Blueprint Labs

Discussion Paper #2018.03

The Impact of a STEM Focused Summer Program on College and Major Choices Among Underserved High-Achievers

Silvia Robles

August 2018

MIT Department of Economics
77 Massachusetts Avenue, Bldg. E53-390
Cambridge, MA 02139

National Bureau of Economic Research
1050 Massachusetts Avenue, 3rd Floor
Cambridge, MA 02138

The Impact of a STEM-Focused Summer Program on College and Major Choices Among Underserved High-Achievers

Silvia Robles

Blueprint Labs Discussion Paper #2018.03
August 2018

ABSTRACT

This study presents evidence that a STEM-focused summer program for high-achieving, underserved high school students that is held annually at a selective, private university increases application and enrollment rates at selective universities and persistence in STEM. The study uses records from the program admission process to reduce selection bias by focusing on applicants who advanced to the penultimate stage of admissions and controlling for observables using OLS and propensity score techniques. Results show the program triples the rate of enrollment at the host institution. Students are shifting into the host institution from less selective universities on average with no detectable difference in graduation rates, allaying fears of college mismatch.

* Robles: Gerald R. Ford School of Public Policy, University of Michigan, 725 State Street, Ann Arbor, MI 48104 (email: srobles@umich.edu).

I. Introduction

Low-income students and underrepresented minorities (URM)[1] are less likely to enroll in selective universities than more affluent students or students of other ethnicities, even conditioned on standardized test scores and other markers of academic achievement (see Hoxby and Avery, 2013). These students are also less likely to major in a science, technology, engineering or math (STEM) field, and after starting a STEM major, less likely to graduate with a STEM degree. This is despite being just as likely to report interest in STEM at college entry (Anderson and Kim, 2006).

Underrepresentation in selective universities and STEM has strong implications for future earnings. Mounting evidence suggests that a STEM premium in earnings exists relative to most other majors (Altonji, Blom and Meghir, 2012; Altonji, Arcidiacono and Maurel, 2015; Arcidiacono, 2004), that STEM majors are causally associated with higher earnings (Hastings, Neilson and Zimmerman, 2013; Kirkebøen, Leuven and Mogstad, 2015), and that the STEM-premium has increased over time (Gemini and Wiswall, 2014). Meanwhile, Chetty et al. (2017) shows that an economically disadvantaged student is just as likely to rise to the top quintile of the income distribution by age 30 as an affluent student who attended the same college.

This study measures the effect of a STEM-focused summer program for high school students to see if it increases application and matriculation rates at top universities, STEM major rates, and persistence. The summer program is held annually at a selective private university that awards a majority of its degrees in STEM fields.[2] The mission of the program is to increase access to STEM careers among students who are traditionally underserved (such as first-generation college-goers, low-income, or URM students). It brings academically high-achieving high school

[1] Black, Hispanic, and Native American students are defined as underrepresented minorities (URM), while students of Asian ancestry, though a minority of the population of the US, are not considered underrepresented at selective colleges. Non-Hispanic White students and students in all other ethnic categories are also excluded from the URM designation.

[2] The host institution will hereafter be referred to with the abbreviation HI.

students from across the United States to the HI campus for six-weeks in the summer and simulates the first year of undergraduate education at the HI.[3]

Although it is not obvious how to encourage high-achieving students from underserved backgrounds to apply to selective schools or study STEM, a summer program before senior year of high school may be a natural policy choice. There is evidence that slack in application rates to elite schools (as well as financial aid) can be remedied with interventions right before college application season, particularly those that alleviate frictions such as gathering accurate information about net costs of college or making it easier to send standardized test score reports (Hoxby and Turner, 2013; Carrell and Sacerdote, 2013; Pallais, 2009; Bettinger et al., 2012). Furthermore, the program acts in a college counseling capacity. While the literature suggests counseling improves application and matriculation outcomes (Avery and Kane, 2004; Avery, 2010), the impact of counseling on STEM major rates is still an open question.

In addition to being a specific example of a university-led intervention, the summer program offers an opportunity to examine the role of universities in addressing disparities in access to selective institutions and persistence in STEM. In particular, it is still unclear whether and to what extent university policies matter for graduating with a STEM degree. Some studies show that access to selective universities may hurt certain students. For example, Griffith (2010) finds that high research expenditures relative to educational expenditures is associated with lower persistence in STEM at selective schools. A recent paper by Arcidiacono, Aucejo and Hotz (2016) suggests that URM students with low test scores and grades may be more likely to persist in STEM if they enroll at less selective schools in the University of California system. These papers fit a mismatch narrative, where attending a selective university is detrimental to persistence in STEM for less prepared students.

