NBER WORKING PAPER SERIES

RACE AND COLLEGE SUCCESS:
EVIDENCE FROM MISSOURI

Peter Arcidiacono
Cory Koedel

Working Paper 19188
http://www.nber.org/papers/w19188

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 2013

We thank Jeff Smith, Sarah Turner, and seminar and conference participants at the Institute for Research on Poverty Summer Workshop, NBER Education Meetings, the Society of Labor Economists, and the University of Missouri for helpful comments. We thank the Missouri Department of Higher Education for providing access to data. The usual disclaimers apply. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Race and College Success: Evidence from Missouri
Peter Arcidiacono and Cory Koedel
NBER Working Paper No. 19188
June 2013
JEL No. I23,J15

ABSTRACT

Conditional on enrollment, African American students are substantially less likely to graduate from 4-year public universities than white students. Using administrative micro data from Missouri, we decompose the graduation gap between African Americans and whites into four factors: (1) racial differences in how students sort to universities, (2) racial differences in how students sort to initial majors, (3) racial differences in school quality prior to entry, and (4) racial differences in other observed pre-entry skills. Pre-entry skills explain 65 and 86 percent of the gap for women and men respectively. A small role is found for differential sorting into college, particularly for women, and this is driven by African Americans being disproportionately represented at urban schools and the schools at the very bottom of the quality distribution.

Peter Arcidiacono
Department of Economics
201A Social Sciences Building
Duke University
Durham, NC 27708
and NBER
psarcidi@econ.duke.edu

Cory Koedel
Department of Economics
University of Missouri
118 Professional Building
Columbia, MO 65211
koedelc@missouri.edu

1. Introduction

At around 40 percent, six-year graduation rates for African Americans are over twenty percentage points lower than for whites (DeAngelo et al., 2011, National Center for Education Statistics, 2012). African American men in particular have low college enrollment rates coupled with high dropout rates (Aucejo, 2012). African Americans' low graduation rates are cause for concern given the substantial and well-documented returns to receiving a 4-year college degree (Heckman, Lochner, and Todd 2006; Bound and Turner 2011), particularly for African Americans (Arcidiacono 2005; Arcidiacono, Bayer, and Hizmo 2010).

In this paper we decompose the differences in 4-year college completion rates between African Americans and whites conditional on college enrollment. We seek to understand how much of the racial disparities are due to each of four factors. First is where students enroll in college. Whites and African Americans attend colleges of different qualities – due to some combination of preferences, access to resources and information, and pre-entry skills – and this may in turn affect persistence. Higher quality colleges may produce higher persistence rates for all students, or certain schools may be a better match for students of certain abilities.[1] Second is what majors students pursue upon college entry. Conditional on pre-entry skills African Americans are more likely to initially pursue STEM majors, which have lower graduation rates. Third, whites and African Americans attend very different schools prior to postsecondary entry, which can lead to differences in academic preparation through access to, for example, more advanced courses. Fourth, whites and African Americans differ in other observed measures of academic preparation.

[1] Dillon and Smith (2012) discuss the determinants and consequences of being under and overmatched with one's university. The match between the student and the school has received attention in the literature on mismatch and how it relates to affirmative action. See, for example, Arcidiacono, Aucejo, Coate, and Hotz (2012) and Arcidiacono, Aucejo, and Hotz (2013). See Arcidiacono, Aucejo, Fang, and Spenner (2011) for a discussion of how mismatch can arise with students making rational enrollment decisions.

We take advantage of administrative data from the system of four-year public universities in Missouri in order to understand these educational disparities. These data offer several advantages. One advantage is that we have panel data on the entire system, which allows us to examine a wide range of college qualities – the graduation rates across the universities in the system range from 30 to 80 percent. A second advantage is that, in addition to measures of pre-entry skills such as high school class ranks and ACT scores, we know the high schools from which students graduated and have large numbers of students from each high school. The information on high school of attendance is important as it improves the value of information on students' class ranks, which are inherently relative measures of student achievement (also see Fletcher and Tienda, 2010). Finally, the data allow us to track students throughout the entire system over time, and provide information on initial major.

We specify a flexible logit model of the probability of graduating college within eight years conditional on attending particular postsecondary institutions and entering with particular majors, allowing for match effects between students and school-major combinations (Arcidiacono, Aucejo, and Hotz 2012; Light and Strayer 2000). The model allows graduation gaps between African Americans and whites to arise as a result of each of the four factors described above. We examine the importance of differences in university sorting between African Americans and whites by first estimating a multinomial logit to predict the probabilities of white students attending each of the universities in the system given their pre-entry skills and conditional on entering major (STEM or non-STEM). We then use the parameters from the sorting model estimated on white students to assign African Americans to counterfactual universities. These counterfactual university assignments reflect sorting patterns for African Americans if African Americans sorted to universities in the same way as observationally similar whites. Given the counterfactual assignment scenario and corresponding estimated graduation probabilities from the graduation model, we can assess how

African American graduation rates would change if African Americans sorted to universities in the same way as observationally similar whites. We use a similar procedure to construct counterfactual major assignments for African Americans.

Next we use empirical estimates of high school quality from the graduation model to quantify the importance of differences in high school quality between African Americans and whites. We predict the gains in African American graduation rates that would accompany an improvement in high school quality so that African Americans and whites attend high schools of the same quality. Finally, we predict the effect of raising African Americans' class ranks and entrance exam scores to align them with whites'.

We find that African American graduation rates, in total, would increase by 1.8 percentage points (from a base of approximately 48 percent) under the counterfactual sorting scenario where African Americans sort to universities like observationally similar whites. This improvement comes with no change in African Americans' observed pre-entry skills or high school quality. The gains from counterfactual sorting are larger for African American women than for men, explaining 14.7 and 5.6 percent of the racial graduation gaps by gender, respectively. Despite initial STEM majors having lower graduation rates, the role of major sorting is negligible for both genders, explaining 1.5 percent of the gap for women and 0.3 percent of the gap for men.

A number of recent studies have raised concerns about the processes by which African Americans sort to universities (e.g., see Bowen, Chingos and McPherson, 2009; Hoxby and Avery, 2012; Hoxby and Turner, 2013; Roderick et al., 2008), with a common theme being that African Americans tend to "undermatch" (i.e., choose universities that are less selective than the universities that they are qualified to attend). While much of the literature on the returns to college quality for African Americans is couched within the context of affirmative action policies, and consequently focuses primarily on highly-selective universities (Arcidiacono, 2005; Card and Krueger, 2005;

Howell, 2010; Long, 2004a/b), we show that differences in enrollment patterns between African Americans and whites across groups of less prestigious colleges are the primary drivers behind the counterfactual sorting gains. In particular, it is moving African Americans out of urban schools and the very bottom schools that result in the graduation gains.

The remaining graduation gaps are explained by racial differences in high school quality and other observed pre-entry skills. For women and men respectively, differences in high school quality explain 18.5 and 8.4 percent of gap.[2] The disparity is owing to the fact that African American women who attend college are more likely to come from lower-quality high schools than their male counterparts. Pre-entry skill differences explain what is left of the graduation gap for each gender – 65.3 percent of the gap for women and 85.7 percent of the gap for men – indicating that most of the degree-attainment gap is the result of differences in pre-entry skills that emerge prior to adulthood.[3] The importance of pre-entry skills is consistent with the larger literature looking at black-white success gaps in other settings (e.g., see Cameron and Heckman, 2001; Neal and Johnson, 1996; Rivkin, 1996).

The rest of the paper proceeds as follows. Section 2 provides details about the Missouri higher education system as well as descriptive statistics by race for each of the universities. Section 3 presents our strategy for unbundling the racial gap in degree attainment. Section 4 examines the academic background measures that affect graduation rates and shows how well particular schools graduate students with different levels of preparation. Section 5 performs the decomposition, examining the roles of university sorting, major sorting, high school quality, and pre-entry skills in explaining the racial degree-attainment gap. Section 6 examines the mechanisms behind the gains to

[2] Both these estimates are likely upper bounds. See Section 3.5 for a discussion.
[3] Pre-entry skills explain an even larger fraction of the racial differences in graduation rates when the outcome measure is five-year or six-year graduation rates rather than eight-year graduation rates. See Section 6.3.

reassigning African Americans across colleges, evaluates the robustness of our results, and considers alternative outcome measures to eight-year graduation rates. Section 7 concludes.

2. Background and Data

The four-year public university system in Missouri consists of 13 campuses; Figure 1 shows their geographic locations. The major population centers in Missouri are in Kansas City and St. Louis and their surrounding areas (the metro areas are located in the middle of the state vertically, and at the western and eastern borders respectively). We divide the 13 universities in the public system into four broad groupings for expositional purposes:

> Group 1: The three most selective universities: Truman State University, the University of Missouri at Rolla and the University of Missouri at Columbia (the latter being the state flagship)
>
> Group 2: The two urban universities: the University of Missouri at Kansas City and the University of Missouri at St. Louis.
>
> Group 3: The four moderately-selective, non-urban universities: Missouri State University, Northwest Missouri State University, Southeast Missouri State University and the University of Central Missouri
>
> Group 4: The four least-selective universities: Missouri Southern State University, Western Missouri State University, Lincoln University and Harris Stowe State University (the latter two universities are historically black universities (HBUs)).

We use administrative data from the Missouri Department of Higher Education for our analysis. The data track students beginning with entry into the system, and subsequently on a semester-by-semester basis through potential graduation. We focus on African American and white students because the Hispanic and Asian populations in Missouri are small.[4] We restrict our analytic sample to include full-time, state-resident, non-transfer students who entered the public university

[4] The 2000 Missouri census reports that Missouri's Hispanic population share was 2.1 percent. The African American share was 11.7 percent, just below the national average of 12.9 percent. Asians are also underrepresented in Missouri, making up just 1.4 percent of the population.

system between 1996 and 2001 as college freshman.[5] We track students for up to eight years after initial entry into the system to determine whether they graduated at any of the four-year public institutions in Missouri. Graduation outcomes can be tracked regardless of the university from which the degree is obtained so long as students remain in the system.[6] Details about the construction of the analytic sample are provided in Appendix Table B.1.

The administrative data track students during college and also include detailed information about pre-entry qualifications. In addition to the standard entrance exam scores (from the ACT), two notable data elements are students' high schools of attendance and class ranks. Most records for state residents contain this information, including students from private schools.[7] An advantage of working with a large student sample from within a single state, rather than a nationally-representative but thinly-spread dataset (e.g., the NELS or NLSY), is that we can condition directly on students' high schools of attendance in our econometric models. We need not rely on proxies for high school quality, which prior research suggests may be insufficient (Roderick et al., 2008). Empirically, students' high schools of attendance and class ranks are strong predictors of success in college. A simple linear model that predicts students' college-graduation outcomes as a function of high school indicator variables and a continuous class rank variable, for example, explains 16 percent of the total variance in outcomes. Alternatively, a model that uses ACT math and reading scores instead explains less than a third as much; just 5 percent.[8]

---

[5] A small number of students who enter a university with sophomore status but no prior university experience are also included. These are students who have collected a full year's worth of college credits while in high school.

[6] Aggregate data from the Missouri Department of Higher Education indicate that approximately three percent of individuals enrolled in a system school transfer to an out-of-system four year university annually (i.e., private in-state, private out-of-state, public out-of-state). We cannot track graduation outcomes for out-of-system transfers. However, based on internal transfer data, we suspect that out-of-system transfers graduate at a lower rate than individuals who do not change universities, who comprise the bulk of our sample. Within the system, university changers graduate within eight years at a rate of 51 percent whereas individuals who do not change universities graduate at a rate of almost 65 percent. The disparity increases if look at five- and six-year graduation rates.

[7] Approximately 6 percent of in-state students do not have either an assigned high school or class ranking.

[8] These results are consistent with other studies showing that entrance exam scores have low predictive power over attainment outcomes– e.g., see Bowen, Chingos and McPherson (2009) and Fletcher and Tienda (2010). Rothstein (2004) shows that SAT scores are much weaker predictors of college GPAs than are high school GPAs.

