COLLEGE MAJOR COMPETITIVENESS AND ATTRITION FROM THE SCIENCES

Marc Luppino and Richard Sander
October 27, 2012

Abstract

This paper examines how the competitiveness of distinct college majors at a student's college affects her major choice and other college outcomes. To mitigate the selection problem, we control for very flexible application-admissions pattern fixed effects to account for student unobservables, as well as school-specific fixed effects to account for typically unobservable institutional characteristics that are plausibly correlated with peer quality and student outcomes. We find that students initially interested in pursuing a science major respond to the competitiveness of both the broad science and non-science major tracks. Weaker, non-minority students typically respond to greater competition in the sciences by shifting their major choice. Under-represented minorities tend to persist in the sciences regardless of competition, but suffer – often substantially -- in terms of college grades and the likelihood of graduating. JEL Codes: I21, J24

* The views in this paper are those of the author and do not necessarily reflect those of the Federal Trade Commission.

I. INTRODUCTION

In the United States, ethnicity is a strong predictor of who becomes a scientist. While young whites are not quite twice as likely as young blacks to earn a bachelor's degree, whites are nearly seven times as likely as blacks to earn a doctorate in science (Bureau of the Census, 2003). A number of social scientists, and more recently the U.S. Commission on Civil Rights, have suggested that at least part of this disparity can be explained by a phenomenon called "science mismatch." The argument is that student learning, interest, and performance are all affected by one's peers; stronger peers often have positive effects, but if a student's peers in, say, college courses are on average much better prepared than the student is, then the student will fall behind, learn less, and perform badly. In the sciences, this problem may be particularly severe because professors presuppose a particular level of technical skill in each course, courses are graded rigorously, and many courses build on prerequisites in the same sequence. Large racial preferences granted by elite college admissions offices may thus have the effect of placing a large proportion of the most talented black and Hispanic science students in environments where they are likely to struggle and either fail to graduate or switch to a non-science major.

A good deal of research has documented the high attrition of blacks (and sometimes Hispanics) from the sciences at elite schools (e.g., Astin and Astin 1993; Elliott and Strenta, 1996). But there are significant challenges in fully testing the mismatch hypothesis: one must compare students facing different levels of peer competition while at the same time controlling for the academic preparation of those students and for the unique characteristics of each college campus. In this paper, we take advantage of a very large dataset covering all freshmen enrollees at the eight undergraduate campuses of the University of California over a nine-year period.

These campuses have many functional similarities – they are all public, they are similar in size, they have common entrance requirements and a common administration – but they embrace a wide range of student competitiveness. At UC Berkeley, the median student has SAT scores that place her at the 91st^{h} percentile of all American students taking the SAT; at the least elite UC campuses, the median student has SAT scores that place her at the 62ndth percentile. While all UC students are academically stronger than the average American college student, this range of entering credentials is much broader than that observed in most other studies of peer effects. Because of the extremely rich data we have on application and admission patterns across all eight campuses, we are able to effectively address selection problems by building upon and improving a strategy used by Dale and Krueger (2002) in their well-known study of the effects of college eliteness on earnings. Specifically, we control for very flexible application-admissions pattern fixed effects to account for student unobservables, as well as school-specific fixed effects to account for typically unobservable institutional characteristics that are plausibly correlated with peer quality and student outcomes. We find strong support for the role of peer effects, and significant support for the mismatch effect as it is usually defined. We also suggest issues to pursue in future research.

II. RELATED LITERATURE

The notion that at college it might be better to be a big frog in a little pond, rather than the reverse, has been around in sociology since the 1960s (Davis, 1966), and in economics since at least the early 1970s (Sowell, 1972). But empirical research on peer effects in college only began in the mid-1990s (e.g., Loury and Garman, 1995), and has accelerated in recent years.

A range of studies have consistently found that black, Hispanic, and white high school seniors are more or less equally likely to aspire to careers in the sciences or engineering, often referred to generally as "STEM" fields (for "Science, Technology, Engineering and Mathematics) (Astin and Astin, 1993).[1] Elliott et al (1996) and Arcidiacono et al (2011) have both shown that white, black, and Hispanic college freshmen at very elite schools all intend to major in STEM fields at similar rates, but that Hispanics and particularly blacks leave those fields at much higher rates. Consequently blacks end up being about half as likely as whites at these elite schools to secure a bachelor's degree in a STEM field.

One potential explanation for these different attrition rates by race/ethnicity is that racial preferences in college admissions have tended to push minorities into academic environments in which their lower level of high school preparation has made it more difficult for them to compete. This disparity between a student's own level of academic preparation and that of her peers is typically referred to as 'mismatch' within the literature. Smyth and McArdle (2004), focus in some depth on the question of whether mismatch particularly affects college student success in the sciences. Using longitudinal data from students at twenty-eight colleges, ranging from moderately elite to very elite, they conclude that preference programs tend to put STEM-interested black and Hispanic students into college environments where it was harder for them to compete effectively; had all the black and Hispanic students in their sample enrolled at schools where their credentials were close to the class-wide averages, then 45% more of the female minorities, and 35% more of the male minorities, would have completed STEM degrees. Conversely, Ost (2010) examines longitudinal data from a single large elite research university

[1] Asian-American high school seniors are substantially more likely than other groups to aspire to STEM careers, and go on to very high rates of college study, college graduation, and graduate study in STEM fields

and finds that the students that are less likely to persistent in the sciences particularly benefit by taking courses with more persistent science peers.

Studies that attempt to identify mismatch effects face the strong challenge of dealing with endogeneity resulting from the college admission and enrollment process. Specifically, mismatch is typically measured by using observable measures of student relative ability and academic preparation. However, students with weaker observables that attend more competitive institutions are more likely to have stronger unobservables than students with comparable observable characteristics who attend less competitive schools.

Because of concerns about potential selection bias resulting from non-random college enrollment patterns, research on more general peer effects in higher education have tended to examine particular academic settings where certain peer groups are randomly or quasi-randomly assigned. Using data from a middle-sized public university in southern Italy, Brunello et al. (2010) find that students with academically stronger roommates achieved significantly higher grades in the hard sciences; they also found no measureable peer effect upon grades achieved in the humanities and social sciences. Similarly, Carrell et al. (2009) find positive academic peer effects resulting from squadron assignment at the United States Air Force Academy, with such effects being largest in math and science courses. Additionally, these authors find that lower ability students appear to particularly benefit from having higher-quality peers in their assigned squadron. Other studies, such as Sacerdote (2001) and Foster (2006), that have looked at roommate and residential peer groups find no or limited resulting effects at other particular academic institutions.

