Math-Oriented Fields of Study and the Race Gap in Graduation Likelihoods at Elite Colleges

Supplementary Appendix

I. Results from multilevel logit models

Table SA1: Coefficients from logit multilevel models predicting six-year graduation likelihoods, by individual-level and major-level characteristics

Model #	(1)	(2)	(3)	(4)
Population:	All students		White students	Black students
Race (black=1)	-0.584**	-0.624**		
	(0.084)	(0.098)		
Math intensity (average math SAT score in major)	-0.003	-0.004	-0.004	-0.008**
	(0.002)	(0.002)	(0.002)	(0.002)
Math intensity^2		-0.000*	-0.000*	-0.000*
		(0.000)	(0.000)	(0.000)
Race*math intensity	-0.004	-0.003		
	(0.002)	(0.002)		
Race*math intensity^2		0.000		
		(0.000)		
Individual-level attributes	Yes	Yes	Yes	Yes
Intercept	2.447**	2.575**	2.683**	1.780**
	(0.312)	(0.314)	(0.350)	(0.675)
Observations	23,637	23,637	21,625	2,012
Number of major	26	26	26	26
Rho	0.0366	0.0287	0.0351	9.03e-07
Sigma u	0.354	0.312	0.346	0.002
Log likelihood	-5485	-5482	-4660	-810.7

Standard errors are in parentheses.
** p<0.01, * p<0.05

II. Results from HLM Model using STEM vs. Non-STEM classification of majors

We examine the robustness of our results by estimating multilevel models that are comparable to those estimated in Table SA1, but that use a more common approach to the classification of fields: namely, STEM vs. non-STEM (Table S2). Figures SA1 and SA2 plot the predicted six-year graduation likelihoods of black and white students by race and field of study based on these models, where the former is standardized to reflect white students' distribution of background characteristics, and the latter is standardized to reflect black students' distribution of background characteristics.

The results from this model are quite similar to those obtained under the math intensity classification scheme for majors, although the standard error for the coefficient of the interaction term between STEM and race is large relative to the coefficient. This is likely due to the curvilinear effect of math intensity on graduation likelihoods observed in Figures 1a and 1b, which is not captured by the STEM-based classification scheme. The predicted probabilities obtained from Model 1 in Table SA2 (Figure SA1 and SA2) are consistent with the results reported in Figures 1a and 1b: black students suffer a greater penalty in STEM fields relative to white students, under both standardizations. The race gap, consequently, is larger in STEM fields.

Table SA2: Coefficients from logit multilevel models predicting six-year graduation likelihoods, by individual-level and major-level characteristics

Model #:	(1)	(2)	(3)
Population:	All students	White students	Black students
Race (black=1)	-0.539**		
	(0.094)		
Fields of study: STEM fields =1	-0.433*	-0.447*	-0.571**
	(0.170)	(0.187)	(0.159)
Race*STEM fields	-0.104		
	(0.159)		
Individual-level attributes	Yes	Yes	Yes
Constant	2.553**	2.662**	1.841**
	(0.310)	(0.344)	(0.669)
Observations	23,637	21,625	2,012
Number of major	26	26	26
Rho	0.030	0.037	0.002
Sigma u	0.318	0.354	0.075
Log likelihood	-5485	-4661	-812

Standard errors are in parentheses.
** p<0.01, *p<0.05

Figure SA1: Predicted six-year graduation probabilities for black and white students, by race and attendance in STEM fields
(standardized to **white** students' distribution of social, academic and instituitonal characteristics)