[3]The freshman year curriculum at the HI is generally composed of three core courses in physics, calculus, and either biology or chemistry, along with one course in the humanities, arts, and social sciences.

Other research has found positive effects of selectivity on completion rates or earnings (for example Dale and Krueger, 2002, 2014; Hoekstra, 2009; Cohodes and Goodman, 2014, among others). This study is well-positioned to examine whether shifting a student on the margin between attending a less selective or highly selective university hurts her persistence in STEM, as the data point to large shifts in college attendance patterns away from less selective universities and towards the HI due to the summer program.

The data for this study come from summer program applications between 2005 and 2011, the National Student Clearinghouse (NSC), and 31 selective, private universities (see Table A1 for a list of participating schools). The latter are particularly informative on the effect of the program on application behavior and admissions, which is usually unobservable.

The study relies on detailed internal knowledge about the program selection process which is usually unavailable, including a rich set of covariates covering most of the information considered by the selection committee, and data on the round of selection where rejected students were cut. By comparing students offered admission to the program to those denied admission conditional on progressing past several screenings, the study reduces the variation in student characteristics and the self-selection problem. Effects are estimated using ordinary least-squares (OLS) and propensity score methods for applicants who advanced through the first rounds of admission to the summer program.

Estimates show the program causes students to apply to the HI 70% more often, and increases admission rates at the HI by 151% relative to comparison students. The admissions effects are large enough that they cannot be explained by the increased application rates alone (at least some applicants must also have higher admission rates conditional on applying). The estimates also show that the program increases the probability of attending the HI by about 31 percentage points relative to a base of 14% (an increase of 218%). There are no detectable increases in enrollment at any other university. The increase in enrollment at the

HI does not come primarily from shifting students between other elite institutions. Thus, on average students matriculate at more selective schools relative to the comparison. Similar to the variance in matriculation behavior found in Hoxby and Avery (2013), about a quarter of students in the comparison group attend the bottom three least selective categories in Barron's Profiles of American Colleges or do not attend any college.

Summer program students graduate from college with STEM degrees significantly more often than comparison students. Estimates show an increase in the likelihood of declaring a STEM major of 5.2 to 9.1 percentage points, and an 8.5 percentage point increase in the probability of graduating with a STEM degree within four years (relative to comparable students who applied to the summer program but were not admitted). Finally, admitted summer program applicants do just as well, if not better, in terms of persistence and overall graduation rates compared to similar students. This is not in line with mismatch theory and contrasts with recent findings in Arcidiacono, Aucejo and Hotz (2016).

This study contributes to the literature evaluating college-led initiatives to increase access to selective universities and STEM majors among URM, low-income, and first-generation college-goers, particularly summer outreach programs. To my knowledge, few summer programs of this nature have been rigorously studied (see for example Price, 2005; Becker, Rouse and Chen, 2016, for evaluations of the American Economic Association summer programs). Recent work by Andrews, Imberman and Lovenheim (2016) finds large positive effects on enrollment, graduation rates and earnings from a program that recruits high-achieving, low-income students to flagship public universities in Texas and provides financial aid and support for students who enroll, but it is not focused on STEM and does not produce any detectable results on STEM major rates.

II. Background and Data

The summer program is offered by an office at the HI dedicated to outreach, and held on campus every year. It immerses rising high school seniors in rigorous science and engineering classes for six weeks during the summer. Students take a course in math, physics, life-sciences, and humanities, as well as an elective course in hands-on, project-based classes such as digital design and genomics. In addition to providing academic preparation, the program exposes students to workshops with leaders of industry and academia and advises students on the college admissions process.

The program covers all costs to the student except for transportation to and from the HI. Students are recruited nationally using mailings and high school visits, as well as word of mouth through the large network of program alumni and online forums.

Selection of program participants occurs based on academic ability and interest in STEM, as indicated through grades, test scores, letters of recommendation, and application essays. In addition, due to the mission of the program to help traditionally underrepresented populations, the following risk factors are considered on a holistic basis during the selection process, though no element in isolation guarantees admission:

1) The individual would be the first in the family to attend college;

2) There is an absence in the individual's family of science and engineering backgrounds;

3) The individual's high school has historically sent less than 50% of its graduates to four-year colleges;

4) The applicant attends a school that presents challenges for success at an urban elite university (e.g., rural or predominantly minority);

5) The individual is a member of a group that is under-represented in the study

and fields of science and engineering (African American, Latino or Native American).