A necessary limitation of studies that focus on roommates or smaller peer groups is that they may not provide us with relevant information that would allow us to project how an individual's outcomes would differ if his or her entire cohort of peers were to change (as would result from a change in college attended). Specifically, the peer dynamics that occur between roommates and within smaller college peer groups (which might promote group study, development of good study habits, and peer tutoring) are likely to be quite different from those that occur at the more macro peer group level (where peer competition in the classroom is more likely to be a driving force). In our analysis we focus on major groupings as the relevant peer groups that determine persistence in the sciences. The key contribution of our work is that we are able to take advantage of a rich dataset that allows us take a number of steps to ensure that our findings are not contaminated by selection bias. Furthermore, this data allows us to examine major group peer effects in a broader range of academic institutions.

III. DATA

The University of California Office of the President (UCOP) maintains extensive and largely consistent databases on every student that has enrolled at the university since 1992. In 2007, a group of economists approached UCOP about studying the effects of Proposition 209 (which banned race preferences at the university for cohorts admitted in 1998 and later) upon academic outcomes. In 2008 and 2010, UCOP released a public-use version of its database. Most relevant for our purposes, the dataset contains nearly forty variables on every freshman applicant to a UC campus from 1995 through 2003. Among the variables are the identity of each campus to which a student applied, which campuses admitted the student, and whether (and where) the student enrolled. The data also includes information on each student's planned field of study and a range of information about high school and standardized test performance. For all enrollees, the dataset includes information on college grades, final field of study, and time to graduation (if the student graduated).

For privacy reasons, UCOP collapsed student observations in a variety of ways. SAT scores are reported in fifty-point ranges, for example, and college GPAs are reported in one-tenth point increments. However, this data does include an exact academic index score, which UCOP constructed as a linear combination of each student's high school GPA, SAT I verbal, and SAT I math scores based on pre-assigned weights. A potential issue with this index score is that the way that it weights these three measures of high school credentials is not necessarily appropriate for an investigation of the determinants of success in the sciences.[2] Therefore, we use these two sets of information on high school credentials to impute precise high school GPA, SAT verbal, and SAT math scores and use these imputed variables in the analysis that follows. Details on the methodology used to impute high school grades and SAT scores are reported in the Appendix.

Another data issue, also arising from UCOP's privacy concerns, is the grouping of student observations into three-year cohorts (the cohorts of interest here are freshmen entering in 1995-97, 1998-2000, and 2001-03). The measures of peer characteristics that we use in our analysis thus pertain to these three-year cohorts instead of more refined entering-class cohorts. UCOP does provide annual data on average entering-class characteristics via its StatFinder webpage.[3] A regression analysis of this data suggests that in using peer variables based on three-year cohorts we lose less than 2 percent of the total variation of comparable peer measures based

[2] In particular, the index weights verbal and math SAT scores equally. However, we argue that math skills are more predictive of success in the sciences and that verbal skills are more predictive of success in other disciplines (and, therefore, possibly of selection out of the sciences).
[3] http://statfinder.ucop.edu/statfinder/default.aspx

on entering-class cohorts. Therefore, it does not appear that the aggregation of students into three-year cohorts meaningfully weakens our analysis.

Finally, the data does not contain information on gender and does not distinguish between black and Hispanic students, categorizing both as under-represented minorities. We would definitely expect student ability and preferences towards the sciences to vary by gender and possibly by minority subgroups. Because we do not observe this information in the data, it is again important that we use an identification strategy that accounts for unmeasured student qualifications and/or preferences that might influence both the selection of a student's enrollment campus and his or her subsequent outcomes.

We have restricted our sample to students that are not missing information on personal characteristics and college outcomes. Summary statistics for this sample are presented in **Table I**. On average, UC students that intend to major in the sciences have stronger high school credentials (in terms of SAT I scores and high school grades) than students intending to pursue a non-STEM major. However, these same students achieve a slightly lower graduation rate and lower cumulative college GPA than students intending to pursue a non-STEM major.

Only 49 percent of intending science majors actually graduate with a science degree. The outflow of students from the sciences is not counterbalanced by a comparable inflow of students from other fields. Only 10 percent of students that do not initially intend to pursue a STEM major end up doing so.

IV. EMPIRICAL STRATEGY

The following framework guides our thinking about how major peer groups affect college outcomes.

IV.A. Conceptual Framework

In deciding on their college major, students are faced with a degree progress production function:

$$d_{ij}^{*} = d(a_{ij}, \overline{a}_{-i(j),j}, \eta_{ij}) \tag{i}$$

where the degree progress of student *i* in major *j* is determined by her own ability (a) in major *j*, the average major-specific ability of the other students studying major *j* ($-i(j)$), and a random shock η. The student successfully obtains a degree in major *j* if her degree progress exceeds the minimum standard (*k*) for that major set by the college:

$$d_{ij} = 1 \text{ if } d_{ij}^{*} \geq k_{ij} \tag{ii}$$

A student receives current and future utility (*V*)*:*

$$V_{ij} = V(Y_{ij}(d_j, e_{ij}), \alpha_{ij}(d_j, u_{ij})) \tag{iii}$$

where this utility is a function of the pecuniary (*Y*) and non-pecuniary (α) benefits that student *i* receives from studying major *j*. Specifically, we can think of *Y* as a vector of the lifetime earnings stream that student *i* receives from having pursued studies in major *j*. Similarly, we can think of α as a vector of the lifetime non-pecuniary benefit stream that *i* experiences as a result of studying major *j*. Each of these benefit streams potentially depend on whether the student successfully obtains a degree in major *j* and a series of random shocks *e* and *u*. A student chooses a major *j** that maximizes expected current and future utility:

$$j_i^{*} = \arg\max_j E[V_{ij}] \tag{iv}$$

To further simplify things, let us assume that students choose between two majors: STEM and non-STEM. Whether a student obtains a STEM degree is dependent on both her utility-maximizing decision and how successful the student is in meeting the minimum requirements of her desired major. It follows that the reduced form model of this outcome is:

$$d_{i,STEM} = f(a_{i,STEM}, a_{i,NON}, \overline{a}_{-i(STEM),STEM}, \overline{a}_{-i(NON),NON}, k_{i,STEM}, k_{i,NON}, v_{i,STEM}, v_{i,NON}, \varepsilon_i) \qquad \text{(v)}$$

where $v_{i,STEM}$ and $v_{i,NON}$ represent preference parameters of student *i* for each major and ε_i represents a composite random shock.

IV.B. Empirical Specification

In order to estimate an empirical model of STEM degree attainment, we use SAT I math and verbal scores as measures of major-specific skills for STEM and non-STEM majors.[4] The basic specification that we use in our empirical analysis is:

$$y_{ikca} = \alpha_1 math_i + \alpha_2 verbal_i + \alpha_3 \overline{math}_{-i(STEM)kc} + \alpha_4 \overline{verbal}_{-i(NON)kc} + \Gamma' X_{ic} + \mu_k + A_{ac} + \varepsilon_{ikca} \qquad \text{(vi)}$$

where we look at a number of outcomes of interest, y_{ikca}. Our analysis focuses on four key outcomes: whether a student graduates, whether a student intending to major in science persists in science, whether students achieve both science-persistence and graduation, and what GPA a student attains.