Table 1 provides basic descriptive statistics for each university in the system relative to the system as a whole, and internally. The universities are ordered by the average value of a pre-entry preparation index for incoming students, which we use as a measure of selectivity.[9] Several features of the Missouri system are notable. Beginning with how enrollment is distributed across the system, forty percent of the students in the analytic sample enter into just two universities: the University of Missouri at Columbia and Missouri State University. Several universities have enrollment shares at or near 10 percent, and 5 of the 13 universities enroll fewer than five percent of the students in our data. The enrollment shares presented in Table 1 are not entirely representative of total enrollment shares because we exclude transfer students from community colleges as well as part-time students, and these students are not evenly distributed across the system. Still, the enrollment shares in Table 1 are broadly reflective of the relative sizes of the public universities in Missouri.

The third column of Table 1 shows the distribution of students who enter with intended majors in a science- or mathematics-related field (STEM) across universities. STEM majors include students who initially enter college with a major in the natural or physical sciences, engineering, computer science, mathematics, or economics. All other students are assigned as non-STEM majors.[10] STEM majors are heavily concentrated at the three most-selective institutions, which account for 58 percent of incoming STEM majors despite accounting for just 35 percent of total enrollment.

---

[9] The preparation index for each student is a weighted average of his/her ACT scores in math and reading, high school class rank, and high school quality where the weights are empirically determined. See Section 3 for details. Further information about selectivity and student sorting across universities is provided in Appendix C.

[10] Appendix Table B.2 provides information about the most common major codes for STEM and non-STEM majors. Economics is included with the STEM group for two reasons (1) ACT math scores for economics majors align much more closely with ACT math scores for STEM than non-STEM majors, and (2) the grade distributions in economics courses look similar to the grade distributions in STEM fields (Koedel, 2011). However, note that economics majors make up such a small share of the STEM-major group that excluding them from our analysis, or shifting them to the non-STEM group, does not affect our findings. Even more, below we consider models that ignore information about initial major entirely (so that designating entrants as STEM/non-STEM is irrelevant). Our results from those models are substantively similar to what we find in our main analysis; a common finding throughout our analysis is that major sorting is not an important factor in determining black-white gaps in graduation rates.

The fourth column shows the distribution of African American students across institutions. Comparing the African American shares to the total enrollment shares reveals their unconditional representation across the system. African Americans are substantially overrepresented at three of the four least selective institutions in the state and at the urban campuses. They are also mildly underrepresented at the two most-selective schools, essentially proportionally represented at the University of Missouri at Columbia (again, the state flagship), and underrepresented at the four moderately-selective non-urban schools.

The second vertical panel of the table provides internal descriptive statistics for each university to complete the system overview. Among the statistics provided is the eight-year graduation rate, which maps fairly closely to the pre-entry preparation index. The most notable differences between the entering index values and graduation rates occur at the urban campuses, which have much lower graduation rates than would be predicted by students' index values alone. The low graduation rates at the urban campuses are consistent with findings from Bowen, Chingos and McPherson (2009).[11] Also note the sharp drop in graduation rates at the four least-selective schools.

3. Model and Decomposition Procedure

We are interested in examining the importance of the four above-described factors in explaining racial differences in college graduation rates within the Missouri system. In particular, we want to understand the importance of racial differences in initial college attended, c, initial major, m, high school, h, and academic background, x. Conditional on attending a college in the Missouri system, the unconditional probability of a student of race r graduating can be expressed as:[12]

---

[11] Bowen, Chingos and McPherson (2009) find that graduation rates are negatively related to the commuter share. A distinguishing feature of the urban campuses is that they have larger commuter populations relative to the other universities in the system.

[12] All of our decompositions are done within gender. We do not condition on gender in equation (1) to conserve on notation.

$$\Pr(y=1|r) = \sum_{x\in X}\sum_{h\in H}\sum_{m\in M}\sum_{c\in C}\Pr(y=1|c,m,h,x,r)\Pr(c,m,h,x,r) \tag{1}$$

$$= \sum_{x\in X}\sum_{h\in H}\sum_{m\in M}\sum_{c\in C}\Pr(y=1|c,m,h,x,r)\Pr(c|m,h,x,r)\Pr(m|h,x,r)\Pr(h|x,r)\Pr(x|r)$$

Equation (1) suggests one way of decomposing the effects of *c*, *m*, *h*, and *x* on college graduation rates:

1. Conditional on initial major, high school quality, and academic background, how much does the different ways that African Americans and whites choose colleges account for differences in graduation rates?
2. Conditional on high school quality and academic background, how do different choices between African Americans and whites over initial majors affect graduation rates, both directly and through the choice of college?
3. Conditional on academic background, how do differences in high school quality between African Americans and whites affect college graduation rates, both directly and through choice of college and major?
4. And finally, how does equalizing academic backgrounds across African Americans and whites affect college graduation rates both directly and through the choice of college and major?

While we believe this is the most natural way to perform the decomposition, there are alternatives. As an example, we could reverse the ordering so that high school quality and pre-entry skills change before college and major re-sorting (we explore this alternative in Section 6.2). The rest of this section outlines how we estimate each of the four conditional probabilities and describes the corresponding decomposition calculations.

3.1 Reducing the State Space

One way of proceeding with the decomposition exercise is to be completely flexible in our specifications of each of the four probabilities. However, given small sample sizes – particularly at the high school cross college level – as well as the fact that we do not observe all of the relevant academic background characteristics, we place some structure on these relationships. In particular, we assume there is a function that maps observed academic background characteristics, *x*, high school quality, *h*, and race, *r*, into an academic index, *AI*.

We make two assumptions about how the academic index interacts with the choice of university and major. First, we assume that the probability of graduating is independent of *x*, *h*, and *r* once we condition on *c*, *m*, and *AI*:

$$\Pr(y=1 \mid c, m, AI, h, x, r) = \Pr(y=1 \mid c, m, AI) \quad \forall\{h, x, r\} \tag{2}$$

This assumption treats differences in graduation rates between African Americans and whites conditional on the same college and major as operating through the academic index. Alternatively, we could separate out race from the academic index, in which case we would be assuming that our measures of preparation are such that any additional differences by race are due to black-white differences in response to the college environment.

Second, we assume that the effects of *x* and *h* on the choice of college and major operate through the academic index:

$$\Pr(c \mid m, AI, h, x, r) = \Pr(c \mid m, AI, r) \quad \forall\{h, x\} \tag{3}$$

$$\Pr(m \mid AI, h, x, r) = \Pr(m \mid AI, r) \qquad \forall\{h, x\} \tag{4}$$

We make this assumption primarily because our measures of high school quality will come from high school fixed effects and sample sizes are problematic at the high school cross college level in some cases. The reduction of the state space afforded by the academic index also makes our analysis more tractable. Note that we still allow for individuals of different races to react differently to their academic index in how they choose their colleges and initial majors.

3.2 Graduation Probabilities

We now turn to specifying the first of the four conditional probabilities in equation (1): the conditional probability of graduation. This conditional probability depends on the university attended, *c*, the initial major, *m*, and the academic index, *AI*. We further allow it to depend on the students cohort, *t* (where *t* = 1996, 1997,…, 2001). The latent utility of graduating (within 8 years) for student *i* who enters the system with year-cohort *t* (where *t* = 1996, 1997,…, 2001) is given by:

$$y_i^* = \sum_c \sum_m \sum_t I(c,m,t \mid i)\delta_{0cmt} + \sum_c \sum_m I(c,m \mid i) AI_i \delta_{1cm} + \varepsilon_i \qquad (5)$$

where $I(c,m,t \mid i)$ and $I(c,m \mid i)$ are indicator variables for whether *i* attended university *c* with initial major *m* and, in the former case, whether *i* was part of cohort *t*. ε_i is an unobserved (to the econometrician) preference shock. Individuals who graduate, $y_i = 1$, have latent indexes greater than zero with $y_i = 0$ otherwise.

We form the academic index as a combination of student characteristics and high school indicator variables. The academic index for student *i* includes the student's gender, $f_i = 1$ if female and zero otherwise, the student's race, $b_i = 1$ if African American and zero otherwise, the student's ACT math and reading scores, $actm_i$ and $actr_i$, the student's high school class rank, g_i, and the student's high school, h_i.[13] h_i is a vector of length equal to the number of high schools in the data.

[13] The class-rank variable is normalized to align it with our procedure for adjusting high school quality, which we discuss below and in Appendix A.

The element of h_i that corresponds to the high school i attended is one and the other elements are zero.[14] The academic index is then given by:

$$AI_i = \gamma_0 + \gamma_1 g_i + \gamma_2 actm_i + \gamma_3 actr_i + \gamma_4 f_i + \gamma_5 b_i + \gamma_6 f_i b_i + h_i \theta \qquad (6)$$

The specification in equation (5) allows for school-major combinations to differ along three dimensions. First, colleges may make graduation attractive (be it through high monetary returns or lower effort costs) to all individuals. In this case δ_{0cmt} will be high regardless of initial major. Second, colleges may make graduation attractive for particular majors, for example due to a high return to one major but not another. Third, colleges may differ in their rewards for higher academic preparation and these rewards may also differ across initial majors. This would be the case, for example, if some schools and majors were better fits for the most prepared students (low δ_{0cmt}, high δ_{1cm}) but others were better fits for the least prepared students (high δ_{0cmt}, low δ_{1cm}).[15] Note that we allow the college-major intercept terms to vary over time to take into account changes in college and major quality. We specify ε_i as having a Type I extreme value distribution, implying that the probability of graduation from the perspective of the econometrician follows a logit.

3.3 University sorting

Next we consider how African American and white students sort into universities. The latent utility of initially enrolling in university *c* depends on the student's race, *r*, initial major, *m*, year-cohort, *t*, academic index, *AI*, and an unobserved preference η.

$$\begin{aligned} U_{ic} = \sum_t I(m,t\,|\,i)[&\phi_{0cmt} + \phi_{1cmt} f_i + \phi_{2cmt} b_i + \phi_{3cmt} f_i b_i \\ &+ AI_i(\phi_{4cmt} + \phi_{5cmt} f_i + \phi_{6cmt} b_i + \phi_{7cmt} f_i b_i)] + \eta_{ic} \end{aligned} \qquad (7)$$

[14] Students from high schools from which fewer than five students are observed over the course of the data panel as full-time, non-transfer college entrants are omitted. Only a small number of observations are dropped from the analytic sample for this reason. See Appendix Table B.1 for more information

[15] See Arcidiacono, Aucejo, Coate and Hotz (2012) and Arcidiacono, Aucejo and Hotz (2013) for similar specifications.

We assume that η follows a Type I extreme value distribution implying multinomial logit probabilities. Note that our sample includes only students who enrolled in a university in the Missouri system, implying the coefficients for one university must be normalized to zero.

This specification implies that we can estimate separate multinomial logits for each cohort of a particular race and gender who enter the system as STEM or non-STEM majors. In this way we allow student sorting to change over time as well as allowing the preferences over majors to influence the attractiveness of attending particular universities. Clearly some individuals will not be able to obtain admission to all schools. Hence, this specification can be thought of as an approximation to the combination of the school and student decision.

3.4 Major Sorting

Sorting into majors is handled in a similar manner to sorting into colleges. The latent utility of initially choosing major *m* is given by:

$$\begin{aligned} V_{im} = \sum_t I(t\,|\,i)[\tau_{0mt} + \tau_{1mt} f_i + \tau_{2mt} b_i + \tau_{3mt} f_i b_i \\ + AI_i(\tau_{4mt} + \tau_{5mt} f_i + \tau_{6mt} b_i + \tau_{7mt} f_i b_i)] + \varsigma_{im} \end{aligned} \tag{8}$$

where ς is distributed Type I extreme value. The probability of student *i* choosing major *m* then takes a logit form. We normalize the coefficients for non-STEM majors to zero and estimate separate models for the choice of initial major for each race-gender cohort.

3.5 High School Quality

We align high school quality between African Americans and whites by constructing a counterfactual distribution of high school fixed effects for African Americans (estimated as part of the academic index) that matches the distribution of the high school fixed effects for whites. The counterfactual distribution preserves the relative ordering of African Americans in terms high school quality, but shifts the entire distribution to align it with the distribution for whites. The distributional shift is a two-step process following Bound, Lovenheim and Turner (2010). In the first step we assign each African American and white student a race-specific percentile ranking in the distribution of high school quality. In the second step we map African Americans to counterfactual high school fixed effects by matching them to fixed effects that are consistent with their distributional rankings, but in the white rather than African American distribution.