The summer program emphasizes STEM, and targets an underserved population at low cost to the student.[4] This makes it relatively unique in the space of summer outreach interventions.

At the end of the program, participants are given written evaluations from their instructors and teaching assistants. Students are advised to submit these evaluations with all of their college applications. The evaluations produce a signal of how a student might perform as a freshman in college; however, the signal is most relevant to the HI since the program simulates the the HI freshman curriculum so closely.

A. Data

The study uses data from program applications, the NSC, 31 private colleges, the Integrated Postsecondary Education Data System (IPEDS), high school surveys from the National Center for Education Statistics (NCES), and Barron's Profiles of American Colleges. The outreach office has digitized program applications since 2003.[5] The application provides detailed identifying information, demographic characteristics including race, gender, and household size; parent characteristics such as occupation, ethnicity, place of birth, and educational history; and household income measures in the form of an indicator for income under $50K and for eligibility for the federal free or reduced price lunch program. In addition, the application mimics the college application process, collecting high school transcripts, standardized test scores, and a history of extracurricular activ-

[4]Not surprisingly the number of applicants for the program has been rising steadily every year, and acceptance rates in 2010 were lower than acceptance rates at the HI. In response, the outreach office instituted two new interventions to serve more students with the same aim of increasing STEM participation: a one-week residential program, and a six-month online forum. The effects of these additional initiatives is outside the scope of this paper and is addressed in an experimental evaluation that is currently underway and detailed in Robles (2016).

[5]Paper applications for rejected candidates from previous years were not archived.

ities and awards.[6] High schools from the application are matched to high school characteristics in the NCES for the sample of students who advanced to the last stage of admission.

Descriptive statistics for students who applied to program between 2005 and 2011 are presented in Table 1 in column (1). Almost 70% of applicants identify as Black or Hispanic, and 44% report income under $50,000. The average PSAT scores are about 64 for math and 60 for critical reading. These are high relative to the U.S. average of 49 for math and 48 for critical reading. In general, program applicants are likely to be doing well in school, underrepresented minorities of modest or moderate means, and have a declared interest in STEM fields. See Section III for further discussion of selection into the summer program.

The NSC tracks college enrollment and degrees earned from participating institutions for educational reporting and verification purposes. Students are matched to the NSC database using the student's name and date of birth. Institutions who are not in the NSC in earlier years tend to be two-year schools or less selective schools. However, there are some notable exceptions including the HI, which did not join the NSC until 2008. To remedy this gap, NSC data is supplemented with data from the HI and 30 other private, selective universities.[7] These 31 universities, referred to hereafter as peer schools, collect data on application, admissions, enrollment, financial aid, and student performance.

Information on college rankings comes from Barron's Profiles of American Colleges. Schools that are not ranked by Barron's are put into the lowest Barron's category. These are usually vocational schools. Students who do not match to the NSC or peer schools are also added to the lowest category. In order to get a continuous measure of school ranking that does not distinguish between research universities, regional universities, liberal arts colleges, or other types of institutions, the analysis uses revealed preference rankings from Avery et al. (2013) to

[6] Essays and letters of recommendation have only been kept for admitted applicants. See Section IV and Online Appendix A.A2 for a discussion of omitted variable bias.

[7] See Online Appendix Table A1 for a list of peer schools, defined as institutions that participated in the Consortium on Financing Higher Education (COFHE) during the study years.

establish an ordering among the top 110 universities. Schools unranked by Avery et al. (2013) are given a ranking of 111, and students who do not attend any university in the data are given a ranking of 112. Information on whether a school is public, private, four-year, or two-year is contained in the NSC, and information on whether the university is a technical school is hand-coded based on whether over 50% of graduates are in a STEM field according to the IPEDS or whether it is listed as a technical school on Wikipedia.

Analysis includes the 2005 through 2011 cohorts of program applicants.[8] The cohort year refers to the summer of program participation. Figure 1 shows an example timeline given a typical on-time progression. Figure 2 presents the distribution of the number of applications for cohorts between 2005 and 2011.[9] While application rates fluctuate over time and are generally increasing, the number of program slots and applicants who advance to the final round of selection remains roughly constant. Match rates between applicants and the union of NSC and peer school data are presented in Table 2, Panel A. On average, over 95% of students who make it to the final round of selection into the summer program match to either the NSC or a peer school.