The main independent variable of interest, $\overline{math}_{-i(STEM)kc}$, is the average math score of other intending STEM majors at campus *k* in cohort *c*. Similarly, $\overline{verbal}_{-i(NON)kc}$ is the average

[4] Bettinger (2010) finds that math ACT score is the strongest predictor of persistence in the STEM majors, even after controlling for science ACT score. He also finds that the English and reading component of overall ACT score is negatively correlated with the likelihood of staying in a STEM major. Arcidiacano (2004) also finds that high math ability leads college students to select in natural science and business majors, while high verbal ability leads students to select into social science and humanities majors.

verbal score of those intending to study non-STEM majors. We define major peer groups according to the intended major each student states on her college application to her enrollment campus. Given that the within-campus-cohort variance in these peer measures is extremely small, we cluster our standard errors by campus-cohort.

Our main concern in trying to identify the effects of major peer groups on different college outcomes is that students selectively enroll in colleges with different peer characteristics. Peer characteristics are likely to be correlated with other institutional characteristics (e.g. campus resources, faculty quality, academic standards, etc.) that affect student outcomes. In order to account for these institutional differences, our specification includes enrollment campus fixed effects (μ_k).

In addition, the preferences and/or abilities of a student that drive the selection of the enrollment campus are also likely to directly affect his or her subsequent outcomes. Our empirical model includes a vector of observable student characteristics (X_{ic}) that control for high school GPA, parental education, and family income.[5] We also include in our specification very flexible application-admissions pattern fixed effects (A_{ac}) to account for student unobservables. The UCOP data provides us with information on which UC campuses each student submitted applications to, whether the stated major on each application was a STEM major, and whether the application was accepted. For each of the 8 UC campuses, each student has seven possible application-admission outcomes: (1) did not apply, (2) applied with an intended non-STEM major and rejected, (3) applied with an intended STEM major and rejected, (4) applied with an

[5] Family income is reported as a categorical variable and in nominal terms. Therefore, we include in our specification a set of family income category by cohort fixed effects. These fixed effects additionally account for any general time trends in the data.

intended non-STEM major and accepted, (5) applied with an intended STEM major and accepted, (6) applied with an intended non-STEM major and admissions outcome missing, and (7) applied with an intended STEM major and admissions outcome missing. While there are 7 to the 8th power application-admission patterns that are theoretically possible, we only observe 11,281 unique combinations for intending STEM majors and 30,803 for all students in our sample. Additionally, we allow these fixed effects to vary by whether the student was eligible for an admissions preference and by cohort, since the admissions regime employed by the UC system changed over time.[6]

V. RESULTS

We begin by examining the effect of major competiveness on the likelihood that students graduate with a science degree. These results are presented in **Table II**. A student's own high school credentials are highly statistically significant predictors of whether he or she graduates with a science degree. As expected, students with higher math SAT scores are more likely to obtain science degrees while students with higher verbal SAT score are less likely.

When a student attends a campus where the sciences are more competitive, he or she is less likely to graduate with a science degree. Specifically, we find that increasing the average math SAT score of intending science majors by 10 points decreases the likelihood of graduating with a science degree by a little over one percentage point. Conversely, attending a college

[6] Prior to the adoption of Prop 209 and continuing through the 1997 cohort, all underrepresented minorities received admissions preferences which varied by campus. Beginning with the fall of 2001, the UC system guaranteed the top 4 percent of students in the graduating class of every California high school UC eligibility if they had completed 11 specific college prep courses by the end of their junior year. This policy, known as "Eligibility in the Local Context" or ELC, was implemented to encourage students who had excelled academically in disadvantaged high schools to attend UC campuses; we are able to observe the ELC status of students after this policy is adopted.

where the non-sciences are more competitive increases the likelihood that students pursue and obtain a STEM degree. Increasing the average verbal SAT score of students intending non-STEM majors by 10 points increases the likelihood of graduating with a STEM degree by slightly more than one percentage point. Both of these peer effects are highly statistically significant.

The results in Table II and subsequent tables are able to improve upon the identification strategy used in Dale and Kruger (2002). Dale and Kruger used information on student applications, and on which colleges accepted them, to compare students who were accepted by similar schools but in fact attended schools with differing levels of eliteness. We are able to go a step further. Because we have such a large number of observations in the UC dataset, and successive cohorts at each of the institutions we study, we are able to compare students who applied to, and were accepted by, the exact same sets of schools, and we also use college fixed effects, which accounts for potential differences that influence both the enrollment decision of students and their subsequent outcomes. It is worth noting that our results are robust to alternative strategies of accounting for student unobservables.

In Table II, the coefficient estimates for the racial (or ethnic) categories reflect any residual differences across groups that are not explained by the student and institutional characteristics for which we have otherwise controlled. The coefficients suggest that URMs (blacks and Hispanics) are more likely to earn a science degree than comparable white students; but it appears this result is driven by differences across groups in initial “field of study” preferences. In other words, blacks and Hispanics who enter a UC school with strong credentials – credentials close to the white and Asian averages – are much more likely to intend to major in

STEM fields, and thus receive STEM degrees in proportionately higher numbers than whites.

In Table III, we use the model with the most stringent set of controls in Table II (Model 4), and do two things: first, we break students into those who, as high school seniors, intended to major in STEM fields, and those who did not; second, we examine several college outcomes: GPA, graduating, and graduating with a STEM degree, and ending one's UC career (with or without a degree) as a STEM major.

These results tell us several valuable things. First, we can see that the peer-effect results in Table II – although they hold for all students – are largely driven by students who initially intend, upon starting college, to major in STEM fields.[7] That is to say, the coefficients for the "peer ability effects" in Table III are generally highly significant for intending STEM majors and not so for other students.

Table III also allows us to explore the multiple ways in which a student may fail to obtain a science degree. One possibility is that an intending science major changes fields but still goes on to graduate from college. Another is that the student fails to graduate in any field.[8] Whether a student continues with her intention to major in the sciences appears to depend on the ability of similarly interested students at the college. Specifically, increasing the average math SAT score of other intending STEM majors by 10 points increases the likelihood that a student switches to a

---

[7] In these analyses, we have classified a student's intended major based on the choice she made on the college application to the campus in which she eventually enrolled. Some students express different intentions on different applications – they may list a STEM field in their application to UC Davis, but humanities in their application to Berkeley. If one classifies as an "intending STEM" major anyone who indicates a STEM preference on any UC application, this picks up more students – but it does not have much effect on the results in Table III. Conversely, if one classifies as an intended STEM major only those student who indicate a STEM preference on **all** UC applications, this produces a small number of students, but again the results we have shown hold steady.

[8] The UCOP data does not tell us the timing or history of changes to a student's intended field of study. We only observe the student's entering preferences and her final major, which is the last official major registered by the student before she exited the UC system (either by graduating or dropping out).

non-STEM field by two percentage points. Intending science majors also appear to weigh the competitiveness of outside fields at their enrollment campus when making the decision to either stay in or exit the sciences. Increasing the average verbal SAT score of students intending non-STEM majors by 10 points increases the likelihood that an intending STEM major stays in STEM by almost three percentage points.