The procedure for aligning high school quality requires an additional adjustment. One of our key academic background variables – high school class rank – is measured relative to the average academic background of the student body. Hence, if African Americans come from more disadvantaged family backgrounds than whites, moving a student from a predominantly African American high school to a predominantly white high school will result in that student having a lower class rank. We describe our method for adjusting class rank to account for this effect in Appendix A.

The net high school quality adjustment comes from the combination of (1) shifting up the distribution of high school fixed effects for African Americans, and (2) correspondingly decreasing their class ranks. We also allow for complementarities between high school quality and college/major sorting by re-running African Americans through the college and major sorting steps after the adjustment for high school quality.

There are two reasons why our approach is likely to yield an upper bound estimate of the contribution of differences in high school quality to racial graduation gaps. First, high school quality may be correlated with things such as family income that are not measured in our data, which in turn may be associated with higher completion rates. Second, the procedure we use to adjust African Americans' class ranks assumes that cross-school heterogeneity in academic backgrounds can be captured entirely by differences across schools in racial composition (see Appendix A). To the extent that average academic backgrounds across high schools vary because of other factors our class-rank adjustment will be too small. For these reasons, our estimates of the influence of high school quality on the graduation gaps are best viewed as upper bounds.

3.6 Academic Background

To see how the graduation gaps additionally close when we close the remaining gaps in academic preparation, we align the distributions of what remains of the academic index between African Americans and whites after adjusting for high school quality, again following the approach of Bound, Lovenheim and Turner (2010). We then re-run African Americans through the college and major sorting steps to capture the indirect effects of improving pre-entry skills associated with postsecondary sorting.

For our primary decompositions, which are based on the graduation model where we include the race-gender indicators as a part of the index (as in equation 5), this final adjustment closes the graduation gap by construction. However, our ability to close the graduation gap does not depend on this feature of our approach. Consistent with decompositions of black-white success gaps in other settings (e.g., Cameron and Heckman, 2001) we show that the racial graduation gap can be explained entirely by observable measures of pre-entry skills given sufficiently rich data. In our application, the key observable measures that fully explain the graduation gap (when combined with university and major sorting and high school quality) are the high school class rank and the high school fixed effects.[16]

3.7 Summary

Given the above specifications for graduation, university choice and major choice; along with the procedures to align high school quality and pre-entry skills between races, we now have the necessary components to answer the questions at the beginning of this section.[17] In particular, the

[16] Cameron and Heckman (2001) show that differences in pre-entry family income explain much of the racial gap in college enrollment, and further, that the mechanism is through income effects on the development of the abilities required to benefit from college. Belley and Lochner (2007) use more recent data to document an increasing role for income conditional on ability. Our decompositions allow for income to influence degree attainment gaps through both avenues. The college re-sorting procedure equalizes differences in college opportunities between whites and African Americans that may stem from income gaps and credit constraints (among other things). The high school quality and pre-entry skills alignment removes the effect of racial income differences on skill formation.

[17] See Appendix A for the equations corresponding to the decomposition.

decomposition begins by examining how racial differences in sorting into universities can account for differences in graduation, taking as given choice of major, high school assignment, and academic background characteristics. Then we evaluate how racial differences in the initial choice of major affect graduation rates, both directly and through university choice. Next we examine how racial differences in high school quality affect graduation rates, both through the academic index as well as through university and major sorting. Finally, the remaining gaps are explained by differences in academic background characteristics, directly and indirectly through sorting.

4. Estimates of the Determinants of College Graduation

We now turn to the empirical results, focusing in this section on the estimates of the graduation model. We first consider the importance of our various measures of academic preparation. Next, we examine graduation probabilities at each of the universities at different values of the academic index. Doing so allows us to see whether some universities are better at graduating students than other universities regardless of the academic background of the student, or whether some universities are best for students with relatively poor academic backgrounds and other universities are best for students with relatively strong academic backgrounds.

4.1 Graduation and the Academic Index

Table 2 shows the raw logit coefficients for the index variables from our preferred specification, where the academic index is specified as in equation (6), as well as from several sparser variants that include subsets of the information in the full index. Full results from the estimation of the parameters in equation (5) are available in Appendix D. As indicated above, the two most important student-level predictors of graduation success are high school of attendance and class rank. To see this, note that the table reports the effect of a one-standard-deviation move in the distribution of high school fixed effects (where the distribution is adjusted for estimation error in the fixed-effect estimates following Koedel, 2009). A one-standard deviation move in the normalized class-rank variable is 0.07, so moving one standard deviation in the distribution of the high school fixed effects is equivalent to moving approximately 0.82 standard deviations in the class-rank distribution. None of the other variables in the index are nearly as important as high school attended or class rank.[18]

The coefficients on the race-gender indicators from the index are also of interest and show that our measures of academic preparation are quite good. Conditional on the other measures of pre-entry preparation, African American men are no less likely than white men to obtain a degree. African American women are conditionally more likely to graduate than their white counterparts.[19] The fact that we can fully explain racial differences in gradation probabilities conditional on our covariates is a reflection of the power of our controls. For example, the alternative index formulations show that African American men are less likely to obtain a degree unconditionally (alternative 1) and conditional on entrance exam scores alone (alternative 3). Unconditionally, African American women perform worse than white men and white women, but outperform white men and perform similarly to white women conditional on ACT scores alone.

Table 3 shows descriptive statistics by race-gender group for the key components of the index. The table shows that white women have class ranks that are substantially higher than the other race-gender groups, which helps to explain their outperforming all other groups unconditionally (Table 2, column 2). Conditional on their superior preparation, however, white women are less likely to graduate than the other race-gender groups (Table 2, column 1). Table 3 also shows large differences in the other index components across races. African Americans attend

[18] The race-gender indicators are binary, and the standard deviations of ACT math and reading scores are 4.7 and 5.5, respectively, for the analytic sample. Combining the variability of the index components with the coefficients shown in Table 2 reveals the relative importance of the high school class rank and high school of attendance. Our finding of the weak predictive power of ACT math and reading scores over attainment, conditional on measures of high school attendance and performance, is consistent with Bowen, Chingos and McPherson (2009) and Fletcher and Tienda (2010).
[19] Our finding that African American men and women are more likely to graduate than whites (or, in the case of men, no less likely to graduate) conditional on pre-entry skills is consistent with previous studies that examine racial differences in college matriculation (Cameron and Heckman, 2001; Rivkin, 1995).

significantly worse high schools than whites and this is especially apparent for African American women. That the average high school effects are different for African American women and men may have to do with the large gender disparity in college enrollment rates for African Americans (Aucejo 2011); African American men are substantially less at risk of going to college.

The academic index combines the information from Tables 2 and 3 into a scalar measure of pre-entry preparation. Figure 2 summarizes the race-gender differences in preparation by plotting the distribution of index values for each race-gender group. Here we can see that African American men lag significantly behind their female counterparts. While white men also have lower index values than white women, the gender gap is much smaller for whites than for African Americans.

4.2 Graduation and the College-Major Match

We next examine how colleges differ in their probabilities of graduating different types of students. Table 4 shows predicted graduation rates for students entering into each university-by-major cell, holding the index value fixed at different points in the index distribution for African Americans (genders combined). The table is divided by entering major type; the first panel is for STEM entrants and the second panel is for non-STEM entrants. Graduation probabilities are reported at the 5th, 10th, 25th, 50th, 75th and 90th percentiles of the index distribution.[20] Reading down the rows in any given column compares observationally equivalent students in different university-by-major cells.[21]

Consistent with Table 1, Table 4 shows that the two urban schools (Kansas City and St. Louis) as well as the four least-selective schools (Missouri Southern, Western Missouri, Lincoln and Harris Stowe) lag significantly behind the others in terms of predicted graduation rates. These

[20] The 10th, 25th, 50th and 75th percentiles of the index distribution for whites (8.28, 9.04, 9.72, 10.26) are at approximately the 25th, 50th, 75th and 90th percentiles of the index distributions for African Americans.

[21] The parameters that underlie the predictions in Table 4 are provided in Appendix D. The table shows predicted graduation rates in each cell holding non-index values fixed at their sample averages (e.g., year cohorts). Note that although the graduation model allows for differential graduation rates by entering major, the outcome is not degree-specific. Some of the predictions reported in the table are extrapolated out of sample (in particular, graduation rates at top schools for individuals with extremely low index values).

schools are behind regardless of initial major or where in the academic index distribution we look. Consider the median African American student. The bottom panel of Table 4 shows this student's predicted graduation probability as a non-STEM major at Central Missouri would be 56 percent. This is 12 percentage points higher than at Lincoln and 15 percentage points higher than at Kanas City.[22] In fact, moving from any of the least-selective or urban schools to any of the moderately-selective schools corresponds to a large increase in her likelihood of degree attainment. More generally, a takeaway from Table 4 is that the universities in which African American students are most overrepresented in the system – the least-selective and urban campuses – are also the ones with the lowest graduation rates conditional on students' pre-entry preparation.[23]

There is also some evidence that the match between the student's academic preparation and the quality of the school is important, and more so for those who begin in the sciences. At the 90th percentile of the African American distribution the three most-selective colleges have higher graduation rates than the moderately-selective colleges, which in turn have higher graduation rates than the bottom four schools, regardless of initial major. At lower percentiles, however, this is not the case. At the 25th percentile, three of the four moderately-selective colleges have higher graduation rates for initial STEM majors than Columbia, ranging from four to eleven percentage points higher. For non-STEM majors, matching is less important with only one of the four moderately-selective schools, Northwest Missouri, having a higher predicted graduation rate, two percentage points higher, than Columbia at the 25th percentile of the index distribution.

[22] An issue that may be of interest to some readers is whether African American enrollment shares across universities are associated with differential African American success rates. We examine this issue in Appendix E and find no evidence of an association.

[23] The estimates in Table 4 are broadly consistent with a number of studies showing the importance of college quality as a determinant of completion (see, for example, Cohodes and Goodman, 2013; Black and Smith, 2006; Loury and Garman, 1995).

5. Breaking out the Determinants of Racial Differences in Graduation Rates

With the estimates from the graduation model in hand, we now turn to the decomposition. The decomposition entails estimating the assignments of students to colleges and majors as well as aligning the distributions of high school quality and academic background. Table 5 shows our primary decomposition results. The top panel of the table reports the graduation gaps that we aim to explain with our models, and the bottom panel shows the share of the gaps can be accounted for by each aspect of our decomposition procedure. In the rest of this section we provide some context for each component of the decomposition and compare their relative importance in explaining the graduation gap.

5.1 College and Major Sorting

The college re-sorting procedure uses the parameter estimates from the sorting model (equation 7), estimated for white students only, to produce counterfactual university assignments for African Americans. The first row in Table 5 shows that differences in college sorting between African Americans and whites explain 14.7 and 5.6 percent of the graduation gap for women and men, respectively. These gains come with no changes to pre-entry skills for African Americans.

The second row of Table 5 shows that additionally re-sorting African Americans to majors leads to a slight increase in graduation rates – major sorting explains 1.5 and 0.3 percent of the graduation gap for women and men, respectively. The small increases occur because African Americans are conditionally over-represented as STEM entrants and the major sorting procedure shifts African American enrollment from STEM to non-STEM fields. Based on actual sorting patterns, 21 percent of African American students enter as STEM majors. Under the counterfactual, 16 percent of African Americans enter as STEM majors. The shifts in initial major suggested by our sorting models are predicted to modestly increase graduation rates for African Americans because conditional on pre-entry preparation, non-STEM majors are more likely to graduate (see Table 4 and Appendix D).[24]

Table 6 shows how African American enrollment patterns across universities change in moving from actual to counterfactual sorting (the sorting model mechanically fits the data for whites nearly perfectly; whites' actual and predicted assignments are the same). For ease of presentation we collapse majors within universities in the table (Appendix C shows details broken down to the university-by-major level). For African American women and men separately, the table first shows the average index value and enrollment share at each university based on where students actually entered the system. Then it shows these same calculations where we replace each student's actual assignment with the vector of predicted assignment probabilities generated by the college and major sorting models.[25] The predicted student shares for each university are the summations of the predicted probabilities across all students, and the index values are weighted averages where the predicted probabilities serve as the weights.