III. Selection into the Summer Program

Selection into the program has historically been done in three steps. The first step is a minimum screening on academic credentials where only a small number of applicants are declined. In the second step, the outreach office subcontracts the process to consultants who reduce the pool by over 50%. The consultants recommend candidates for consideration to a selection committee. In the final step, the selection committee, composed of stakeholders and leaders who have a longstanding relationship with the outreach office, make final recommendations

[8]The 2003 and 2004 cohorts of applications are dropped because of the sparse nature of NSC and peer school data in those years. The 2012 and 2013 cohorts are also dropped due to the lag in reporting for both the NSC and peer schools.

[9]Some of the variation in total application rates in 2007 and 2008 appears to be from a data entry policy whereby applications were recorded with more fidelity if they advanced in the screening process. This should not affect the integrity of the analysis for students who advanced past the second screening.

for admission to the program.

Table 1 describes who applies to the summer program and who is accepted along academic, demographic, and socioeconomic measures, as well as high school characteristics. It shows applicants are positively selected relative to the general population along their PSAT scores. The applicant pool also has a large fraction of students with family income under \$50,000 or who qualify for free or reduced price lunch. About 30% of applicants have no college educated parent, and 70% of applicants identify as URM. This makes sense since rational and forward-looking applicants should internalize the selection criteria and choose to apply more often if they have good grades and test scores, an interest in STEM, and feel they fit the mission of the program.

Columns (2) and (3) detail differences between accepted candidates and rejected candidates. The summer program admissions process is, by design, supposed to select candidates. Therefore the large and significant differences along observable characteristics between columns (2) and (3) are unsurprising. Admitted candidates are more likely to report taking math and science classes, belong to URM groups, have family income under \$50,000, have no college-educated parent in the household, and have higher test scores. In addition, the program has historically tried to maintain gender parity in its class composition, while the share of female applicants is 39%.

The analysis includes candidates who made it past the first two screenings, and onto the desks of the selection committee. Rejected applicants who made it to the final round are hereafter referred to as "post second screening"; column (5) of Table 1 reports descriptive statistics for this group. Column (6) illustrates how, along most measures of observable characteristics, the difference between admitted and rejected candidates post second screening is smaller than the difference between admitted students and all rejected candidates.[10] However, these

[10]The exceptions are parental education, the fraction of students who are White, and the share of minorities at a student's high school. For these three measures, the difference between admitted candidates and students rejected after the second screening gets larger relative to the difference between admitted

differences are still jointly significant. The final column of Table 1 further adjusts for selection by controlling for the estimated propensity score. Along individual measures, as well as jointly, adjusting for the propensity score shrinks the differences between admitted and rejected applicants to levels that are not statistically significant.

IV. Methods

Causal interpretation of the estimates in this study rely on satisfying the conditional independence assumption. Conditional on observed characteristics, each student's potential outcomes must be independent of admission into the summer program. In addition, controlling for variables related to the probability of receiving treatment satisfies conditional independence (Rosenbaum and Rubin, 1983). These are exactly the variables in the application to the program.

The main threat to a causal interpretation of the results is the omitted variable bias from application variables that are hard to code or unobserved, specifically letters of recommendations and application essays. This concern is addressed in two ways. First, by restricting all analysis to the sample of students who survived past the second screening, I compare students who must be more similar to each other on all variables that matter for being admitted to the summer program, including unobserved characteristics such as essays and letters of recommendation. Intuitively, the selection problem is reduced by limiting analysis to students who were more likely candidates ex ante.

In addition, Section A.A2 presents bounds on the estimates and calculates how much variation would have to be explained by unobservable characteristics such that one could no longer reject the null hypothesis (Oster, 2013; Altonji, Elder and Taber, 2005).

candidates and all rejected applicants.

Estimation

The primary estimates of the effect of the summer program come from an OLS regression of the form

$$Y_i = \sum_j \alpha_j + \beta Treat_i + X_i\gamma + \varepsilon_i$$

where Y_i is the outcome, $Treat_i$ is an indicator for program admission, and X_i is a vector of student level covariates from the program application. The α_j are cohort fixed effects, and ε_i is an error term. The full list of covariates is listed in Table A3 and includes demographic characteristics, parent characteristics, household income measures, high school transcripts, standardized test scores, and a history of extracurricular activities and awards. Using admission to the summer program as the dependent variable is effectively calculating an intent to treat estimate and abstracts from students deciding to accept offers of admission. However, over this time period few students reject offers of admission (two or three every year) and treatment on the treated estimates should be similar.