The strength of one's STEM peers at a college also affects the chances that an intending STEM major will graduate from college with any degree (including a non-STEM degree). Increasing the average math SAT score of other intending science majors by 10 points decreases the likelihood that an intending STEM major graduates from college (in any field) by around 1.8 percentage points. The magnitude of this effect is similar to the effect upon major-switching (leaving a STEM field). One way to think about this is that students who face difficult competition in a STEM field must make a choice: they can switch to a non-STEM field, which will probably entail staying in college longer; they can persist in the major, perhaps with low grades, or they can drop out. One would expect to see each of these results for some students facing a high level of mismatch.

Unfortunately, the UCOP data does not include direct measures of grades within a major or an alternative measure of degree progress within a major that might enable us to closely scrutinize the interaction of grades with the student's choices. While we do know the student's "final GPA" – measured, like final major, when a student graduates or exits from UC – the GPA data is based on very heterogeneous college paths, and therefore is a very noisy measure of learning and achievement. Consider, for example, a student who starts out as a science major, runs into stiff peer competition, and then switches to a major in communications, with a resulting boost to her grades; here stronger science peers could lead to higher final GPAs. Perhaps not surprisingly, we therefore find small peer effects on student grades. While the coefficients on peer effects are negative, as predicted, they are small, and even when they are statistically significant they are driven by subsets of students (as we shall discuss below, in examining Table V).

A priori, we might expect that having strong non-science peers at a college would have mixed, conflicting effects on the likelihood that intending STEM majors will graduate. The greater non-science competition makes it less appealing to bail out of the STEM fields when the going gets rough – perhaps increasing STEM persistence – but correspondingly may make it harder to survive if one does switch to a non-STEM field. And in fact, the coefficients in Table III show a small, positive, and marginally statistically significant effect on the likelihood that intending science majors graduate from college.

There are several interesting conditional racial differences in Table III. While Asian students are more likely to graduate than comparable whites, they tend to earn lower GPAs. This might be driven by the higher proportion of Asian students with language issues (not measured in our data); note the GPA gap is larger in non-scientific fields, where English proficiency is probably more often crucial. However, it also appears that Asian students are more likely to persist in college than otherwise comparable whites; this, too, could produce lower final GPAs. We find no independent effect of "URM" status on one's likelihood of continuing in the science track. But we do find significantly lower graduation rates and GPAs among URMs than among comparable whites – a pattern we return to below.

Finally, it is worth noting that a given increase in a student's math SAT score predicts a larger increase in her cumulative GPA if she intends to major in science than does a comparable increase in their verbal SAT score. The opposite is true for students intending non-STEM majors. Since intending STEM majors are likely to take more science courses and other students more non-science courses, these results seem to validate our use of math and verbal scores as measures of, respectively, science and non-science skills.

Our results suggest that intending science majors have a 4 percentage point greater likelihood of graduating with a science degree (e.g., 50% versus 46%) when they face peer competition levels similar to those found at UCLA during the period of our study, compared with those found at Berkeley. When we break our analysis down by racial/ethnic groups, our results suggest that a similar shift in peer competitiveness increases the likelihood that blacks and Hispanics persist and graduate with science degrees by 11 percentage points.[9] This is striking, because both Berkeley and UCLA are almost universally viewed as "elite" schools, and yet the difference in academic preparation among "peer" students appears to be sufficient to have rather important effects on science completion.

Another interesting lens for considering our results is to consider UC Davis with UC Santa Barbara. These two schools are both viewed as moderately elite colleges, and indeed the academic preparation of science majors at the two schools is very similar. However, the non-science students at UC Santa Barbara have, on average, significantly strong academic preparation than comparable students at UC Davis, and our models suggest that this has significant, positive effects upon science completion at UC Santa Barbara: specifically, a 5.5 percentage point increase in persistence for science-intending students overall, and an 8 percentage point increase in the rate at which blacks and Hispanics persist and graduate with

[9] Corresponding model estimates for distinct racial/ethnic groups are presented in Table V.

science degrees.

V.A. Heterogeneous Effects by Student Ability (Evidence on Mismatch Effects)

To this point, we have found that the college outcomes of intending science students are influenced by peer effects on average. However, we have not considered the possibility that these peer effects might be moderated by a student's own ability. The mismatch literature has specifically suggested that placing students in strong academic environments might have large and negative effects on less academically prepared students. To test for this possibility, we include interaction terms between "own ability" and peer ability in our specifications.

In accord with the mismatch hypothesis, we find that a student's persistence in science is particularly hurt by stronger peers when the student's own math ability is relatively low. As shown in Table IV, the interaction term between a student's own math SAT and the SAT of her peers is substantial and negative. If the math SAT scores of a student's peers rises by 10 points, the student's likelihood of having a STEM "final major" drops 2.4 percentage points if the student's own math SAT is 550, but only 1.9 points if the student's own math SAT is 650. Since admissions preferences within the UC system often have the effect of increasing the math SAT scores of one's peers by 100 points, this is a substantial effect.

While own math ability does appear to moderate the effect that peer math ability has on persistence in science, it does not appear to influence peer effects on the likelihood of graduating or on cumulative college grades. It also does not appear to be the case that own ability moderates the effect of attending a campus with stronger non-STEM peers. Here, we have evaluated whether peer effects are heterogeneous by own ability by including a simple linear interaction term. We find very similar results when we express own ability as a series of dummy variables and interact these dummy variables with average peer ability.

V.B. Heterogeneous Effects by Race

Up until this point, our findings have suggested that weaker students are more likely to switch majors when faced with greater competition in the sciences. Black and Hispanic ("URM") students might seem to be particularly vulnerable; they tend to have lower credentials than their peers, and our earlier analyses showed grade and graduation gaps that may reflect unobserved racial differentials in college preparation. However, our analyses suggest that URM students are less likely than other students to switch from STEM to non-STEM fields; they are more likely to persist than their non-URM peers, but they pay a price in lower graduation rates….

Because of racial preferences in college admissions, it is also important to evaluate how different racial groups are affected by major competitiveness in terms of their college outcomes. Results, broken down by race, are presented in **Table V**. Faced with stronger peers in the non-STEM majors, intending science majors from all racial-ethnic groups are more likely to persist in the sciences. However, white students are much more likely to exit the sciences when faced with stronger peer competition compared to Asians and, particularly, under-represented minorities.

