p < .01$

Table B.4: Adult Log Earnings by Graduation Major and Race: Alternative Measures of Major Difficulty, American Community Survey Sample

	All		White male	
	Wage	Pctl	Wage	Pctl
	(1)	(2)	(3)	(4)
Major	0.861***	0.630***	0.895***	0.614***
	(0.031)	(0.026)	(0.044)	(0.034)
Major × Black	-0.329***	-0.252***	-0.357***	-0.249***
	(0.050)	(0.035)	(0.049)	(0.034)
Worker Characteristics	X	X	X	X
Year FE	X	X	X	X
State × Race FE	X	X	X	X
Observations	2,701,293	2,701,293	2,701,293	2,701,293

Source - U.S. Census Bureau, 2011-2021 American Community Survey, Public Use Microdata
Notes - Robust standard errors clustered at the major level in parenthesis. Worker characteristics include age fixed effects and a gender indicator. Column (1) uses the Wage Return difficulty measure computed using all prime age workers. Column (2) uses the Percentile Return difficulty measure computed using all prime age workers. Column (3) uses the Wage Return difficulty measure computed using white male prime age workers. Column (4) uses the Percentile Return difficulty measure computed using white male prime age workers. Standard errors clustered by graduation major are reported in parenthesis: $*p < .1, **p < .05, ***p < .01$

Table B.5: Adult Log Earnings by Graduation Major and Race: Alternative Measures of Major Difficulty, Baccalaureate and Beyond Sample

	All		White male	
	Wage	Pctl	Wage	Pctl
	(1)	(2)	(3)	(4)
Major	0.584***	0.415***	0.534***	0.362***
	(0.129)	(0.097)	(0.139)	(0.097)
Major × Black	-0.146*	-0.103*	-0.157**	-0.110**
	(0.074)	(0.053)	(0.074)	(0.051)
Worker Characteristics	X	X	X	X
Year FE	X	X	X	X
Carnegie Classificiation FE	X	X	X	X
Observations	26,350	26,350	26,350	26,350

Source - U.S. Department of Education, National Center for Education Statistics, 2008/18 Baccalaureate and Beyond Longitudinal Study (B&B:08/18)
Notes - Robust standard errors clustered at the major level in parenthesis. Column (1) uses the Wage Return difficulty measure computed using all prime age workers. Column (2) uses the Percentile Return difficulty measure computed using all prime age workers. Column (3) uses the Wage Return difficulty measure computed using white male prime age workers. Column (4) uses the Percentile Return difficulty measure computed using white male prime age workers. Standard errors clustered by graduation major are reported in parenthesis: $*p < .1, **p < .05, ***p < .01$

Table B.6: Major Selection and Log Earnings Regressions with Institution Fixed Effects

	Major Difficulty		Log Earnings	
	(1)	(2)	(3)	(4)
	Panel A: Major Wage Return			
Black	0.046***	0.054***	-0.070***	-0.072***
	(0.007)	(0.009)	(0.020)	(0.019)
Black × SAT		0.005		
		(0.003)		
Major			0.550***	0.557***
			(0.107)	(0.107)
Black × Major			-0.126	-0.128
			(0.081)	(0.086)
	Panel B: Major Percentile Return			
Black	0.066***	0.078***	-0.028	-0.028
	(0.009)	(0.012)	(0.033)	(0.033)
Black × SAT		0.007		
		(0.005)		
Major			0.391***	0.395***
			(0.084)	(0.084)
Black × Major			-0.087	-0.089
			(0.059)	(0.063)
	Panel C: Major STEM Courses			
Black	0.040***	0.050***	-0.034	-0.033
	(0.009)	(0.013)	(0.035)	(0.035)
Black × SAT		0.006		
		(0.005)		
Major			0.348***	0.350***
			(0.096)	(0.098)
Black × Major			-0.079	-0.082
			(0.085)	(0.087)
Student Characteristics	X	X	X	X
Institution FE	X	X	X	X
SAT Fixed Effects	X	X		X
Observations	11,460	11,460	26,350	26,350

Source - U.S. Department of Education, National Center for Education Statistics, 2008/18 Baccalaureate and Beyond Longitudinal Study (B&B:08/18)

Notes - For columns (1) and (2), the outcome variable for Panel A is the average wage return from the ACS for white graduates by major, for Panel B is the percentile ranking of the average wage return from the ACS for white graduates by major, and for Panel C is the fraction of course credits in STEM courses by major. For columns (3) and (4), the outcome variable is log earnings and the definition of the major variable is given in the panel title. Student characteristics include an indicator for gender and student age at graduation. Students not identified as either black or white are excluded from the analysis. Standard errors clustered by institution in columns (1) and (2) and clustered by graduation major in columns (3) and (4) are reported in parenthesis: $*p < .1, **p < .05, ***p < .01$