Table 6 shows that there is movement of African Americans across most schools in the system, reflecting a combination of the intercept and slope parameters from the sorting model. Consistent with the descriptive statistics provided thus far, counterfactual sorting shifts a large fraction of African American enrollment out of the urban and least-selective schools and into the moderately-selective schools. Unsurprisingly, the differences between actual and counterfactual sorting are particularly stark at the two historically black universities (Lincoln University and Harris Stowe State University). Note that the gains in predicted graduation rates in Table 5 are not occurring because of increased representation at the top schools. While Truman State does see an

[24] If STEM majors are more difficult (Arcidiacono, Aucejo, and Spenner 2012; Koedel 2011) and major switching is costly it would explain this result. However, the shift in African American enrollment from STEM to non-STEM fields explains very little of the graduation gap because differences in how African Americans and whites choose their initial fields of study are small.
[25] The enrollment shifts shown in Table 6 are substantively similar in analogous calculations that do not account for major sorting.

increase in the share of African Americans, the enrollment gain there is more than offset by losses at UM-Columbia.

The larger gains for African American women relative to African American men are the result of three factors. First, relative to their male counterparts, African American women are more overrepresented at urban schools. Second, African American women have higher index values and hence can benefit more from the increases in college quality. Finally, African American men are more likely to be STEM majors where the gains in graduation probabilities from improving college quality are lower.

5.2 High School Quality

We next consider improving African American high schools so that they attend schools of equal quality to whites (in terms of empirically predicting success), holding the pool of African American entrants fixed. To align high school quality between African Americans and whites we adjust African Americans' high school fixed effects, and correspondingly, their class ranks, as discussed in Section 3. Adjusting high school quality in turn affects assignment to colleges and majors.

The third row of Table 5 shows that the net high school quality adjustment reduces the graduation gap by 2.8 and 1.5 percentage points, or 18.5 percent and 8.4 percent of the gap, for women and men respectively. As discussed earlier, our method for obtaining these effects is likely an upper bound, effectively assuming away selection into higher quality high schools. The effect for women is higher than the effect for men, consistent with the descriptive statistics provided in Table 3. While both male and female African Americans attend lower quality high schools than their white counterparts conditional on enrolling in one of the four-year colleges in the Missouri system, this is particularly true for women and reflects the large gender gap in college enrollment among African Americans.

To get a sense of what our high school quality measures are picking up we incorporate additional data from the 2000 census. In particular, we merge in data on the median household income, share with a high school degree, share with a college degree, racial composition, and population density for the zip code in which the high school is located. We then regress our estimates of high school quality on these variables. After making an adjustment for estimation error in the high school fixed effects following Aaronson, Barrow and Sander (2007), we estimate that the census information explains just over 25 percent of the true variance in the high school fixed effects.[26] The most important predictors by far are the two controls for local-area education levels, which combine to explain 22 percent of the variance by themselves.

5.3 Academic Background

Finally, the last row of Table 5 shows how the graduation gap is affected by aligning pre-entry skills. Note that aligning pre-entry skills affects graduation rates directly as well as through the assignment to colleges and majors. The alignment of pre-entry skills, which as a practical matter is driven almost entirely by changing African American class ranks, explains most of the graduation gap. Specifically, it reduces the graduation gap by 9.8 and 15.4 percentage points, or 65.3 percent and 85.7 percent of the total gap, for women and men respectively.

Our findings on the importance of pre-entry skills complement prior studies looking at the determinants of black-white success gaps for other outcomes. Notable examples include Cameron and Heckman's (2001) examination of the determinants of gaps in college attendance and Neal and Johnson's (1996) examination of gaps in labor market outcomes. Until significant changes occur in the pre-entry skills gap, large differences in college graduation rates will remain.

---

[26] The adjustment is required because the high school fixed effects are estimated with error. Even if the Census variables explain 100 percent of the true variance in high school quality they would not explain 100 percent of the variance in the estimates of high school quality. See Aaronson, Barrow and Sander (2007) for details.

6. Extensions and Sensitivity Analysis

We now extend our analysis in three ways. First, we examine the sources of the gains from reallocating African Americans to different universities. Second, we look at the sensitivity of our results to changing the order of the decomposition and to removing information about initial majors from our models. Finally, we see how our decomposition results change when we consider alternative graduation outcomes.

6.1 The Gains from College Sorting

In this section we briefly extend the college-sorting exercise to examine which aspects of the counterfactual re-sorting scenario are driving the predicted improvements in graduation rates for African Americans. African Americans are overrepresented among the most-selective schools (primarily due to UM-Columbia), the urban schools, and at three of the four least selective schools. We examine how shifting African American enrollment from each of these three sets of schools affects graduation rates, considering three specific scenarios:

1. Holding African American enrollment fixed at the urban and least selective schools, how does reallocating the remaining African American students according to the white assignment rules affect graduation rates?
2. Holding African American enrollment fixed at the most selective and least selective schools, how does reallocating the remaining African American students according to the white assignment rules affect graduation rates?
3. Holding African American enrollment fixed at the most selective schools and the urban schools, how does reallocating the remaining African American students according to the white assignment rules affect graduation rates?

The first scenario allows us to examine how college quality at the top end of the distribution affects graduation rates. This would be somewhat similar to removing affirmative action at top schools. The second and third scenarios focus on removing African American students from the groups of colleges with the lowest graduation rates.

Given the limited role that major sorting plays in our decompositions, we hold initial major fixed and focus on the gains from re-sorting via equation (7). Let C_n indicate the set of colleges being considered under scenario *n*. The conditional probability of an African American student being assigned to school *c* in the set C_n under the white assignment rules is:

$$\Pr_i(c) = \frac{\exp(U_{ic}^*)}{\sum_{c \in C_n} \exp(U_{ic}^*)}$$

where U_{ic}^* is taken from (7) and African American interactions and the error term are removed:

$$U_{ic}^* = \sum_t I(m,t \mid i)[\phi_{0cmt} + \phi_{1cmt} f_i + AI_i(+\phi_{4cmt} + \phi_{5cmt} f_i)]$$

We calculate these probabilities for all African Americans who attended schools in C_n. The effect of the alternative "partial sorting" assignment policies on graduation rates is then given by:

$$\sum_i{}^{'} \sum_{c \in C_n}{}^{'} \Pr_i(y \mid c) \Pr_i(c) - \sum_i{}^{'} \sum_{c \in C_n}{}^{'} I(c \mid i) \Pr_i(y \mid c) \tag{9}$$

where $\Pr_i(y \mid c)$ is calculated using the parameter estimates in (5), $I(c \mid i)$ is an indicator variable for whether *i* chose *c*, and where the first sum is taken only for African Americans who attended one of the colleges in C_n.

Results for each of the scenarios are presented in Table 7. In the first scenario, moving African American students out of the top colleges and into the moderately selective colleges has a small, negative effect on graduation rates. The moderately selective schools – particularly given the academic backgrounds of African American students – produce graduates at a rate that is not all that different from the very best schools in the state. The next two partial-sorting scenarios increase predicted African American graduation rates. Reallocating African American women away from the urban and least selective schools increases graduation rates by over one percentage point in each scenario. The graduation-rate changes for men in these scenarios are also positive, but smaller given the smaller role that sorting plays for men in general.

6.2 Alternative Decompositions

Next we consider the sensitivity of our findings to adjustments to the decomposition procedure. We show results from two alternatives to our preferred approach (in Table 5). First, we eliminate majors from the exercise entirely. This requires re-estimating the graduation model without initial major interactions. Results are presented in Table 8. The previous results suggest that initial majors affect graduation rates but that differences in initial majors across races are not an important driver of racial graduation gaps. Table 8 shows that removing information about majors from the analysis has little effect on the relative importance of the other factors in explaining racial differences in graduation rates.

We next consider an alternative ordering of the decomposition. Prior research in other contexts indicates that ordering can be important (Dinardo, Fortin and Lemieux, 1996). Our preferred ordering is driven by the increased interest from economists and others in improving college-choice outcomes for disadvantaged populations (Bowen, Chingos and McPherson, 2009; Roderick et al., 2008; Hoxby and Avery, 2012; Hoxby and Turner, 2013). The first step in the decompositions, as presented thus far, gives an indication of how much interventions at this level can be expected to narrow the black-white graduation gap in the absence of earlier interventions to rectify skill gaps. Still, it is of interest to consider how the decomposition results change in response to reordering the steps.

Table 9 presents an alternative decomposition that first shifts high school quality, then shifts other pre-entry skills, and concludes by re-sorting African Americans across universities and majors. Our findings are similar qualitatively to what we show in Table 5. That said, it is of some interest that the weights on the various aspects of the decomposition shift slightly away from the high school quality and pre-entry skill components and toward college re-sorting. The reason is that the gains from improving high school quality and other pre-entry skills for African Americans are muted if we hold their university assignments fixed, again owing to the complementarity between pre-entry skills and college quality. The gains from college re-sorting are larger if the re-sorting occurs after pre-entry skills are aligned.

6.3 Alternative Graduation Outcomes

So far we have focused on eight-year graduation rates from any university in the system as our outcome of interest. In this section we examine graduation outcomes over a shorter time horizon – within five and six years – as well as eight-year graduation rates from the initial school to see how our findings are affected by transfers.

Table 10 reports results from decompositions for the new outcome measures following the same procedure as outlined in Section 3. A key takeaway from Table 10 is that regardless of which outcome we consider, racial differences in pre-entry skills are the primary driver of the graduation gaps. In fact, pre-entry skills become more important as we shorten the graduation window. Table 10 also shows that high school quality and college sorting explain smaller shares of the graduation gaps when we shorten the time horizon, and major sorting explains a larger share. Switching majors may lead to delays in graduation and individuals are more likely to switch out of STEM majors than into STEM majors. Because African Americans are conditionally more likely to choose STEM majors upon entry, major sorting takes on an increased (albeit still small) role in the models where we shorten the graduation window.

7. Conclusion

Differences in college graduation rates between African Americans and whites are stark. In the Missouri system, conditional on 4-year college enrollment the gap for women is 15 percentage points and for men it is 18 percentage points. These gaps are in line with nationwide gaps in BA completion at four-year public universities as reported by Lynch and Engle (2010).

We show that racial differences in graduation rates can be partially diminished by re-sorting African Americans across universities so that their enrollment decisions are similar to comparable white students. The gains result from shifting African Americans away from urban campuses and the least selective schools, both of which have relatively low graduation rates at all skill levels. The graduation gaps can also be partly explained by differences in high school quality, to which some differences in pre-entry skills are directly attributable. However, racial differences in other pre-entry academic preparation are the primary driver for graduation gaps for both women and men.

Although our analytic sample is restricted to students who initially enroll in a four-year public university in Missouri, there are reasons to believe that our findings provide insights about black-white graduation gaps that will generalize more broadly. For example, the Missouri system includes universities that vary considerably in quality (measured by selectivity of admissions and graduation rates), and the U-shaped pattern of high African American enrollment at the most and least selective universities, and low enrollment at moderately selective schools, has been found elsewhere (Arcidiacono, Vigdor and Khan, 2011). Still, a limitation of our study is that we do not analyze the full universe of students who interact with the higher-education sector, even in Missouri. Our study does not speak directly to postsecondary graduation gaps as they pertain to students who enter the system through community colleges, or to students who attend private or out-of-state colleges.[27]

We also note that our college re-sorting exercise ignores potential general-equilibrium effects of the student reallocation. Given that 94 percent of the students in the analytic sample are white, so that only 6 percent of the student population is being moved, the general equilibrium consequences

[27] It is possible to analyze graduation pathways through community colleges with our data. We leave this for future research.

in the present application are probably small. In an omitted analysis we verify that with a handful of exceptions at the smallest universities in the system (and in particular the two historically black universities), the hypothetical re-sorting of African American students that we consider has essentially no effect on the overall selectivity of the universities in Missouri. The effects on total enrollment at most of the schools would also be small.

Finally, the validity of our counterfactual calculations depends on our ability to draw causal inference from the graduation model. Analogously to Bound, Lovenheim and Turner (2010), it must be the case that the relationship between pre-entry skills and college completion indicated by our model reflects the effect of these skills on this outcome. Our reliance on variation within university-by-major cells to identify the pre-entry skill parameters in equation (5) lends some credence to a causal interpretation. Perhaps less obvious is whether the counterfactual college and major assignments for African Americans would affect them in the same way as whites. In Appendix E we show that our model does not make systematic errors in predicting African American graduation rates across universities based on the sorting we observe, which suggests a limited role for race-specific heterogeneity in university effectiveness and generally supports a causal interpretation. But, of course, our data are not sufficient to provide conclusive evidence on this point. At the least, our study provides a clear framework for describing the factors that align with observed racial gaps in college completion.