Being more persistent in the face of greater peer competition comes with a cost for minority student groups. The Table V results imply that attending a college with stronger science peers leads Asians, blacks, and Hispanics who are interested in science to have lower cumulative college GPAs and lower odds of graduating from college. The negative effect of strong peers is particularly severe for under-represented minorities (URMs). Increasing the average math SAT score of other intending science majors by 10 points decreases the likelihood that URM intended STEM majors graduate from college by around 4.6 percentage points. The same change in science peer ability decreases the cumulative college GPA of black and Hispanic intended science majors by 0.042 points

Much of the peer-effect literature measures outcomes in terms of first-year grades, while the principal grade outcome we examine is final GPA. Carrel et al (2009) is an exception; these authors found that a 100-point increase in the Verbal SAT scores of freshmen squandron members raised the cumulative grades of U.S. Air Force Academy cadets by approximately 0.25 grade points. Our findings imply that a 100-point increase in the Verbal SAT scores of non-science peers at a University of California campus decreases the cumulative GPAs of black and Hispanic intending science majors by 0.15 grade points. A 100-point increase in the Math SAT scores of science peers lowers the final GPA of black and Hispanic intending science majors by 0.46 points. As noted earlier, our findings are not at all necessarily in conflict with those like Carrel et al, because a squadron is a relatively small and cohesive group, and a squadron member is likely to have many different types of interactions, and thus many more diverse effects, upon one's academic performance than a college classmate on a large campus.

V.C. Robustness to Alternative Definitions of Peer Group and Peer Ability

Our analysis to this point has defined relevant peer groups based on intended major and has defined peer ability by Math SAT scores of science students and Verbal SAT scores of non-science students.—that is, measures of skill relevant to the respective major fields. Our results are not sensitive to these particular definitions. **Table VI** presents results using broader definitions of the relevant peer group(s) and peer ability. The general magnitude, sign, and statistical significance of our estimated effects do not vary much with these measures.

A significant finding of this paper is that peer effects stand out more clearly when we use narrow, rather than broad measures of both peer groups and peer skills. Past research on peer effects has often examined outcomes for entire college classes, using general measures of competitiveness such as overall SAT scores or high school grades. In our analysis, such broad definitions would mask the most important peer effects, which are strong and significant when we split students into science and non-science majors, and consider the differential effects of peers with high math, or high verbal, SAT scores. Scholars examining these issues should be careful to test for disaggregated effects like these.

VI. CONCLUSION

Using a rich dataset on the universe of University of California students who enrolled between 1995 and 2003, we examine how the competitiveness of distinct college majors at a student's campus affects his or her major choices and other college outcomes. We find that students initially interested in pursuing a science major respond to the competitiveness of both the broad science and non-science major tracks. Attending a campus with a stronger group of intending science majors lowers the likelihood that students graduate with a science degree. While some of these students who leave the sciences simply shift their course of study, others fail to graduate at all. We take strong measures to ensure that these findings are not driven by peer group endogeneity.

Consistent with mismatch theories, we find that weaker students are particularly adversely affected by attending colleges where the sciences are more competitive. Perhaps most striking, we find that minorities – especially underrepresented minorities – are much less likely than whites to respond to mismatch by switching majors. Instead, minority students interested in science are much more likely to drop out when they are placed among more competitive peers, and, if they do graduate, they take a very substantial hit on their GPAs. What accounts for this race effect? Perhaps minority students, relative to whites, have a stronger commitment to pursuing science and are willing to bear the risk of not graduating in order to pursue their dream of becoming a scientist. Or perhaps, minority students are less likely to know how to maneuver the college landscape. Perhaps they underestimate the risks of mismatch until they feel they are too far invested in a STEM major to switch, or perhaps they interpret the stiff competition they face as a sign of their own generalized academic weakness. Our data does not allow us to evaluate these alternate hypotheses. Future research in this area should evaluate whether these particular patterns of minority attrition hold in other instances of "science mismatch." The general problem of science attrition, as shown by this and other research, is sufficiently serious that universities should attempt careful measurement of evolving student attitudes and outcomes – especially of students vulnerable to mismatch – as they enter and advance through college.

VII. REFERENCES

Antonovics, Kate and Richard H. Sander (2011). "Affirmative Action Bans and the Chilling Effect." Working Paper, University of California – San Diego.

Arcidiacono, Peter (2004). "Ability Sorting and the Returns to College Major." Journal of Econometrics, 121(1-2): 343-375.

Arcidiacono, Peter, Esteban Aucejo, Patrick Coate, and V. Joseph Hotz (2011). "The Effects of Proposition 209 on College Enrollment and Graduation Rates in California." Working Paper.

Arcidiacono, Peter and Sean Nicholson (2005). "Peer Effects in Medical School." Journal of Public Economics, 89: 327-350.

Astin, Alexander and Helen Astin (1993). Undergraduate Science Education: The Impact of Different College Environments on the Educational Pipeline in the Sciences. Higher Education Research Institute, UCLA.

Bettinger, Eric (2010). "To Be or Not to Be: Major Choices in Budding Scientists." In Charles T. Clotfelter (Ed.), *American Universities in a Global Market*. Chicago :University of Chicago Press.

Betts, Julian R. and Jamie L. Shkolnik (1999). "Key Difficulties in Identifying the Effects of Ability Grouping on Student Achievement." Economics of Education Review, 19(1): 21-26.

Bowen, William G. and Derek Bok (1998). The Shape of the River: Long-Term Consequences of Considering Race in College and University Admissions. Princeton University Press, Princeton, NW.

Brunello, Giorgio, Maria De Paola, and Vincenzo Scoppa (2010). “Peer Effects in Higher Education: Does the Field of Study Matter?” Economic Inquiry, 48(3): 621-634.

Carrell, Scott E., Richard L. Fullerton, and James E. West (2009). “Does Your Cohort Matter? Measuring Peer Effects in College Achievement.” Journal of Labor Economics, 27(3): 439-464.

Dale, Stacy Berg and Alan B. Kreuger (2002). “Estimating the Payoff to Attending a More Selective College: An Application of Selection on Observables and Unobservables.” Quarterly Journal of Economics, 117(4): 1491-1527.

Davis, James (1966). “The Campus as a Frog Pond: An Application of the Theory of Relative Deprivation to Career Decisions of College Men.” 72 American Journal of Sociology 17

Elliott, Rogers, A Christopher Strenta, Russell Adair, Michael Matier, and Jannah Scott (1996). “The Role of Ethnicity in Choosing and Leaving Science in Highly Selective Institutions,” 37 Research in High Education, 68.

Foster, Gigi (2006). “It’s Not Your Peers, and It’s Not Your Friends: Some Progress Toward Understanding the Educational Peer Effect Mechanism.” Journal of Public Economics, 90: 1455-1475.

Frisancho Robles, Veronica C. and Kala Krishna (2012). “Affirmative Action in Higher Education in India: Targeting, Catch Up, and Mismatch at IIT-Delhi.” NBER Working Paper 17727.

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

Ost, Ben (2010). “The Role of Peers and Grades in Determining Major Persistence in the

Sciences." Economics of Education Review, 29: 923-934.

Sacerdote, Bruce (2001). "Peer Effects with Random Assignment: Results from Dartmouth Roommates." Quarterly Journal of Economics, 116(2): 681-704.

Smyth, Frederick L. and John J. McArdle (2004). "Ethnic and Gender Differences in Science Graduation at Selective Colleges with Implications for Admission Policy and College Choice." Research in Higher Education, 45(4): 353-381.