Propensity Score Methods

Comparing students using the propensity score is a natural alternative in contexts where researchers know more about selection into the program. In this study, the application process is well known, to the point that the analysis can focus on the applicant pool who advanced to the penultimate step in selection, and there is a rich set of covariates that correspond directly to how students were evaluated by the selection committee (because they are drawn from the application). These are favorable conditions where the propensity score and OLS results are likely to adhere closely to an experimental setting (Diaz and Handa, 2006; Cook, Shadish and Wong, 2008).

Calculating the propensity score makes it easy to visually inspect common support issues. Figure 3 shows a histogram of the propensity score for admitted applicants as well as the mirror image histogram of the propensity score for rejected candidates when estimating the model using all program applicants. The majority of applicants have an exceedingly low propensity score and the overlap near the top of the distribution is small. However, Figure 4 shows the corresponding histograms for only the post second screening applicants. There is more overlap in the propensity score distributions, which may ease concerns over whether credible comparisons can be drawn from this group.

Online Appendix A.A3 discusses the propensity score estimation in detail and compares OLS estimates to results from propensity score nearest neighbor matching (with one and three neighbors), as well as inverse propensity score weighting. In general, results closely follow OLS estimates, therefore the core results are reported for OLS.

V. Results

This section details estimates of the effect of the summer program on applications to college, admissions, enrollment, probability of majoring in a STEM field, persistence in college, graduation rates and the probability of graduating with a STEM degree. Results are reported for OLS regression, though Table A4 in the online appendix contrasts estimates using OLS regression with propensity score nearest neighbor matching (with one and three nearest neighbors) and inverse propensity score weighting. By and large, point estimates and standard errors stay within a narrow range of each other between estimation methods that account for covariates. For example, the effect on enrollment in any college fluctuates between 1.5 and 3.1 percentage points and is not significant regardless of the estimation method. The absence of marked differences between the OLS and propensity score estimates is consistent with good covariate overlap. Results are presented for all post second screening applicants.[11]

[11]Results by demographic subgroups are available upon request. There are no detectable subgroup differences.

A. Enrollment

Figure 5 presents a histogram of the unadjusted difference in density of students enrolled in each Barron's category ranking, by program admission status. The top Barron's category, Most Competitive, is broad. The HI, other top 10 schools, and the top 11 through 20 schools (using the Hoxby and Avery, 2013, rankings) are broken out to show increased detail among elite institutions. There are notable masses of summer program admits at the HI and in the other top 10. In contrast, although comparison students are also concentrated in the top 10 schools, there are more comparison students found all along the ranking distribution. The large share of comparison students in category five or no college is almost as large as the share in the top 10.

These are unadjusted differences that could be entirely driven by selection bias. Table 3 presents the OLS results on college enrollment using the union of NSC and peer school data. The columns present estimates of the impact of the summer program on enrollment within one, two, three, or four years of the senior year of high school. Because the data are right-censored, each successive column varies in the cohorts for whom data is available.

The first row shows a positive, nonsignificant increase in college enrollment for summer program admits of 2.6 percentage points on a base of about 91 percent in the first year. The effect of the summer program on enrollment in four-year colleges, in the second row, is of similar size as the effect on enrollment in any college since few students choose to attend a two-year.

The third row of Table 3 shows the program causes an estimated 14.5 percentage point increase in enrollment at private, four-year colleges within one year of high school. The magnitude of the coefficients are larger two, three, and four years after high school, suggestive of higher persistence among summer program students (these differences are not significant). The shift into four-year, private schools is almost perfectly accounted for by a significant decrease in enrollment at public four-year colleges of about 12.2 percentage points one year out, as well as nonsignificant decreases at two-year colleges[12] and the share of students who do not enroll anywhere.

Table 4 shows the main beneficiary of shifting enrollment patterns is the HI. Enrollment at the HI increases by an estimated 30.8 percentage points in the first year after high school, which is roughly stable two, three, and four years after high school. Since more summer program students enrolled to begin with, identical percentage point decreases for admitted and rejected students between year one and year two, for example, would translate to a higher fraction of summer program students enrolled at the HI in year two. Therefore, identical coefficients across each column are consistent with greater persistence for summer program admits.