Table B.7: First-Year Major Selection (wage return) by Race, SAT, and Neighborhood Characteristics

	State Schools			
	(1)	(2)	(3)	(4)
Black	0.034***	0.033***	0.033***	0.034***
	(0.003)	(0.003)	(0.003)	(0.003)
Black × SAT	0.005***	0.004***	0.004***	0.004***
	(0.001)	(0.001)	(0.001)	(0.001)
Median Income (10,000s)		-0.001***		
		(0.000)		
Median Income × SAT		-0.000***		
		(0.000)		
Median Education			-0.002***	
			(0.000)	
Median Education × SAT			-0.000***	
			(0.000)	
Income Mobility				-0.028**
				(0.011)
Income Mobility × SAT				-0.035***
				(0.011)
Institution × Start Year FE	X	X	X	X
Student Characteristics	X	X	X	X
SAT Fixed Effects	X	X	X	X
Observations	620,470	620,470	620,470	620,470

Source - Multiple-Institution Database for Investigating Engineering Longitudinal Development (MIDFIELD)
Notes - The outcome is first-year major wage return from the ACS for white graduates by major. This table is similar to Table 6 where the outcome is the graduation major wage return. Only students who identified as either black or white are included in the analysis. Student characteristics include high school GPA, a female indicator, and transfer student indicator. Median household income and median education, are measured at the ZCTA level and income mobility is measured at the county level. Standard errors clustered by institution and year of college entry are reported in parenthesis: $*p < .1, **p < .05, ***p < .01$

Table B.8: Major Selection (wage percentile) by Race, SAT, and Neighborhood Characteristics

	State Schools					
	First Major			Graduation Major		
	(1)	(2)	(3)	(4)	(5)	(6)
Black	0.044***	0.044***	0.045***	0.055***	0.055***	0.054***
	(0.003)	(0.003)	(0.004)	(0.005)	(0.005)	(0.005)
Black × SAT	0.005***	0.005***	0.005***	0.010***	0.010***	0.009***
	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)
Median Income (10,000s)	-0.001***			0.002***		
	(0.000)			(0.000)		
Median Income × SAT	-0.000***			-0.001***		
	(0.000)			(0.000)		
Median Education		-0.001***			0.003***	
		(0.001)			(0.001)	
Median Education × SAT		-0.000***			-0.001***	
		(0.000)			(0.000)	
Income Mobility			-0.027*			0.128***
			(0.015)			(0.023)
Income Mobility × SAT			-0.053***			-0.111***
			(0.014)			(0.018)
Institution × Start Year FE	X	X	X	X	X	X
Student Characteristics	X	X	X	X	X	X
SAT Fixed Effects	X	X	X	X	X	X
Observations	620,470	620,470	620,470	311,520	311,520	311,520

Source - Multiple-Institution Database for Investigating Engineering Longitudinal Development (MIDFIELD)
Notes - The outcome is major wage percentile from the ACS for white graduates by major. This table is similar to Table 6 where the outcome is the graduation major wage return. Only students who identified as either black or white are included in the analysis. Student characteristics include high school GPA, a female indicator, and transfer student indicator. Median household income and median education, are measured at the ZCTA level and income mobility is measured at the county level. Standard errors clustered by institution and year of college entry are reported in parenthesis: $*p < .1, **p < .05, ***p < .01$

Table B.9: Major Selection (STEM Courses) by Race, SAT, and Neighborhood Characteristics

	State Schools					
	First Major			Graduation Major		
	(1)	(2)	(3)	(4)	(5)	(6)
Black	0.031***	0.031***	0.033***	0.028***	0.029***	0.031***
	(0.003)	(0.003)	(0.003)	(0.005)	(0.005)	(0.005)
Black × SAT	0.006***	0.006***	0.006***	0.010***	0.010***	0.010***
	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)
Median Income (10,000s)	-0.001***			-0.002***		
	(0.000)			(0.000)		
Median Income × SAT	-0.001***			-0.000***		
	(0.000)			(0.000)		
Median Education		-0.003***			-0.003***	
		(0.001)			(0.001)	
Median Education × SAT		-0.001***			-0.001***	
		(0.000)			(0.000)	
Income Mobility			-0.105***			-0.106***
			(0.016)			(0.023)
Income Mobility × SAT			-0.082***			-0.071***
			(0.010)			(0.014)
Institution × Start Year FE	X	X	X	X	X	X
Student Characteristics	X	X	X	X	X	X
SAT Fixed Effects	X	X	X	X	X	X
Observations	620,470	620,470	620,470	311,520	311,520	311,520