Sowell, Thomas (1972). "Black Education: Myths and Tragedies. David McKay Company, Inc. New York.

Stinebrickner, Todd R. and Ralph Stinebrickner (2011). "Math or Science? Using Longitudinal Expectations Data to Examine the Process of Choosing a College Major." NBER Working Paper 16869.

VIII. TABLES

TABLE I
Sample Means (Standard Deviations in Brackets)

	All Students	Intended Major Science	Intended Major Non-science
Asian	38%	46%	34%
Black or Hispanic	17%	14%	18%
White	37%	32%	39%
Other	8%	8%	9%
Math SAT Score	613	636	601
	[89]	[87]	[88]
Verbal SAT Score	578	580	577
	[93]	[94]	[92]
UC-Adjusted HS GPA	3.77	3.86	3.72
	[0.43]	[0.41]	[0.44]
Declared Science as Final Major	28%	61%	11%
Graduated from College	82%	81%	82%
in the Sciences	24%	49%	10%
Cumulative College GPA	3.01	2.92	3.06
	[0.59]	[0.60]	[0.58]
Observations	241,062	84,466	156,596

Notes: The sample consists of three cohorts of students that enrolled in one of 8 UC campuses for the periods 1995-1997, 1998-2000, and 2001-2003.

TABLE II
Determinants of Graduating with a Science Degree

VARIABLES	(1)	(2)	(3)	(4)
Peer Ability:				
(Average Math SAT Score of Intended Science Majors)/10	-0.0139***	-0.0154***	-0.0192***	-0.0112***
	[0.00359]	[0.00397]	[0.00309]	[0.00227]
(Average Verbal SAT Score of Intended Non-science Majors)/10	0.0159***	0.0188***	0.0178***	0.0125***
	[0.00417]	[0.00397]	[0.00382]	[0.00335]
Own Characteristics:				
(Math SAT Score)/10	0.0121***	0.0118***	0.0118***	0.00691***
	[0.000647]	[0.000636]	[0.000692]	[0.000443]
(Verbal SAT Score)/10	-0.00358***	-0.00352***	-0.00388***	-0.00237***
	[0.000286]	[0.000265]	[0.000241]	[0.000176]
UC-Adjusted High School GPA	0.174***	0.163***	0.153***	0.102***
	[0.0118]	[0.0104]	[0.00925]	[0.00717]
Asian	0.0752***	0.0563***	0.0564***	0.0382***
	[0.00611]	[0.00535]	[0.00584]	[0.00470]
Black or Hispanic	0.0328***	0.0222***	0.0189***	-0.00336
	[0.00473]	[0.00473]	[0.00470]	[0.00430]
Other Non-white Race/Ethnicity	0.0319***	0.0250***	0.0247***	0.0191***
	[0.00426]	[0.00375]	[0.00399]	[0.00370]
College Fixed Effect	Yes	Yes	Yes	Yes
Application Pattern Fixed Effect	No	Yes	No	No
Application-Admission Pattern Fixed Effect	No	No	Yes	No
Application-Admission Pattern Fixed Effect (w/ Application Major Preferences)	No	No	No	Yes
Observations	241,062	241,062	241,062	241,062
R-squared	0.113	0.128	0.160	0.403

Notes: Each column contains coefficient estimates from separate regressions. Each regression also contains controls for parental education and family income by cohort. Robust standard errors, clustered by campus-cohort, are reported in brackets. *** p<0.01, ** p<0.05, * p<0.1.

TABLE III
Determinants of College Outcomes by Intended Major

VARIABLES	Intending STEM Majors Graduated in Sciences	 Final Major Science	 Graduated	 Cumulative GPA	Not Intending STEM Majors Graduated in Sciences	 Final Major Science	 Graduated	 Cumulative GPA
Peer Ability:								
(Average Math Score of Intended STEM Majors)/10	-0.0247***	-0.0200***	-0.0175***	-0.0135	-0.00253	-0.00332	0.00239	0.00562
	[0.00390]	[0.00529]	[0.00356]	[0.00794]	[0.00209]	[0.00223]	[0.00391]	[0.00509]
(Average Verbal Score of Intended Non-STEM Majors)/10	0.0279***	0.0273***	0.00837**	-0.0119*	0.00316	0.00241	0.00144	-0.00928
	[0.00453]	[0.00542]	[0.00344]	[0.00595]	[0.00304]	[0.00302]	[0.00397]	[0.00852]
Own Characteristics:								
(Math SAT Score)/10	0.0115***	0.0113***	0.00158***	0.00850***	0.00494***	0.00527***	0.000491**	0.00265***
	[0.000588]	[0.000615]	[0.000290]	[0.000453]	[0.000396]	[0.000398]	[0.000192]	[0.000381]
(Verbal SAT Score)/10	-0.00242***	-0.00270***	-0.000107	0.00616***	-0.00219***	-0.00241***	7.82e-05	0.00989***
	[0.000322]	[0.000370]	[0.000262]	[0.000394]	[0.000218]	[0.000236]	[0.000242]	[0.000308]
UC-Adjusted HS GPA	0.212***	0.121***	0.162***	0.542***	0.0604***	0.0589***	0.122***	0.462***
	[0.00857]	[0.0117]	[0.00735]	[0.0100]	[0.00632]	[0.00626]	[0.00674]	[0.00872]
Asian	0.0437***	0.0277***	0.0224***	-0.0475***	0.0335***	0.0349***	0.0218***	-0.0741***
	[0.00696]	[0.00678]	[0.00477]	[0.00819]	[0.00413]	[0.00433]	[0.00436]	[0.00618]
Black or Hispanic	-0.0242**	-0.00947	-0.0307***	-0.106***	0.00118	0.000997	-0.0226***	-0.101***
	[0.00905]	[0.00983]	[0.00764]	[0.0134]	[0.00375]	[0.00394]	[0.00579]	[0.0104]
Other non-White Race/Ethnicity	0.0252***	0.0197**	0.00993	-0.0250***	0.0157***	0.0177***	0.00200	-0.0237***
	[0.00835]	[0.00778]	[0.00712]	[0.00800]	[0.00351]	[0.00371]	[0.00206]	[0.00511]
Observations	84,466	84,466	84,466	83,683	156,596	156,596	156,596	153,010
R-squared	0.306	0.289	0.306	0.435	0.266	0.273	0.200	0.392

Notes: Each column contains coefficient estimates from separate regressions. Each regression also contains controls for parental education, family income by cohort, enrollment campus, and UC application-admission pattern fixed effects that account for the intended major listed on each application. Robust standard errors, clustered by campus-cohort, are reported in brackets. *** p<0.01, ** p<0.05, * p<0.1.