Table 4 also shows enrollment by different measures of school selectivity and for other schools. The summer program reduces the chance of attending another tech school by about four percentage points one year out of high school, with estimates increasing over time. The fall in other technical schools is much smaller than the rise in HI enrollment, resulting in a net gain in tech school enrollment.

With regards to selectivity, changes in enrollment due to the summer program exhibit a U-shaped pattern, with about half of the difference arising relative to comparison students enrolled in a selective school and half from students enrolled in a non-selective school or no school. Although a non-negligible fraction of the shift towards the HI comes from the most competitive Barron's category, the category is broad and encompasses approximately the top 40 schools in the Avery et al. (2013) rankings. Rows three, four, and five of Table 4 show that the summer program is not detectably drawing students from other top 10 institutions, is causing a 5.8 percentage point reduction in enrollment at schools in the top 11 through 20, and a 5.8 percentage point reduction in enrollment at other Most Competitive schools outside of the top 20. On average, the summer program is causing an increase in the matriculating school's selectivity, especially for students

[12] Results on two-year schools available upon request.

who would have gone to schools in the bottom four Barron's categories, or not gone to a school in the NSC.

B. *Enrollment, Application and Admission Rates at Peer Schools*

A natural question is whether the increase in enrollment selectivity comes from a shift in where students apply, where they are admitted, the student's decision to enroll conditional on an offer of admission, or some combination of all three. Table 5 gives a more detailed picture for peer schools where both applications and admissions are observed.[13] Peer schools are selective, private schools who share a subset of their administrative data with each other, and are listed in Online Appendix Table A1.

Table 5, Panel A presents results on whether a student applies to a peer school within two years of high school. Program admits are, on net, 3.7 percentage points more likely to apply to any peer school; this increase represents new applicants who would not have otherwise applied to any peer school. All of the positive effects are concentrated at the HI (31.8 percentage point increase). There is some evidence of application crowd out. Columns (4) and (5) of Panel A report a significant nine percentage point decrease in the application rate at other peer schools outside the Ivy League, and a significant, four percentage point decrease in application rates to liberal arts peer schools. Ivy League application rates are not significantly higher, with a three percentage point positive coefficient.

In Panel B, program admits are 12.8 percentage points more likely to be admitted to any peer school within two years of high school. Once again, the results are concentrated at the HI. Students admitted to the summer program are 33.8 percentage points more likely to be admitted to the HI. For Ivy Leagues, Non-Ivys, and Liberal Arts schools, the change in admissions is similar or smaller in magnitude than the change in applications, which means that admissions conditional on applying are not falling. For example, if applications dropped by nine

[13] The categories presented are the only way that peer schools permit data from this database to be stratified. the HI has given special permission to be identified separately.

percentage points as in column (4), and admissions rates conditional on applying also fell, the unconditional admissions drop would be larger than nine percentage points, not smaller. Only the 3.2 percentage point drop in overall admission to Liberal Arts colleges is significant.

The comparison group's admission rate conditional on applying to the HI is about 49%. If summer program students who applied to the HI were admitted 49% of the time, that would only account for about half of the observed increase in the unconditional admissions rate (see Panel A, column (3) of Table 5). Thus at least some of the applicants to the HI must have also had a higher probability of being admitted due to the summer program.

Finally, Panel C presents estimates of the impact of program admission on enrollment at peer schools. The results reinforce the pattern observed in the NSC data. Enrollment gains are observed at the HI, with small or negligible reductions at other peer schools.[14] The largest enrollment drops are at colleges that are neither ivy-league universities nor liberal arts colleges. While these are selective schools, they tend to be less selective on average relative to the other peer schools.

C. STEM Majors

One of the reported goals of the summer program is to encourage the study of STEM fields. Table 6 column (1) shows the estimates of the program effect on the probability of ever declaring a STEM major. Estimates come from applicants to the summer program in 2007 through 2011 matched to information about majors from the NSC.[15]

Missing observations are coded as zeros, comparing the likelihood of declaring a STEM major unconditional on attending college or having a missing observation for major of study. The summer program is associated with a 7.7 percentage

[14]Small differences in estimated effects on HI enrollment between Table 3 and 5 emerge due to slight variation between peer school and NSC data.