References

Aaronson, Daniel, Lisa Barrow and William Sander. 2007. Teachers and Student Achievement in the Chicago Public High Schools. *Journal of Labor Economics* 25(1), 95-135.

Arcidiacono, Peter. 2005. Affirmative Action in Higher Education: How do Admission and Financial Aid Rules Affect Future Earnings? *Econometrica* 73(5), 1477-1524.

Arcidiacono, Peter, Esteban Aucejo, Patrick Coate, and V. Joseph Hotz. 2012. Affirmative Action and University Fit: Evidence from Proposition 209. working paper.

Arcidiacono, Peter, Esteban Aucejo, Hanming Fang, and Kenneth Spenner. 2011. What Happens After Enrollment? An Analysis of the Time Path of Racial Difference in GPA and Major Choice. *Quantitative Economics* 2(3), 303-333.

Arcidiacono, Peter, Esteban Aucejo, and V. Joseph Hotz. 2013. University Differences in the Graduation of Minorities in STEM Fields: Evidence from California. Working paper.

Arcidiacono, Peter, Esteban Aucejo, and Kenneth Spenner. 2012. What Happens After Enrollment? An Analysis of the Time Path of Racial Difference in GPA and Major Choice. *IZA: Journal of Labor Economics* Vol. 1, Article 5.

Arcidiacono, Peter, Pat Bayer, and Aurel Hizmo. 2010. Beyond Signaling and Human Capital: Education and the Revelation of Ability. *American Economic Journal: Applied Economics*, 2(4), 76-104.

Arcidiacono, Peter, Jacob Vigdor, and Shakeeb Khan. 2011. "Representation versus Assimilation: How do Preferences in College Admissions Affect Social Interactions? *Journal of Public Economics* 95(1-2): 1:15.

Aucejo, Esteban. 2012. Explaining Cross-Racial Differences in the Educational Gender Gap. Working paper.

Belley, P. and L. Lochner. 2007. The Changing Role of Family Income and Ability in Determining Educational Achievement. *Journal of Human Capital* 1(1): 37–89

Bound, John, Michael Lovenheim, and Sarah E. Turner. 2010. Why Have College Completion Rates Declined? An Analysis of Changing Student Preparation and Collegiate Resources. *American Economic Journal: Applied Economics*, 2(3): 1-31.

Black, Dan A. and Jeffrey A. Smith. 2006. Estimating the Returns to College Quality with Multiple Proxies for Quality. *Journal of Labor Economics* 24(3): 701–728.

Bound, John and Sarah Turner. 2011. "Dropouts and diplomas: The divergence in collegiate outcomes," in E. Hanushek, S. Machin & L.Woessmann, eds, *Handbook of the Economics of Education*, Vol. 4, Elsevier.

Bowen, William G., Mathew M. Chingos and Michael S. McPherson. 2009. *Crossing the Finish Line*. Princeton, New Jersey: Princeton University Press.

Cameron, Stephen V., and James J. Heckman. 2001. The Dynamics of Educational Attainment for Black, Hispanic, and White Males. *Journal of Political Economy* 109(3): 455-499.

Card, David and Alan B. Krueger. 2005. Would the Elimination of Affirmative Action Affect Highly Qualified Minority Applicants? Evidence from California and Texas. *Industrial and Labor Relations Review* 58(3): 416-434.

Cohodes, Sarah and Joshua Goodman. 2013. Merit Aid, College Quality and College Completion: Massachusetts' Adams Scholarship as an In-Kind Subsidy. Unpublished manuscript, Harvard University.

DeAngelo, Linda, Ray Franke, Sylvia Hurtado, John H. Pryor, and Serge Tran. 2011. Completing College: Assessing Graduation Rates at Four-Year Institutions. Higher Education Research Institute report.

Dillon, Eleanor and Jeffrey Smith. 2012. The Determinants of Mismatch Between Students and Colleges. Working paper.

Dinardo, John, Nicole M. Fortin and Thomas Lemieux. 1996. Labor Market Institutions and the Distribution of Wages, 1973-1992. A Semiparametric Approach. *Econometrica* 64(5): 1001-1044.

Fletcher, Jason and Marta Tienda. 2010. "Race and Ethnic Differences in College Achievement: Does High School Attended Matter?" *The ANNALS of the American Academy of Political and Social Science* 627: 144-166.

Heckman, James, Lance Lochner, and Petra Todd. 2006. "Earnings functions, rates of return and treatment effects: the Mincer equation and beyond", in E. Hanushek & F. Welch, eds, *Handbook of the Economics of Education*, Vol. 1, Elsevier.

Howell, Jessica S. 2010. Assessing the Impact of Eliminating Affirmative Action in Higher Education. *Journal of Labor Economics* 28(1): 113-166.

Hoxby, Caroline and Christopher Avery. 2012. The Missing "One-Offs": The Hidden Supply of High-Achieving, Low Income Students. National Bureau of Economic Research Working Paper No. 18586.

Hoxby, Caroline and Sarah Turner. 2013. Expanding College Opportunities for High-Achieving, Low Income Students. Stanford Institute for Economic Policy Research Discussion Paper No. 12-014.

Koedel, Cory. 2011. Grading Standards in Education Departments at Universities. *Education Policy Analysis Archives* 23(1).

Koedel, Cory. 2009. An Empirical Analysis of Teacher Spillover Effects in Secondary School. *Economics of Education Review* 28(6): 682-692.

Light, Audrey and Wayne Strayer. 2000. Determinants of College Completion: School Quality or Student Ability. *The Journal of Human Resources* 35(2): 299-332.

Light, Audrey and Wayne Strayer. 2002. From Bakke to Hopwood: Does Race Affect College Attendance and Completion? *The Review of Economics and Statistics* 84(1): 34-44.

Long, Mark. 2004a. College Applications and the Effect of Affirmative Action. *Journal of Econometrics* 121(1-2): 319-342.

Long, Mark. 2004b. Race and College Admission: An Alternative to Affirmative Action? *Review of Economics and Statistics* 86(4): 1020-1033.

Loury, Linda Datcher and David Garman. 1995. College Selectivity and Earnings. *Journal of Labor Economics* 13(2): 289-308.

National Center for Education Statistics. 2012. *The Condition of Education*. Washington, DC: United States Department of Education.

Neal, Derek A. and William R. Johnson. 1996. The Role of Premarket Factors in Black-White Wage Differences. *Journal of Political Economy* 104(5): 869-895.

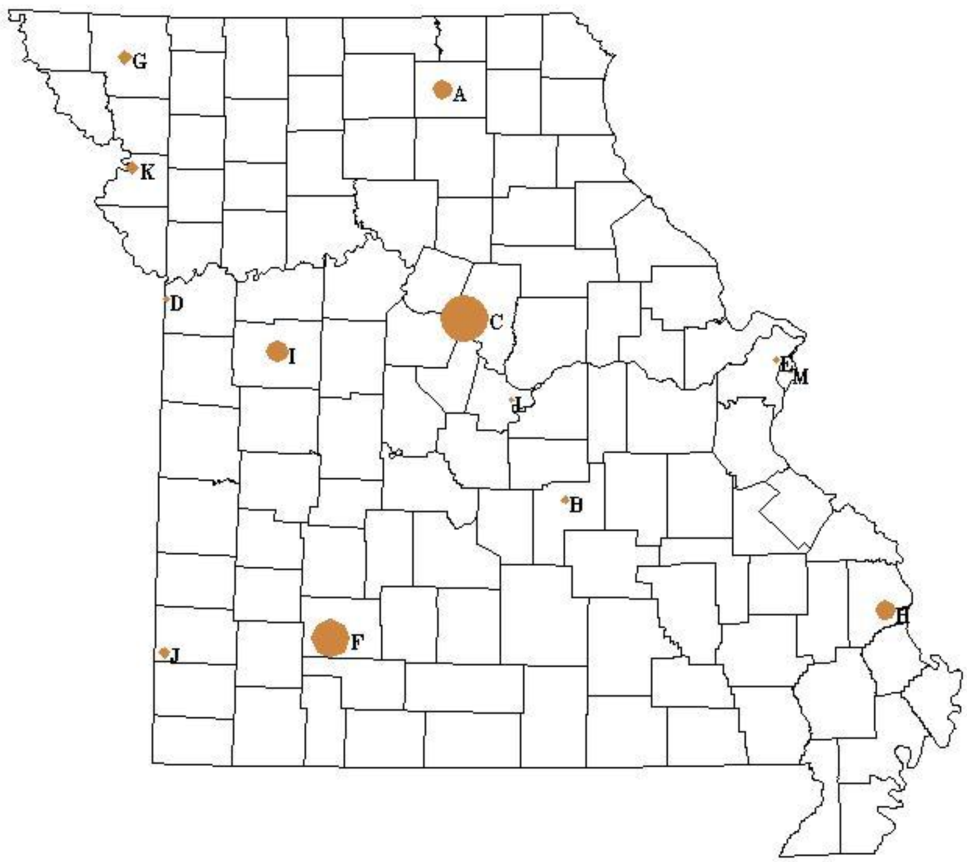
Rivkin, Steven G. 1995. Black/White Differences in Schooling and Employment. *The Journal of Human Resources* 30(4): 826-852.

Roderick, Melissa, Jenny Nagaoka, Vanessa Coca, Eliza Moeller, Karen Roddie, Jamiliyah Gilliam and Desmond Patten. 2008. From High School to the Future: Potholes on the Road to College. Policy Report, Consortium on Chicago School Research.

Rothstein, Jesse M. 2004. College performance predictions and the SAT. *Journal of Econometrics* 121(1-2): 297-317.

Turner, Sarah E. and William G. Bowen. 1999. The Changing (Unchanging) Gender Gap. *Industrial and Labor Relations Review* 52(2): 289-313.

Figure 1. Geographic Locations of Missouri Public Universities.



Legend

A: Truman State University
B: UM-Rolla
C: UM-Columbia
D: UM-Kansas City
E: UM-St. Louis
F: Missouri State University
G: Northwest Missouri State University
H: Southeast Missouri State University
I: University of Central Missouri
J: Missouri Southern State University
K: Western Missouri State University
L: Lincoln University
M: Harris Stowe State University

Note: Circle sizes correspond to enrollment shares from the analytic sample.

Figure 2. Index-Value Distributions by Race-Gender. Median index values for white women, African American women, white men and African American men are 9.8, 9.3, 9.6 and 8.7, respectively.

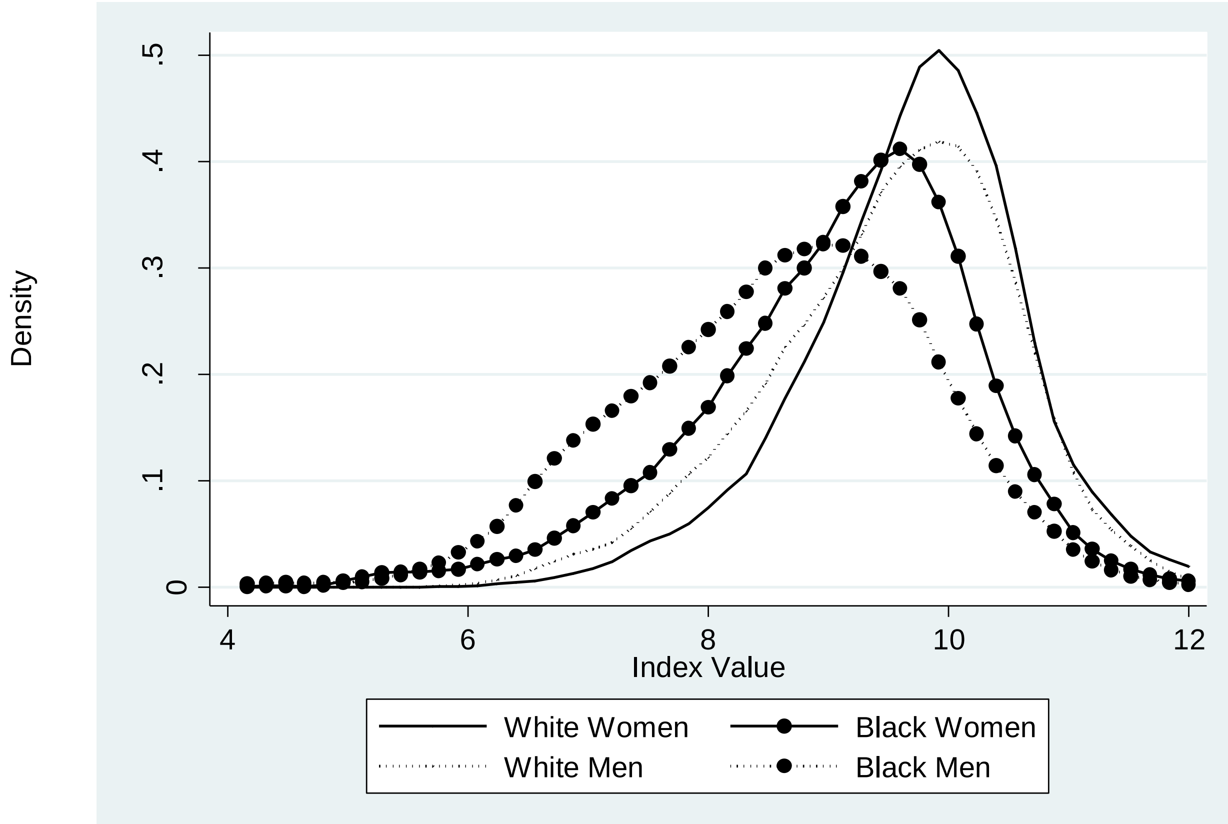


Table 1. University Descriptive Statistics for Analytic Sample.