TABLE IV
Determinants of College Outcomes for Intended Science Majors
w/ Own Ability-Peer Ability Interactions

VARIABLES	Graduated in Sciences	Final Major Science	Graduated	Cumulative GPA
(Average Math SAT Score of Other Intended Science Majors)/10	-0.0586***	-0.0542***	-0.0165**	-0.0225*
	[0.00720]	[0.00920]	[0.00658]	[0.0112]
(Own Math SAT Score)/10	-0.0217***	-0.0223***	0.00253	9.90e-06
	[0.00571]	[0.00747]	[0.00456]	[0.00700]
(Average Math Score of Other Intended Science Majors * Math Score)/100	0.000526***	0.000531***	-1.50e-05	0.000134
	[8.81e-05]	[0.000116]	[7.21e-05]	[0.000108]
(Average Verbal SAT Score of Intended Non-science Majors)/10	0.0363***	0.0363***	0.00891*	-0.0168***
	[0.00632]	[0.00560]	[0.00500]	[0.00514]
(Own Verbal SAT Score)/10	0.00591	0.00622	0.000462	0.00102
	[0.00590]	[0.00675]	[0.00436]	[0.00563]
(Average Verbal Score of Intended Non-science Majors * Verbal Score)/100	-0.000146	-0.000156	-9.80e-06	8.86e-05
	[0.000100]	[0.000113]	[7.40e-05]	[9.51e-05]
Marginal Effect of 10 Point Increase in Average Math Score of Other Intended Science Majors:				
Own Math SAT Score = 550	-0.0297***	-0.0250***	-0.0173***	-0.0151*
	[0.00412]	[0.00531]	[0.00379]	[0.00829]
Own Math SAT Score = 650	-0.0245***	-0.0197***	-0.0175***	-0.0137
	[0.00396]	[0.00517]	[0.00353]	[0.00812]
Observations	84,466	84,466	84,466	83,683
R-squared	0.306	0.289	0.306	0.435

Notes: Each column contains coefficient estimates from separate regressions. Each regression also contains controls for race/ethnicity, UC-adjusted high school GPA, parental education, family income by cohort, enrollment campus, and UC application-admission pattern fixed effects that account for the intended major listed on each application. Robust standard errors, clustered by campus-cohort, are reported in brackets. *** p<0.01, ** p<0.05, * p<0.1.

TABLE V
Determinants of College Outcomes for Intended Science Majors by Race/Ethnicity

	Asian			Black or Hispanic			White		
VARIABLES	Final Major Science	Graduated	Cumulative GPA	Final Major Science	Graduated	Cumulative GPA	Final Major Science	Graduated	Cumulative GPA
Peer Ability:									
(Average Math Score of Intended STEM Majors)/10	-0.0177*	-0.0171***	-0.0154*	-0.00866	-0.0461***	-0.0418**	-0.0213***	-0.0134	-0.00360
	[0.00943]	[0.00581]	[0.00861]	[0.0189]	[0.0117]	[0.0179]	[0.00579]	[0.00875]	[0.0149]
(Average Verbal Score of Intended Non-STEM Majors)/10	0.0263***	0.00582	-0.0131**	0.0346**	-0.00408	-0.0193*	0.0282***	0.0204*	-0.00964
	[0.00718]	[0.00358]	[0.00569]	[0.0131]	[0.00966]	[0.0107]	[0.00701]	[0.00988]	[0.0125]
Own Ability:									
(Math SAT Score)/10	0.0103***	0.00258***	0.0118***	0.0127***	-0.000432	0.00328*	0.0116***	0.000968**	0.00601***
	[0.000584]	[0.000410]	[0.000696]	[0.00120]	[0.00131]	[0.00168]	[0.000904]	[0.000391]	[0.000700]
(Verbal SAT Score)/10	-0.00310***	0.000714*	0.00622***	-0.00165*	0.00120	0.00800***	-0.00294***	-0.000562	0.00686***
	[0.000578]	[0.000352]	[0.000651]	[0.000862]	[0.00105]	[0.00173]	[0.000850]	[0.000413]	[0.000912]
UC-Adjusted HS GPA	0.144***	0.159***	0.548***	0.117***	0.195***	0.481***	0.0967***	0.152***	0.584***
	[0.0123]	[0.0195]	[0.0223]	[0.0251]	[0.0164]	[0.0249]	[0.0152]	[0.0123]	[0.0165]
Observations	38,676	38,676	38,525	12,114	12,114	11,956	27,206	27,206	26,786
R-squared	0.347	0.375	0.481	0.484	0.489	0.561	0.335	0.345	0.458

Notes: Each column contains coefficient estimates from separate regressions. Each regression also contains controls for parental education, family income by cohort, enrollment campus, and UC application-admission pattern fixed effects that account for the intended major listed on each application. Robust standard errors, clustered by campus-cohort, are reported in brackets. *** p<0.01, ** p<0.05, * p<0.1.

TABLE VI
Determinants of College Outcomes for Intended Science Majors
w/ Different Peer Group/Ability Measures

VARIABLES	Graduated in Sciences	Final Major Science	Graduated	Cumulative GPA
Baseline Peer Group/Ability Definition:				
(Average Math SAT Score of Intended Science Majors)/10	-0.0247***	-0.0200***	-0.0175***	-0.0135
	[0.00390]	[0.00529]	[0.00356]	[0.00794]
(Average Verbal SAT Score of Intended Non-science Majors)/10	0.0279***	0.0273***	0.00837**	-0.0119*
	[0.00453]	[0.00542]	[0.00344]	[0.00595]
Alternate Peer Group Definition:				
(Average Math SAT Score of All Matriculants)/10	-0.0283**	-0.0218*	-0.0265***	-0.0139
	[0.0107]	[0.0125]	[0.00871]	[0.0130]
(Average Verbal SAT Score of All Matriculants)/10	0.0275***	0.0279***	0.0105*	-0.0150
	[0.00852]	[0.00976]	[0.00558]	[0.00876]
Alternate Peer Ability Definition (I):				
(Average Combined SAT Score of Intended Science Majors)/10	-0.0218***	-0.0158***	-0.0141***	-0.0110**
	[0.00359]	[0.00466]	[0.00219]	[0.00513]
(Average Combined SAT Score of Intended Non-science Majors)/10	0.0298***	0.0258***	0.0119***	-0.00233
	[0.00487]	[0.00594]	[0.00258]	[0.00547]
Alternate Peer Ability Definition (II):				
(Average UC-Adjusted HSGPA of Intended Science Majors)/0.05	-0.0298***	-0.0257***	-0.0147***	-0.0125
	[0.00680]	[0.00848]	[0.00299]	[0.00890]
(Average UC-Adjusted HSGPA of Intended Non-science Majors)/0.05	0.0312***	0.0241***	0.0173***	0.0122
	[0.00638]	[0.00760]	[0.00461]	[0.0106]

Notes: Each of the peer group/ability definition groupings and each column contains coefficient estimates from separate regressions. Each regression contains controls for race/ethnicity, SAT I math and verbal scores, UC-adjusted high school GPA, parental education, family income by cohort, enrollment campus, and UC application-admission pattern fixed effects that account for the intended major listed on each application. Robust standard errors, clustered by campus-cohort, are reported in brackets. *** p<0.01, ** p<0.05, * p<0.1.