[15]The HI joins the NSC in time for summer program applicants in 2007 who enroll one year after high school to be included, making this the most complete set of years.

point increase in the probability of declaring a STEM major. There is likely to be bias present due to missing observations. Majors are not the principal data collected by the NSC and 33% of observations on majors in the NSC are missing for the study sample.[16]

Columns (1) and (3) of Online Appendix Table A5 report the results of predicting the likelihood of missing a major in the NSC using observable characteristics. They show that missing data is not related to program admission. This is true conditional on other observable variables in column (3), and without conditioning on observables in column (1).

By inspection, the majority of students who have missing majors in the NSC attend either the HI or an Ivy League school that prevents the NSC from reporting its records; both of these schools are available in the peer school data. Table 6 column (2) shows what happens to the OLS estimates of the probability of majoring in STEM if gaps in the NSC data are supplemented with data from peer schools using the 2007 to 2011 cohorts (the same sample years as in column 1). Column (3) reports results for the full sample. The estimated impact of the summer program on the likelihood of declaring a major in these specifications is an increase of 10.2 and 9.1 percentage points, respectively.

Finally, Table 6, column (4) examines the sensitivity of the estimates to missing observations. Observations for summer program students with missing majors are assigned the STEM major rate of non-missing control observations. Control students are given the STEM-major rate of non-missing summer program admits. This is a conservative imputation. Estimates remain positive and significant, indicating a lower bound of 5.2 percentage points for the effect of the program on the likelihood of majoring in STEM, relative to an imputed control mean of 69.4%.

[16]See supplemental Table A6.

D. Graduation Rates

The idea of mismatch is that underprepared students who attend a more selective university may be worse off if they becomes overwhelmed by the higher academic standards. Table 7 looks for evidence of mismatch by comparing estimates of the effects on graduation rates at four-year colleges and the probability of graduating with a STEM degree. There are no detectable adverse effects on graduation rates. In the first row, the coefficients on program admission are positive and not significant four, five, and six years after high school, when most people would expect to graduate. The estimates rule out a negative impact on four-year graduation rates larger than about 2.7 percentage points.

Estimates in the first row aggregate graduation across all majors and institutions. The coefficients in the second row for four-year colleges are similar in magnitude to the results for all colleges. Overall, students admitted to the summer program do not graduate from college at a detectably different rate when aggregating all majors, and if anything, may graduate more often. The magnitude of the coefficients is similar to the initial difference in enrollment, which is consistent with higher persistence rates for summer program admits, though the difference is not detectable. However, persistence effects are not large enough to be detectable.

The summer program has a larger impact on graduation with a STEM degree. Students admitted to the summer program are 8.5 percentage points more likely to graduate with a STEM degree within four years using the conservative imputation from Section V.C, a magnitude that is not detectably different for graduates who finish five and six years after high school, although the six year coefficient is not significant. All students who applied to the summer program presumably did so because they were interested in STEM fields, and 40.6% of comparison students are estimated to graduate with a STEM degree within four years. Despite the high share of the comparison group that earns a STEM degree, the program manages to cause an estimated 21% increase in the four-year rate of graduation with a STEM degree.

VI. Discussion

A. Application, Admission, and Enrollment at the HI

Admission to the summer program more than triples the rate at which students attend the HI; this boosts college selectivity for the treatment group on average. The shift into the HI comes from all over the selectivity distribution, but enrollment in other top ten institutions stays constant. Notably, almost ten percent of comparison students either failed to enroll in college, or enrolled in a school that does not report to the NSC, which means that even among high-achievers there is space for improving college-going rates.[17]

Summer program students are increasing their enrollment rates at the HI partially by increasing their application rates. The extent to which this crowds out applications to institutions outside of the peer schools is unknown, but it is unlikely that the average selectivity of the bundle of colleges that students apply to goes down given how highly ranked the HI is. There are also large gains in the admissions rate at the HI for summer program students.

The mechanisms for these effects are less clear. For example, the program could make students more desirable to admissions offices through increasing human capital, or by sending a clearer signal about their suitability for admission. Moreover, students who apply to peer schools in the absence of the program are likely to be different from students who are induced to apply by the program. the intervention need not change the student at all to induce an admissions effect as long as it has already induced an application increase.

All three possibilities would be consistent with undermatching in the sense that a student might not realize her potential to be admitted to the HI in the absence of the program. However, the policy implications are quite different depending on

[17] Compare this to the analysis in Smith, Pender and Howell (2013) which shows that among all SAT test takers, students with characteristics that predict "access" to selective universities, or high-achievers, fail to enroll in any school only three percent of the time.

the mechanism. If the summer program induces students to apply who otherwise would not have applied, and these students are precisely the most desirable candidates to admissions offices, this supports policies that increase application rates among underrepresented students. The outreach studied in Hoxby and Turner (2013) and other targeted recruitment practices are in this vein.