	Avg. Preparation Index	Descriptive Statistics Relative to Full System: Enrollment Share	Initial STEM Share	Minority Share	Internal Descriptive Statistics: Initial STEM Share	Minority Share	Total Grad Rate
All	9.56	1.000	0.218	0.063			
Truman State Univ	10.21	0.09	0.10	0.04	0.23	0.03	0.80
Univ of Missouri -Rolla	10.09	0.04	0.17	0.02	1.00	0.03	0.74
Univ of Missouri-Columbia	10.03	0.22	0.31	0.23	0.30	0.06	0.76
Univ of Missouri -Kansas City	9.77	0.03	0.07	0.07	0.44	0.12	0.54
Univ of Missouri -St. Louis	9.69	0.03	0.02	0.06	0.17	0.12	0.52
Missouri State Univ	9.41	0.18	0.08	0.06	0.09	0.02	0.59
Northwest Missouri State Univ	9.38	0.07	0.05	0.03	0.15	0.03	0.64
Southeast Missouri State Univ	9.32	0.09	0.05	0.08	0.11	0.05	0.59
University of Central Missouri	9.29	0.10	0.08	0.07	0.17	0.04	0.60
Missouri Southern State Univ	8.99	0.05	0.03	0.01	0.13	0.02	0.44
Western Missouri State Univ	8.73	0.06	0.04	0.11	0.14	0.11	0.42
Lincoln Univ	8.69	0.02	0.01	0.16	0.08	0.40	0.40
Harris Stowe State Univ	8.25	0.01	0.00	0.05	0.00	0.63	0.30
N		63,135	13,740	3,952			

Notes: The analytic sample includes full-time, resident, non-transfer students who entered the system between 1996 and 2001 as college freshman (African American and white only). It omits students whose high school of attendance, class rank, and/or ACT scores are unavailable (combined data loss ≈ 6 percent). See Appendix B for more details about the construction of the analytic sample.

Table 2. Index Parameters from Primary and Alternative Specifications for the Index.

	Primary Specification	Alternative Specifications (1)	(2)	(3)	(4)	(5)
High School Class Rank	13.41				7.57	13.26
	(0.043)**				(0.300)**	(0.421)**
ACT Math Score	0.006		0.064	0.076	0.034	0.008
	(0.003)*		(0.005)**	(0.005)**	(0.003)**	(0.003)**
ACT Reading Score	-0.012		0.006	0.003	-0.009	-0.013
	(0.002)**		(0.002)**	(0.002)	(0.002)**	(0.002)**
White Female	-0.062	0.256		0.305	0.069	
	(0.022)**	(0.036)**		(0.024)**	(0.018)**	
African American Male	0.058	-0.461		-0.144	-0.180	
	(0.070)	(0.084)**		(0.050)**	(0.053)**	
African American Female	0.124	-0.089		0.257	-0.067	
	(0.060)*	(0.039)*		(0.040)**	(0.043)	
HS Fixed Effects[a] (Standard Deviation)	0.75					0.75

Note: Estimates are for the index parameters in the graduation model. The remaining parameters for our preferred specification are in Appendix D.

[a] The reported standard deviation for the high school fixed effects is unweighted and adjusted for estimation error in the fixed-effect estimates following Koedel (2009).

** indicates statistical significance at the 1 percent level; * indicates statistical significance at the 5 percent level

Table 3. Race-Gender Sample Averages for Index Components.

	White Women	White Men	African American Women	African American Men
Class Rank	74.59	67.55	64.00	53.15
HS Fixed Effect	0.031	0.028	-0.472	-0.390
ACT Math	22.14	23.83	18.34	19.13
ACT Reading	24.71	24.70	20.35	19.86
N	32680	26503	2486	1466

Notes: HS fixed effects are reported in standard deviation units and centered around the weighted sample average.

Table 4. Predicted Graduation Probabilities at Different Points in the African American Academic Index Distribution (Genders Combined), at each University and for Each Entering Major.

Panel A. STEM Entrants.

	Predicted Graduation Rates by Percentile of the Index Distribution for African Americans					
	5	10	25	50	75	90
	(6.75)	(7.24)	(8.17)	(9.03)	(9.74)	(10.26)
Truman State	0.35	0.42	0.55	0.67	0.75	0.80
UM-Rolla	0.08	0.13	0.29	0.50	0.68	0.79
UM-Columbia	0.06	0.10	0.24	0.47	0.67	0.79
UM-Kansas City	0.08	0.12	0.22	0.37	0.52	0.63
UM-St. Louis	0.14	0.17	0.27	0.38	0.49	0.57
Missouri State	0.05	0.08	0.21	0.41	0.61	0.74
Northwest Missouri State	0.11	0.17	0.32	0.51	0.67	0.76
Southeast Missouri State	0.09	0.14	0.28	0.46	0.63	0.73
Central Missouri	0.14	0.20	0.35	0.53	0.68	0.76
Missouri Southern State	0.03	0.06	0.15	0.32	0.51	0.65
Western Missouri State	0.06	0.09	0.21	0.40	0.59	0.71
Lincoln	0.09	0.13	0.26	0.42	0.57	0.68
Harris Stowe State						

Panel B. Non-STEM Entrants.

Truman State	0.11	0.17	0.36	0.58	0.75	0.84
UM-Rolla						
UM-Columbia	0.12	0.19	0.38	0.59	0.75	0.83
UM-Kansas City	0.11	0.15	0.27	0.41	0.54	0.64
UM-St. Louis	0.07	0.10	0.21	0.37	0.53	0.65
Missouri State	0.10	0.15	0.32	0.52	0.69	0.79
Northwest Missouri State	0.16	0.22	0.40	0.59	0.73	0.81
Southeast Missouri State	0.14	0.20	0.36	0.55	0.69	0.78
Central Missouri	0.15	0.21	0.38	0.56	0.71	0.79
Missouri Southern State	0.09	0.13	0.26	0.44	0.60	0.71
Western Missouri State	0.08	0.13	0.27	0.47	0.64	0.75
Lincoln	0.13	0.18	0.30	0.44	0.57	0.66
Harris Stowe State	0.12	0.16	0.26	0.39	0.50	0.59

Notes: Each column shows predicted graduation rates for index values fixed at different point in the distribution for African Americans, as reported in parentheses at the top of each column. Predictions are made by using sample average values across year cohorts.

Table 5. Decompositions of the Graduation Gaps.

	Women	Men
White predicted graduation rate	66.2	61.1
African American predicted graduation rate	51.2	43.1
Total gap (percentage points)	15.0	18.0
Portion of gap explained by:		
College Sorting	2.2 (14.7%)	1.0 (5.6%)
Major Sorting	0.2 (1.5%)	0.1 (0.3%)
High school quality adjustment	2.8 (18.5%)	1.5 (8.4%)
Pre-entry skills adjustment	9.8 (65.3%)	15.4 (85.7%)

Notes: The decomposition begins by resorting African American students to universities according to white assignment rules. Next, African Americans are resorted into different majors which in turn affects sorting into colleges. High school quality and pre-entry-skills adjustments also lead to re-sorting to universities and majors to allow for complementarities between pre-entry preparation and college and major selection.

Table 6. University-Average Index Values and Enrollment Shares by Race-Gender Group, for Actual and Counterfactual Sorting.

	African American Women				African American Men				Whites (Actual Sorting)			
	Actual		Counterfactual		Actual		Counterfactual		Women		Men	
	Avg Index	Share	Avg Index	Share	Avg Index	Share	Avg Index	Share	Avg Index	Share	Avg Index	Share
Truman State	9.77	0.037	9.94	0.072	9.97	0.036	9.68	0.046	10.26	0.106	10.15	0.084
UM-Rolla	10.04	0.007	9.95	0.009	9.59	0.042	9.59	0.041	10.27	0.014	10.07	0.070
UM-Columbia	9.77	0.243	9.79	0.153	9.34	0.207	9.43	0.151	10.15	0.214	9.95	0.235
UM-Kansas City	9.48	0.073	9.47	0.028	9.12	0.061	9.06	0.024	9.93	0.034	9.69	0.031
UM-St. Louis	9.33	0.064	9.43	0.023	8.99	0.046	9.02	0.024	9.86	0.027	9.64	0.030
Missouri State	9.33	0.063	9.04	0.209	8.89	0.064	8.57	0.194	9.54	0.196	9.24	0.177
Northwest Missouri State	9.11	0.027	9.04	0.077	8.67	0.044	8.50	0.074	9.54	0.073	9.19	0.065
Southeast Missouri State	8.99	0.084	8.94	0.115	8.45	0.081	8.45	0.102	9.48	0.103	9.17	0.085
Central Missouri	8.95	0.071	8.95	0.115	8.47	0.075	8.39	0.128	9.49	0.103	9.10	0.104
Missouri Southern State	8.61	0.010	8.57	0.063	8.03	0.015	7.94	0.078	9.20	0.047	8.75	0.048
Western Missouri State	8.30	0.112	8.26	0.104	7.68	0.119	7.74	0.100	8.99	0.066	8.56	0.053
Lincoln	8.37	0.146	8.73	0.020	7.72	0.179	7.97	0.024	9.32	0.016	8.77	0.015
Harris Stowe State	8.20	0.063	6.08	0.010	7.75	0.030	6.07	0.013	8.70	0.002	8.27	0.002
N		2486				1466				32680		26503

Notes: The "counterfactual" sorting columns for African Americans show predicted enrollment shares and index values when the sorting parameters estimated for whites are applied to African Americans per the procedure discussed in the text. For the counterfactual columns the index values are weighted averages where the predicted enrollment probabilities serve as the weights. The table shows the counterfactual that incorporates college and major sorting prior to adjusting high school quality or pre-entry skills for African Americans. We collapse STEM and non-STEM cells within universities for ease of presentation here; Appendix Table C.2 shows counterfactual sorting changes at the university-by-major level.

Table 7. Predicted African American Graduation Rate Increases From Three Partial Sorting Scenarios.

	Predicted African American Graduation Rate	
	Women	Men
Predicted Baseline Graduation Rate:	51.2%	43.1%
Total Gain from College Sorting Adjustment performed in one step (prior to major sorting and high school quality and pre-entry skills adjustments, see Table 5):	2.2	1.0
Predicted Graduation Rate Increase from Partial-Sorting Scenario 1: Hold African American Enrollment Fixed at Urban and Least Selective Schools and Reallocate Remaining African American Students According to the White Assignment Rules	-0.4	-0.5
Predicted Graduation Rate Increase from Partial-Sorting Scenario 2: Hold African American Enrollment Fixed at Most and Least Selective Schools and Reallocate Remaining African American Students According to the White Assignment Rules	+1.4	+0.6
Predicted Graduation Rate Increase from Partial-Sorting Scenario 3: Hold African American Enrollment Fixed at the Most Selective and Urban Schools and Reallocate Remaining African American Students According to the White Assignment Rules	+1.3	+0.5

Notes: The three partial sorting scenarios are not constrained to sum to the total effect. Section 6.1 describes the procedure used to perform these calculations.