Source - Multiple-Institution Database for Investigating Engineering Longitudinal Development (MIDFIELD)
Notes - The outcome is the fraction of course credits in STEM courses in the major. This table is similar to Table 6 where the outcome is the graduation major wage return. Only students who identified as either black or white are included in the analysis. Student characteristics include high school GPA, a female indicator, and transfer student indicator. Median household income and median education, are measured at the ZCTA level and income mobility is measured at the county level. Standard errors clustered by institution and year of college entry are reported in parenthesis: $*p < .1, **p < .05, ***p < .01$

Table B.10: Major Selection (Other Measures) by Race, SAT, and Neighborhood Characteristics

	B&B					
	Percentile Return			STEM Courses		
	(1)	(2)	(3)	(4)	(5)	(6)
Black	0.095***	0.095***	0.096***	0.057***	0.058***	0.062***
	(0.013)	(0.013)	(0.013)	(0.014)	(0.014)	(0.014)
Black × SAT	0.009*	0.009*	0.009*	0.007	0.007	0.008
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Median Income (10,000s)	-0.001			-0.006***		
	(0.001)			(0.001)		
Median Income × SAT	-0.000			-0.001*		
	(0.000)			(0.000)		
Median Education		-0.003*			-0.013***	
		(0.002)			(0.002)	
Median Education × SAT		-0.000			-0.001**	
		(0.000)			(0.000)	
Income Mobility			-0.030			-0.252***
			(0.075)			(0.070)
Income Mobility × SAT			-0.006			-0.052
			(0.034)			(0.032)
Student Characteristics	X	X	X	X	X	X
SAT Fixed Effects	X	X	X	X	X	X
Carnegie Classification FE	X	X	X	X	X	X
Observations	8,500	8,500	8,500	8,460	8,460	8,460

Source - U.S. Department of Education, National Center for Education Statistics, 2008/18 Baccalaureate and Beyond Longitudinal Study (B&B:08/18)

Notes - The outcome is the fraction of course credits in STEM courses in the major. This table is similar to Table 6 where the outcome is the graduation major wage return. Only students who identified as either black or white are included in the analysis. Student characteristics include high school GPA, a female indicator, and transfer student indicator. Median household income and median education, are measured at the ZCTA level and income mobility is measured at the county level. Standard errors clustered by the institution are reported in parenthesis: $*p < .1, **p < .05, ***p < .01$

Table B.11: Log Earnings by Graduation Major Selection, Race, and Neighborhood Characteristics

	B&B					
	Major = Percentile Return			Major = STEM Courses		
	(1)	(2)	(3)	(4)	(5)	(6)
Black	0.046	0.029	0.025	0.003	-0.015	-0.022
	(0.034)	(0.033)	(0.033)	(0.043)	(0.043)	(0.043)
Major	0.382***	0.229**	0.432***	0.402***	0.158	0.509***
	(0.112)	(0.101)	(0.142)	(0.094)	(0.110)	(0.106)
Black × Major	-0.230***	-0.204***	-0.221***	-0.159	-0.127	-0.144
	(0.059)	(0.058)	(0.058)	(0.110)	(0.107)	(0.105)
Median Income (10,000s)	0.013***			0.018***		
	(0.003)			(0.003)		
Median Income × Major	0.008			0.002		
	(0.006)			(0.008)		
Median Education		0.003			0.010**	
		(0.004)			(0.005)	
Median Education × Major		0.027***			0.032***	
		(0.004)			(0.007)	
Income Mobility			0.912***			1.320***
			(0.344)			(0.217)
Income Mobility × Major			0.121			-0.791**
			(0.573)			(0.379)
Student Characteristics	X	X	X	X	X	X
Year FE	X	X	X	X	X	X
Carnegie Classification FE	X	X	X	X	X	X
Observations	21,930	21,930	21,930	21,810	21,810	21,810

Source - U.S. Department of Education, National Center for Education Statistics, 2008/18 Baccalaureate and Beyond Longitudinal Study (B&B:08/18)

Notes - The outcome variable is log earnings measured at either 1, 4 or 10 years after graduation. Major is defined in columns (1) - (3) as the major wage return percentile from the ACS for white graduates and in columns (4) - (6) as the fraction of credits from STEM courses by major. This table is similar to Table 8 where major is defined as the wage return for the major. Median household income and median education, are measured at the ZCTA level and income mobility is measured at the county level. Standard errors clustered by the graduation major are reported in parenthesis: $*p < .1, **p < .05, ***p < .01$