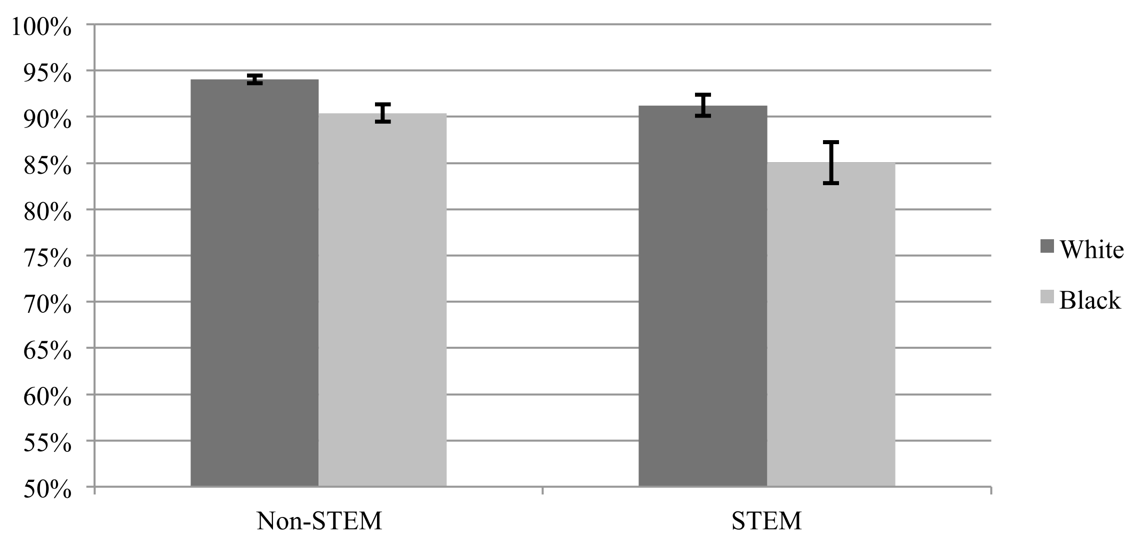
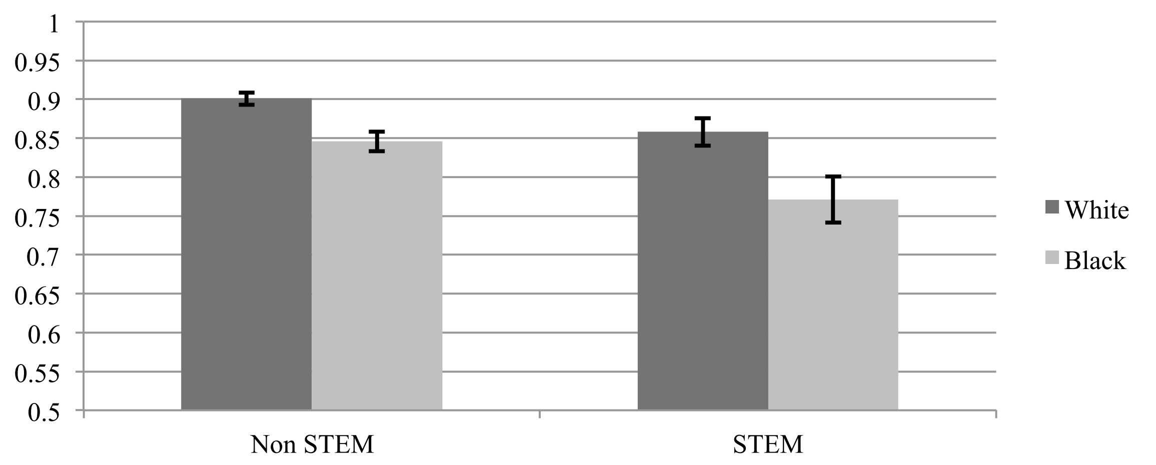


Figure SA2: Predicted 6-years graduation probabilities for black and white students by race and attendance in STEM field
(standardized to **black** students' distribution of social, academic and instituitonal characteristics)



III. Results from a weighted regression

We examine the possibility of negative selection to math-oriented fields by estimating weighted regressions that predict students' graduation likelihoods as a function of their chosen field of study—math-oriented versus non-math-oriented—for black and white students separately. This technique, developed primarily for the estimation of causal effects in observational data, helps uncover underlying patterns in the data by identifying possible heterogeneity in the effect of fields of study on graduation likelihood that may be related to the selection of students into different fields of study. One of the main benefits of this method is that it uses basic properties of regression—weights—in order to balance the sample and generate meaningful comparisons, which are subject to causal interpretation (see Morgan and Winship 2015: Chapter 7). Using weights based on students' propensity to undergo a treatment (in this case, major in math-oriented fields), this method estimates the effect of the treatment on a balanced sample of students. We use two such balanced samples: one that reflects the characteristics of students in the treatment group (which gives us the average treatment effect for the treated, or ATT), and one that reflects the characteristics of students in the control group (which gives us the average treatment effect for the control, or ATC). If the ATT is substantially different than the ATC, this may indicate an underlying selection on unobserved characteristics that is associated with the treatment effect. A greater negative effect of the ATT relative to the ATC would suggest negative selection, while a greater negative effect of the ATC relative to the ATT would suggest positive selection.