Table 8. Decompositions of the Graduation Gaps. Information about Entering Major is Entirely Removed from all Models.

	Women	Men
White predicted graduation rate	66.2	61.1
African American predicted graduation rate	51.3	43.0
Total gap (percentage points)	14.9	18.1
Portion of gap explained by:		
College Sorting	2.0 (13.4%)	1.0 (5.3%)
High school quality adjustment	3.0 (20.1%)	1.8 (9.7%)
Pre-entry skills adjustment	9.9 (66.4%)	15.4 (85.0%)

Notes: See notes for Table 5. The predicted graduation rates for African American women and men differ nominally from the predictions reported in Table 5 as a result of the change to the graduation model.

Table 9. Decompositions of the Graduation Gaps. Alternative Ordering.

	Women	Men
White predicted graduation rate	66.2	61.1
African American predicted graduation rate	51.2	43.1
Total gap (percentage points)	15.0	18.0
Portion of gap explained by:		
High school quality adjustment	2.6 (17.0%)	1.5 (8.5%)
Pre-entry skills adjustment	8.8 (58.5%)	14.2 (79.1%)
College Sorting	3.6 (23.7%)	2.5 (14.1%)
Major Sorting	0.1 (0.8%)	-0.3 (-1.7%)

Notes: With this alternative decomposition, colleges and majors are first held fixed and high school quality and pre-entry skills are adjusted such that African Americans and whites have similar distributions. Next, individuals are sorted into universities given the new high school qualities and pre-entry skills. Finally, individuals are sorted into majors, which in turn affects the sorting into universities.

Table 10. Decompositions for Alternative Graduation Outcomes: Graduation in Five Years and Six Years; and Eight Years from the Initial College.

	Graduation Within 5 Years		Graduation Within 6 Years		Graduation Within 8 Years from the Initial College	
	Women	Men	Women	Men	Women	Men
White predicted graduation rate	56.8	47.3	63.1	56.4	58.2	54.2
African American predicted graduation rate	37.3	25.3	46.0	37.4	44.4	38.4
Total gap (percentage points)	19.5	22.0	17.1	19.0	13.8	15.8
Portion of gap explained by:						
College Sorting	1.5 (7.7%)	0.2 (0.9%)	2.0 (11.7%)	0.7 (3.7%)	1.6 (11.6%)	0.7 (4.4%)
Major Sorting	0.8 (4.1%)	0.5 (2.3%)	0.4 (2.4%)	0.2 (1.1%)	0.4 (2.9%)	0.2 (1.3%)
High school quality adjustment	2.2 (11.3%)	0.5 (2.3%)	2.3 (13.5%)	0.8 (4.2%)	2.2 (15.9%)	1.1 (7.0%)
Pre-entry skills adjustment	15.0 (76.9%)	20.8 (94.5%)	12.4 (72.5%)	17.3 (91.1%)	9.6 (69.6%)	13.8 (87.3%)

Notes: See notes for Table 5.

Appendix A
Model Appendix

A.1 Class Rank Adjustment

Here we describe how we adjust class ranks for African Americans as they are moved to stronger high schools. Let A_{ih} denote the academic background of student *i* who attended school *h* and let $\bar{A}_h$ denote the average academic background at school *h*. Then G_{ih}, the student's (transformed) class rank, can be expressed as:

$$G_{ih} = A_{ih} - \bar{A}_h \tag{A.1}$$

We further decompose academic background in the following way:

$$A_{ih} = A_i^* + \pi b_i + \theta_h^* \tag{A.2}$$

Academic background net of the effect of race (b_i) and the effect of the high school is then given by A_{ih}. Due to African Americans coming from disadvantaged households we would expect π to be negative.

Substituting (A.2) into (A.1) implies that class rank can be expressed as:

$$G_{ih} = \pi(b_i - \bar{b}_h) + (A_{ih}^* - \bar{A}_h^* - \pi\bar{b}_h) \tag{A.3}$$

where $\bar{b}_h$ is the African American share at high school *h*. Given the assumption of random assignment to high schools conditional on race, the expectation of the second term in parenthesis is the same across races. Hence, we can estimate π using the transformed class rank data.

However, there is a selection problem as we only observe class rank for those who enroll in college. We address the selection problem with a standard Heckman selectivity correction. In the first stage we draw on enrollment data from the K-12 public system to predict selection into our sample for each individual as a function of race, high school, and race interacted with high school.[28] We then estimate equation (A.3) with the selectivity term added to the equation.

We use our estimate of π to adjust African Americans' class ranks after improving the quality of their high schools. The class rank adjustment depends on changes to the racial compositions of high schools for African Americans that correspond to aligning the distributions of high school fixed effects between races. The alignment of high school racial compositions is

[28] We use administrative enrollment data from the Missouri Department of Elementary and Secondary Education to estimate the selection model.

concurrent with the alignment of the high school fixed effects.[29] We subtract $\pi(\bar{b}_h^w - \bar{b}_h^b)_i$ from each African American's class rank to make the adjustment, where $\bar{b}_h^w$ and $\bar{b}_h^b$ are the African American enrollment shares at the counterfactual and original high schools, respectively.[30]

A.2 Decomposition

Having described the estimation procedure for the five components of the model – probability of graduating, sorting into universities, sorting into majors, differences in high school quality, and differences in academic background – we now show how these components can be used to decompose differences in African American and white graduation probabilities into four components. In particular, given the assumptions we have made about the academic index and selection into high schools (and therefore obtaining an upper bound estimate on the effect of high schools), equation (1) becomes:

$$\Pr(y=1|r) = \sum_{x\in X}\sum_{h\in H}\sum_{AI}\sum_{m\in M}\sum_{c\in C}\Pr(y=1|c,m,AI)\Pr(c|m,AI,r)\Pr(m|AI,r)\Pr(AI|h,x,r)\Pr(h|r)\Pr(x|r) \quad \text{(A.4)}$$

The total difference in graduation probabilities between African Americans and whites is given by:

$$\begin{aligned} D_r = {} & \sum_{x\in X}\sum_{h\in H}\sum_{AI}\sum_{m\in M}\sum_{c\in C}\Pr(y=1|c,m,AI)\Pr(c|m,AI,b)\Pr(m|AI,b)\Pr(AI|h,x,b)\Pr(h|b)\Pr(x|b) - \\ & \sum_{x\in X}\sum_{h\in H}\sum_{AI}\sum_{m\in M}\sum_{c\in C}\Pr(y=1|c,m,AI)\Pr(c|m,AI,w)\Pr(m|AI,w)\Pr(AI|h,x,w)\Pr(h|w)\Pr(x|w) \end{aligned} \quad \text{(A.5)}$$

The decomposition is then done in the following steps:

Step 1: Differences in the probability of graduating due to *c*:

$$\begin{aligned} D_c = {} & \sum_{x\in X}\sum_{h\in H}\sum_{AI}\sum_{m\in M}\sum_{c\in C}\Pr(y=1|c,m,AI)\Pr(c|m,AI,b)\Pr(m|AI,b)\Pr(AI|h,x,b)\Pr(h|b)\Pr(x|b) - \\ & \sum_{x\in X}\sum_{h\in H}\sum_{AI}\sum_{m\in M}\sum_{c\in C}\Pr(y=1|c,m,AI)\Pr(c|m,AI,w)\Pr(m|AI,b)\Pr(AI|h,x,b)\Pr(h|b)\Pr(x|b) \end{aligned} \quad \text{(A.6)}$$

[29] Specifically, each high school's racial composition is carried through during the process of aligning the distributions of high school fixed effects. This allows, for example, for a move between the African American and white distributions of high school fixed effects at the 10th percentile to be associated with a different change in the African American enrollment share than a move at the 90th percentile.

[30] High school enrollment shares by race are only available for public K-12 students. For private high school attendees we set the African American enrollment shares at the sample averages for public high school attendees, by race. Reasonable adjustments to how we handle private high school attendees do not qualitatively affect our findings because most students in the sample attend public high schools (approximately 12 percent of the analytic sample attended a private high school).

Here we change how African Americans are assigned to colleges but keep their initial majors, high school quality and academic background characteristics the same.

Step 2: Differences in the probability of graduating due to *m:*

$$D_m = \sum_{x\in X}\sum_{h\in H}\sum_{AI}\sum_{m\in M}\sum_{c\in C}\Pr(y=1|c,m,AI)\Pr(c|m,AI,w)\Pr(m|AI,b)\Pr(AI|h,x,b)\Pr(h|b)\Pr(x|b) - \sum_{x\in X}\sum_{h\in H}\sum_{AI}\sum_{m\in M}\sum_{c\in C}\Pr(y=1|c,m,AI)\Pr(c|m,AI,w)\Pr(m|AI,w)\Pr(AI|h,x,b)\Pr(h|b)\Pr(x|b) \quad \text{(A.7)}$$

Here we compare how graduation rates would change due to changing the way African Americans choose their initial majors given the college assignment rules of whites. Note that there are two effects of changing initial major: the direct effect on graduation rates and the indirect effect through college assignment.

Step 3: Differences in the probability of graduating due to *h*:

$$D_h = \sum_{x\in X}\sum_{h\in H}\sum_{AI}\sum_{m\in M}\sum_{c\in C}\Pr(y=1|c,m,AI)\Pr(c|m,AI,w)\Pr(m|AI,w)\Pr(AI|h,x,b)\Pr(h|b)\Pr(x|b) - \sum_{x\in X}\sum_{h\in H}\sum_{AI}\sum_{m\in M}\sum_{c\in C}\Pr(y=1|c,m,AI)\Pr(c|m,AI,w)\Pr(m|AI,w)\Pr(AI|h,x,b)\Pr(h|w)\Pr(x|b) \quad \text{(A.8)}$$

Here we compare how graduation rates would change due to changing the way African Americans are assigned to high schools, allowing for the direct effect of improved high school quality as well as the indirect effect through choice of college and major, assuming African Americans choose initial colleges and majors in the same way as whites.[31]

Step 4: Differences in the probability of graduating due to *x*:

$$D_x = \sum_{x\in X}\sum_{h\in H}\sum_{AI}\sum_{m\in M}\sum_{c\in C}\Pr(y=1|c,m,AI)\Pr(c|m,AI,w)\Pr(m|AI,w)\Pr(AI|h,x,b)\Pr(h|w)\Pr(x|b) - \sum_{x\in X}\sum_{h\in H}\sum_{AI}\sum_{m\in M}\sum_{c\in C}\Pr(y=1|c,m,AI)\Pr(c|m,AI,w)\Pr(m|AI,w)\Pr(AI|h,x,w)\Pr(h|w)\Pr(x|w) \quad \text{(A.9)}$$

Finally, we calculate how differences in academic background characteristics affect the gap in graduation rates, again taking into account the direct and indirect effects. Note that the sum of the four effects exactly accounts for the difference in graduation rates.

[31] Note that selection into high schools is written as not depending on *x*. However, the way this adjustment works is such that African American students who were at the worst high schools will be shifted to the worst white high schools. The key point is that the white and African American distributions across high schools will be the same after this adjustment.

Appendix B
Data Appendix

Appendix Table B.1. Construction of the Analytic Sample.

	Records Lost	Remaining Sample
First-time entering freshman at one of the 13 public, four-year campuses in Missouri between 1996 and 2001[a]		106,747
Assigned to a non-Missouri county of residence, or a foreign country, or county of residence unknown	-17,249	89,498
Not full time upon entry (less than 12 credit hours attempted in first semester)	-10,113	79,385
Older than 20 at the beginning of the fall semester	-3,515	75,870
Unknown high school[b]	-1,886	73,984
Missing high school percentile rank[b]	-2,601	71,383
Missing ACT scores (math or reading)	-566	70,817
Race other than white or African American (including unknown)	-4,008	66,809
Unspecified combination major[c]	-3,418	63,391
Other data restrictions[d]	-256	63,135

[a] Students who enter with sophomore status are included as long as they did not transfer from a previous university (in Missouri or elsewhere). This is to facilitate high school students who take college credits prior to entry.

[b] High school codes are available for most public and private high schools in the state. Not all high schools report class ranks, but most do.

[c] The pre-entry characteristics of unspecified dual majors suggest that this group includes both STEM and non-STEM students. We were unable to convincingly divide this group into major types.