IX. APPENDIX

IX.A. Imputation of SAT I Scores and High School GPA

Instead of reporting the exact high school GPA and SAT I scores for students the UCOP contains an academic index, which is a weighted linear combination of SAT I math (*m*) and verbal (*v*) scores and high school GPA (*g*):

$$Index_i = c + w_m x^*_{m,i} + w_v x^*_{v,i} + w_g x^*_{g,i} \qquad \text{(vii)}$$

Each of the weights (*w*) and the constant (*c*) in this equation are known, while the *x** terms represent the unobserved true values of a student's SAT scores and high school GPA.

The UCOP data also reports categorical ranges for each student's scores and GPA, with these ranges having an upper ($\overline{x}$) and lower ($\underline{x}$) bound:

$$\overline{x}^0_{j,i} \geq x^*_{j,i} \geq \underline{x}^0_{j,i} \text{ for each } j \in [m, v, g] \qquad \text{(viii)}$$

where the zero superscript represents that these bounds are those that are initially reported in the data.

We can rearrange the terms in equation (ix) to express the unknown math score as a function of known and unknown inputs:

$$x^*_{m,i} = \left[\frac{1}{w_m}\right] Index_i - \left[\frac{c}{w_m}\right] - \left[\frac{w_v}{w_m}\right] x^*_{v,i} - \left[\frac{w_g}{w_m}\right] x^*_{g,i} \qquad \text{(x)}$$

Similarly, we can write this expression for verbal score and high school GPA. Given the initial upper and lower bounds reported in the data and equation (xi), we first attempt to tighten the bounds around the true values of each unobserved variable by iteratively running the following algorithm:

$$\overline{x}^{r}_{m,i} = \min\left(\overline{x}^{r-1}_{m,i}, \left[\frac{1}{w_m}\right] Index_i - \left[\frac{c}{w_m}\right] - \left[\frac{w_v}{w_m}\right] \underline{x}^{r-1}_{v,i} - \left[\frac{w_g}{w_m}\right] \underline{x}^{r-1}_{g,i} \right) \quad \text{(xii)}$$

$$\overline{x}^{r}_{v,i} = \min\left(\overline{x}^{r-1}_{v,i}, \left[\frac{1}{w_v}\right] Index_i - \left[\frac{c}{w_v}\right] - \left[\frac{w_m}{w_v}\right] \underline{x}^{r-1}_{m,i} - \left[\frac{w_g}{w_v}\right] \underline{x}^{r-1}_{g,i} \right)$$

$$\overline{x}^{r}_{g,i} = \min\left(\overline{x}^{r-1}_{g,i}, \left[\frac{1}{w_g}\right] Index_i - \left[\frac{c}{w_g}\right] - \left[\frac{w_v}{w_g}\right] \underline{x}^{r-1}_{v,i} - \left[\frac{w_m}{w_g}\right] \underline{x}^{r-1}_{m,i} \right)$$

$$\underline{x}^{r}_{m,i} = \max\left(\underline{x}^{r-1}_{m,i}, \left[\frac{1}{w_m}\right] Index_i - \left[\frac{c}{w_m}\right] - \left[\frac{w_v}{w_m}\right] \overline{x}^{r-1}_{v,i} - \left[\frac{w_g}{w_m}\right] \overline{x}^{r-1}_{g,i} \right)$$

$$\underline{x}^{r}_{v,i} = \max\left(\underline{x}^{r-1}_{v,i}, \left[\frac{1}{w_v}\right] Index_i - \left[\frac{c}{w_v}\right] - \left[\frac{w_m}{w_v}\right] \overline{x}^{r-1}_{m,i} - \left[\frac{w_g}{w_v}\right] \overline{x}^{r-1}_{g,i} \right)$$

$$\underline{x}^{r}_{g,i} = \max\left(\underline{x}^{r-1}_{g,i}, \left[\frac{1}{w_g}\right] Index_i - \left[\frac{c}{w_g}\right] - \left[\frac{w_v}{w_g}\right] \overline{x}^{r-1}_{v,i} - \left[\frac{w_m}{w_g}\right] \overline{x}^{r-1}_{m,i} \right)$$

until all bounds converge (e.g. $\overline{x}^{R}_{j,i} = \overline{x}^{R-1}_{j,i}$ and $\underline{x}^{R}_{j,i} = \underline{x}^{R-1}_{j,i}$ for each $j \in [m, v, g]$). In running this algorithm, we additionally take advantage of the discrete nature of SAT scores and high school GPA to further tighten the revised upper and lower bounds implied by the data for each of our unobserved measures.

One can envision constructing imputed measures as weighted averages of the revised upper and lower bounds:

$$\hat{x}^{*}_{j,i} = a_j \overline{x}^{R}_{j,i} + (1 - a_j) \underline{x}^{R}_{j,i} \quad \text{(xiii)}$$

The question then becomes, "what is the appropriate weight (a)?" If we plug equation (xiv) into equation (xv) for each unobserved variable we get:

$$Index_i = c + u_m \overline{x}^R_{m,i} + l_m \underline{x}^R_{m,i} + u_v \overline{x}^R_{v,i} + l_v \underline{x}^R_{v,i} + u_g \overline{x}^R_{g,i} + l_g \underline{x}^R_{g,i} \qquad \text{(xvi)}$$

where $u_j = w_j a_j$ and $l_j = w_j(1 - a_j)$ for each $j \in [m, v, g]$. It also follows that:

$$a_j = \frac{u_j}{u_j + l_j} \qquad \text{(xvii)}$$

We can estimate equation (xviii) using regression analysis and construct the implied weights based on the corresponding coefficient estimates. Specifically, we estimate the following regression model for each vigintile (q) of the academic index distribution:

$$Index^q_i = \beta^q_0 + \beta^q_{\overline{m}} \overline{x}^R_{m,i} + \beta^q_{\underline{m}} \underline{x}^R_{m,i} + \beta^q_{\overline{v}} \overline{x}^R_{v,i} + \beta^q_{\underline{v}} \underline{x}^R_{v,i} + \beta^q_{\overline{g}} \overline{x}^R_{g,i} + \beta^q_{\underline{g}} \underline{x}^R_{g,i} + \varepsilon_i \qquad \text{(xix)}$$

The correspondence between this regression model and equation (xx) implies that appropriate weights should be constructed such that:

$$a^q_j = \frac{\hat{\beta}^q_{\overline{j}}}{\hat{\beta}^q_{\overline{j}} + \hat{\beta}^q_{\underline{j}}}, \text{ for each } j \in [\underline{m}, \overline{m}, \underline{v}, \overline{v}, \underline{g}, \overline{g}] \qquad \text{(xxi)}$$

Using these weights and the revised upper and lower bounds for each unobserved variable, we then construct imputed values for each student's unobserved high school GPA, math SAT score, and verbal SAT score.