We estimate the weighted regressions in several steps: First, we use fitted logit models that predict field of study selection as a function of student background and institutional characteristics, estimated separately for black and white students, and use the propensity scores from these estimations to construct the ATT and ATC weights.[1,2] Next, we fit several logit models that predict black and white students' graduation likelihoods as a function of their field of study, using the ATT and ATC weights. These models also include the vector of social, academic and institutional variables, in order to account for any remaining imbalance in the samples. The coefficient for math-oriented fields in the ATT model denotes the expected penalty in graduation likelihood associated with majoring in a math-oriented field (versus other fields) for black and white students who are observationally similar to students that major in math-oriented fields. The coefficient for math-oriented fields in the ATC model denotes the expected

[1] The weights are a function of the probability (p1) of a subject being in a treatment or control group. The ATT weights for those in the control group (d=0) are a function of their inverse probability of being in the treatment group (p1/1-p1), while the ATT weights for those in the treatment group is 1. The opposite is done to calculate the ATC weights.

[2] We use the diagnostic routine developed by Morgan and Todd (2008) to evaluate the balance that the ATT and ATC weights achieved in the samples. As expected, both the black and white student samples in math-oriented and non-math oriented fields were highly unbalanced before the weights were applied. The average of standardized mean difference (hereafter, ASMD) was .137 for black students and .155 for white students. The ATT and ATC weights significantly improved the balance of the samples: Using the ATT weights, the ASMD decreased to .011 for blacks, and to .006 for whites. Using the ATC weights, the ASMD decreased to .028 for blacks, and to .011 for whites.

penalty in graduation likelihood associated with majoring in a math-oriented field for black and white students who are observationally similar to students that major in non-math-oriented fields. Table SA3 presents the coefficients for math-oriented fields, and Table SA4 shows the predicted graduation probabilities for black and white students based on the weighted regressions.

The results from the weighted regressions support those obtained from the multilevel models. While the graduation likelihoods of all students are negatively affected by enrolling in math-oriented fields, the disadvantage experienced by black students is larger. The results in Table SA4 indicate that the predicted graduation likelihoods of white students who enroll in math-oriented fields (ATT weights) are 2.2 percentage points lower than those of similar white students in non-math-oriented fields, everything else being equal. The penalty associated with enrolling in math-oriented fields (ATT weights, black students) is triple that size for black students, with a difference of 6.8 percentage points. The relative similarities in the predicted gaps across the different weighting schemas and in the magnitude of the coefficients for fields suggest that there is not much heterogeneity in the effect of math-oriented fields that is related to the propensity of enrolling in math-oriented fields.[3] Put differently, the penalty in graduation likelihoods associated with math-oriented fields, detected in Figures 1a and 1b, does not appear to be driven by negative unobserved selection.

[3] The reader should note that for black students, the average treatment effect for the control group is slightly smaller than the average treatment effect for the treated, which may indicate a slight negative selection to math-oriented fields among blacks. The opposite is true for white students, which may indicate positive selection among whites. Since the differences in magnitude are very small and the samples are not fully balanced, we do not interpret these patterns as indicative of selection, but rather attribute these differences to the remaining imbalance in the sample. See Chapter 7 in Morgan and Winship (2015) for a thorough discussion of the interpretation of the ATC results.

Table SA3: Coefficients from weighted logit regressions predicting six-year graduation likelihoods, by ATT and ATC weights and race

	White students		Black students	
	Math-oriented fields coefficient	se	Math-oriented fields coefficient	se
ATT weights: (Students with characteristics similar to those who major in math-oriented fields)	-0.341**	(0.0911)	-0.509**	(0.220)
ATC weights: (Students with characteristics similar to those who major in non-math-oriented fields)	-0.405**	(0.0705)	-0.485**	(0.195)

Source: C&B 1989
Notes: N= 2012 for black students; N=21,625 for white students. Robust standard errors are in parentheses.
** $p<0.05$

Table SA4: Predicted six-year graduation rates from weighted regressions, by enrollment in STEM fields

Population:	White students			Black students		
Fields:	Non-STEM	STEM	Difference	Non-STEM	STEM	Difference
ATT weights: (Students with characteristics similar to those who major in STEM fields)	93.6%	90.7%	2.9%	84.8%	78.1%	6.8%
ATC weights: (Students with characteristics similar to those who major in non-STEM fields)	93.2%	90.1%	3.1%	83.8%	77.5%	6.2%

Source: C&B 1989