[d] Students from high schools that sent fewer than five students to college over the course of the data panel were dropped, as well as a small number of individuals who were coded as entering non-STEM majors at UM-Rolla (less than 50).

Appendix Table B.2. Top Five Majors in STEM and non-STEM Categories.

	Major Share of Category (STEM or non-STEM)
<u>Top STEM Majors</u>	
General Biology	0.248
General Engineering	0.203
Computer and Information Sciences	0.154
Chemistry	0.040
Pre-Medicine	0.030
<u>Top non-STEM Majors</u>	
Undeclared	0.285
General Business	0.098
Teacher Education	0.054
General Psychology	0.045
Business Administration and Management	0.034

Appendix Table B.3. Comparison of students entering as undeclared majors to those entering with STEM and other non-STEM majors.

	Undeclared	Other non-STEM	STEM
HS Percentile Rank	64.44	70.34	78.12
ACT Math	21.27	22.01	25.58
ACT Reading	23.26	24.23	26.12
Graduation Rate	56.92	64.13	65.95
Science Degree Completion Rate	5.48	1.91	39.00
N	14079	35316	13740

Description of Appendix Tables B.2 and B.3

Appendix Table B.2 shows the top five majors for the STEM and non-STEM groups. Just over 6 out of every 10 STEM majors come from general biology, general engineering and computer/information sciences. The remaining STEM entrants are spread out across a large number of smaller fields.

The most common non-STEM major category includes undeclared entrants. *Ex ante*, it was not clear that these individuals should be categorized as non-STEM entrants. But after examining their pre-entry characteristics we concluded that they were a much better fit as non-STEM than STEM majors, despite being somewhat negatively selected even among non-STEM majors. A notable characteristic of undeclared majors is that they rarely complete a STEM degree (5.48 percent). Although their STEM degree completion rate is higher than declared non-STEM majors, which is perhaps not surprising given that some undeclared majors may have a preference for STEM fields, it is still very far below the STEM degree completion rate for STEM entrants (this is true conditional on general graduation, or unconditionally).

Appendix C
Additional Sorting Details

Table C.1. Average ACT Scores, High School Class Rank, and High School Fixed Effects for African Americans and Whites, by University. Genders and Initial Majors are Combined.

	African Americans						Whites					
	ACT Composite	ACT Math	ACT Reading	Class Rank	HS Fixed Effect	N	ACT Composite	ACT Math	ACT Reading	Class Rank	HS Fixed Effect	N
Truman State	23.44	22.40	24.55	72.33	0.081	146	27.01	25.84	28.24	79.93	0.371	5690
UM-Rolla	23.06	24.05	22.38	73.83	-0.174	80	27.70	28.38	27.75	83.97	-0.021	2317
UM-Columbia	22.25	21.15	23.09	73.34	-0.248	908	25.88	25.05	26.74	78.94	0.217	13226
UM-KC	20.30	19.11	20.71	72.89	-0.595	271	25.02	23.81	25.73	78.46	-0.056	1921
UM-STL	20.31	19.24	21.13	70.19	-0.620	226	24.14	23.29	24.97	71.32	0.232	1662
MO State	19.91	18.60	21.02	62.94	-0.272	251	23.15	21.92	24.04	69.62	-0.128	11104
Northwest	18.82	17.65	19.16	59.95	-0.468	133	22.11	20.99	22.86	67.78	-0.063	4095
Southeast	19.78	18.76	20.31	52.28	-0.131	327	22.66	21.51	23.44	64.96	0.071	5618
Central	19.01	17.90	19.77	60.44	-0.670	286	22.08	21.00	22.84	66.46	-0.097	6129
Southern	18.54	17.11	19.67	48.06	-0.481	46	21.84	20.64	22.74	65.70	-0.455	2807
Western	16.48	15.91	16.68	45.10	-0.687	453	20.35	19.32	21.04	56.75	-0.181	3565
Lincoln	16.41	15.87	16.92	44.96	-0.605	625	20.07	19.46	20.34	53.39	0.370	928
Harris	18.65	17.67	19.13	47.30	-0.698	200	21.12	19.79	22.17	48.54	-0.069	121

Appendix Table C.2. Expansion of the Differences between Actual and Counterfactual African American Sorting (from Table 6), Overall and by Gender.

	Counterfactual Sorting	Actual Sorting	Difference	Difference (Women)	Difference (Men)
Truman State STEM	0.013	0.009	0.004	0.005	0.002
Truman State non-STEM	0.049	0.028	0.021	0.028	0.008
UM-Rolla STEM	0.021	0.020	0.001	0.002	-0.001
UM-Columbia STEM	0.037	0.078	-0.041	-0.043	-0.039
UM-Columbia non-STEM	0.114	0.151	-0.037	-0.049	-0.018
UM-Kansas City STEM	0.010	0.028	-0.018	-0.017	-0.020
UM-Kansas City non-STEM	0.017	0.041	-0.024	-0.028	-0.017
UM-St. Louis STEM	0.003	0.008	-0.005	-0.006	-0.003
UM-St. Louis non-STEM	0.020	0.049	-0.029	-0.035	-0.019
Missouri State STEM	0.016	0.006	0.010	0.007	0.014
Missouri State non-STEM	0.188	0.057	0.131	0.139	0.116
Northwest STEM	0.010	0.006	0.004	0.006	0.001
Northwest non-STEM	0.066	0.027	0.039	0.044	0.029
Southeast STEM	0.010	0.012	-0.002	-0.004	0.001
Southeast non-STEM	0.101	0.070	0.031	0.036	0.020
Central STEM	0.017	0.015	0.002	-0.003	0.012
Central non-STEM	0.103	0.058	0.045	0.047	0.042
Southern STEM	0.006	0.001	0.005	0.003	0.008
Southern non-STEM	0.063	0.011	0.052	0.051	0.054
Western STEM	0.011	0.015	-0.004	-0.005	-0.003
Western non-STEM	0.093	0.099	-0.006	-0.002	-0.016
Lincoln STEM	0.001	0.014	-0.013	-0.010	-0.017
Lincoln non-STEM	0.020	0.144	-0.124	-0.116	-0.139
Harris Stowe non-STEM	0.012	0.051	-0.039	-0.052	-0.017
STEM/non-STEM Majors	0.156 / 0.844	0.212 / 0.788	-0.056 / +0.056	-0.064 / +0.064	-0.044 / +0.044

Notes: This table expands on the differences in African American enrollment shares between the actual and counterfactual sorting scenarios shown in Table 6. See notes for Table 6.

Description of Appendix Tables C.1 and C.2

Appendix Tables C.1 and C.2 provide additional details about college and major sorting in the Missouri system. Table C.1 provides extended selectivity information for each university based on actual enrollment data. Table C.2 expands on Table 6 in the main text. One feature of the re-sorting highlighted by Table C.2 is the system-wide shift in African American enrollment from STEM to non-STEM majors.

Appendix D
Parameter Estimates from the Graduation Model

Appendix Table D.1. Raw Logit Parameter Estimates from the Graduation Model.

Index Parameters	
Normalized High School Class Rank	13.41
	(0.430)**
ACT Math Score	0.006
	(0.003)*
ACT Reading Score	-0.012
	(0.002)**
White Female	-0.062
	(0.022)**
African American Male	0.058
	(0.070)
African American Female	0.124
	(0.060)
Other Parameters	
STEM Entrant	-1.783
	(1.045)†
Truman State Entrant	-0.078
	(0.760)
Truman State STEM Entrant	6.044
	(1.707)**
UM-Rolla Entrant (STEM Only)	1.363
	(1.261)
UM-Columbia Entrant	0.290
	(0.470)
UM-Columbia STEM Entrant	-0.121
	(1.262)
UM-Kansas City Entrant	2.016
	(0.85)*
UM-Kanas City STEM Entrant	0.806
	(1.665)
UM-St. Louis Entrant	0.129
	(0.799)
UM-St. Louis STEM Entrant	4.410
	(1.901)*
Northwest Entrant	1.257
	(0.534)*
Northwest STEM Entrant	1.095
	(1.592)
Southeast Entrant	1.271
	(0.462)**
Southeast STEM Entrant	0.535
	(1.561)
Central Entrant	1.186
	(0.460)*
Central STEM Entrant	1.820
	(1.342)
Southern Entrant	0.792
	(0.599)
Southern STEM Entrant	-0.848
	(2.157)
Western Entrant	-0.129
	(0.539)
Western STEM Entrant	0.873
	(1.682)
Lincoln Entrant	2.576
	(0.609)**
Lincoln STEM Entrant	0.313
	(2.566)
Harris Stowe State Entrant (non-STEM Only)	2.047
	(1.626)

Index Interaction Parameters	
STEM Entrant	0.146
	(0.110)
Truman State Entrant	0.061
	(0.079)
Truman State STEM Entrant	-0.631
	(0.177)**
UM-Rolla Entrant (STEM Only)	-0.087
	(0.130)
UM-Columbia Entrant	0.013
	(0.049)
UM-Columbia STEM Entrant	0.013
	(0.130)
UM-Kansas City Entrant	-0.248
	(0.085)**
UM-Kanas City STEM Entrant	-0.054
	(0.170)
UM-St. Louis Entrant	-0.075
	(0.082)
UM-St. Louis STEM Entrant	-0.465
	(0.194)**
Northwest Entrant	-0.099
	(0.056)†
Northwest STEM Entrant	-0.126
	(0.167)
Southeast Entrant	-0.111
	(0.047)*
Southeast STEM Entrant	-0.105
	(0.160)
Central Entrant	-0.125
	(0.046)*
Central STEM Entrant	-0.165
	(0.141)
Southern Entrant	-0.077
	(0.064)
Southern STEM Entrant	0.064
	(0.225)
Western Entrant	-0.009
	(0.057)
Western STEM Entrant	-0.081
	(0.177)
Lincoln Entrant	-0.274
	(0.062)**
Lincoln STEM Entrant	-0.006
	(0.269)
Harris Stowe State Entrant (non-STEM Only)	-0.330 (0.166)*
Constant	-9.090
	(0.306)**
N	63,135

Notes: Parameters that allow for differential outcomes across university-by-major cells in different years, and high school fixed effects, are suppressed for brevity. The omitted university is Missouri State University. The baseline parameters for STEM-entrant and the STEM-entrant interaction with the index apply to Missouri State. Parameters for the other university-by-major cells are relative to the Missouri State baseline. The net effects for STEM entrants at any university can be obtained by summing the general-entrant effect and the STEM-entrant effect for that university.

Appendix E
African American Predicted and Actual Graduation Rates and African American Representation, by University

Appendix Table E.1. Differences between Actual and Model-Predicted Graduation Rates for African Americans, by University. Genders and Initial Majors Combined.

	Predicted Grad Rate	Grad Rate	Gap	N
Truman State	0.74	0.66	0.08	146
UM-Rolla	0.65	0.71	-0.06	80
UM-Columbia	0.67	0.66	0.01	908
UM-Kansas City	0.46	0.46	0.00	271
UM-St. Louis	0.43	0.37	0.06	226
Missouri State	0.54	0.56	-0.02	251
Northwest Missouri State	0.54	0.53	0.01	133
Southeast Missouri State	0.49	0.48	0.01	327
Central Missouri	0.50	0.54	-0.04	286
Missouri Southern State	0.33	0.44	-0.11	46
Western Missouri State	0.29	0.32	-0.03	453
Lincoln	0.32	0.32	0.00	625
Harris Stowe State	0.27	0.31	-0.05	200
Totals	0.48	0.48	-0.00	3952

Notes: Predictions are based on actual African American Enrollment.

Description of Appendix Table E.1

In Table E.1 we compare actual and predicted African American graduation rates across the 13 universities in the system. Note that our graduation model does not explicitly allow for racial differences in the intercepts, or for the returns to the academic index, by university. There is little evidence to suggest that universities with higher African-American representation produce better outcomes for African Americans. Take the comparison of Western Missouri State and Lincoln University as an example. These two universities are similar along many dimensions, but Lincoln University has a much higher proportion of African American students. The differences between actual and predicted graduation rates for African Americans at these schools are similar. In fact, if anything African Americans are more likely to outperform the model's prediction at Western Missouri State. Missouri State University is another interesting example. African Americans who enter Missouri State do better than the model predicts despite the small African American enrollment share.