If applicants whom admissions officers encounter with lower frequency, such as low-income or minority candidates, are difficult to evaluate and are accepted less often (as in Autor and Scarborough, 2008), a testable implication is that more precise information in these students' applications would raise admissions rates. The relevant policy would be more involved screening mechanisms.

B. Graduation

Increasing participation and retention in STEM fields is a major goal of the program. Importantly, there is no evidence that students are harmed from matriculating in a higher ranked school; results on graduation rates suggest that summer program students are not more likely to drop out and are more likely to complete a STEM degree. I estimate a positive and significant lower bound for the effect of the program on the likelihood of declaring a STEM major of about 5.2 percentage points.

Once again, the relevant policy depends on the mechanism. If a student is unsure of what to major in, the school's comparative advantage in STEM fields could lead her to choose a STEM field as she explores possible areas of study. In other words, diverting students to tech schools could in itself cause students to earn more STEM degrees. Alternatively, the STEM major effect may be specific to this summer program, perhaps because a large cohort of low-income and minority students attend the HI together afterward, which provides a source of peer group support. It could be necessary to foster a community for underrepresented students, as in the Posse scholarship program, to increase STEM persistence. Finally, the program could directly increase STEM major rates by increasing the level of enthusiasm for STEM. Understanding how and why the institution plays a role in STEM persistence seems like a fruitful area for future work.

The increase in STEM persistence due to the summer program contrasts with the findings in Arcidiacono, Aucejo and Hotz (2016), but it is difficult to draw a direct comparison since participating in a STEM-focused summer program is not the same conceptual exercise as plucking one student out of a less selective school and enrolling her in the HI. For one, the summer program could directly increase STEM persistence rates enough to negate any mismatch-style adverse effects. In this case, summer programs like the one in this study or other supportive interventions could mitigate the impact of being academically vulnerable at a selective university.

Arcidiacono, Aucejo and Hotz (2016) also find that decreases in STEM persistence are concentrated among students with the lowest grades and test scores. If there is some threshold above which a student is sufficiently prepared to take on the challenges of selective schools, these summer program applicants may be above that threshold even before the program. In this case, the need for more research is evident. Thresholds could vary considerably between schools and populations. A better developed screening process, one that considers more precise signals for whether a student will thrive, could be necessary for successfully increasing diversity at selective institutions. The summer program provides such a signal by acting as a low-stakes test run for the HI. Other selective colleges could theoretically offer similar trial periods for candidates that are on the margin for being admitted to their schools.

C. Conclusion

This study uses detailed knowledge of the summer program admissions process to estimate the impact of the program and mitigate any selection bias. While propensity score analysis demonstrates good covariate overlap after focusing on students who advanced to the penultimate stage of summer program admissions, there will always be concern when using a selection on observables strategy that the analysis is missing some important element. However, some aspects of this study have been replicated with early results from an ongoing randomized trial. Experimentally obtained point estimates of the effect of the intervention on application rates to the HI are significant, positive, and broadly similar in magnitude (see Robles, 2016), which is reassuring.

Although the experimental analysis of the summer program faces fewer identification challenges, the benefit of looking at retrospective data is the ability to examine longer-term outcomes which will take many years to collect prospectively, such as graduation rates and STEM degree rates. Overall, this study finds that students who are admitted to the program go to a more selective college, are no more likely to drop out, and are more likely to graduate with a STEM degree than they otherwise would have been. This research shows that among high-achieving, low-income or URM students, there is still space to increase application and enrollment rates at selective colleges, a point made in the literature already (Hoxby and Avery, 2013). However, the documented potential for university policies to positively influence STEM persistence in this population is a new and exciting development. Moreover, it stands as a counterpoint to the evidence on mismatch, suggesting either that universities can support potentially vulnerable students to mitigate mismatch using similar summer programs, or that mismatch is only a problem in certain settings.

Acknowledgments

This work was supported by the National Science Foundation (Award 1260898), The Lab for Economic Applications and Policy at Harvard University, and the Alfred P. Sloan Foundation. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the author and do not reflect the views of any other entity. I thank Larry Katz, Josh Angrist, Amanda Pallais, Sarah Cohodes, Brian Jacob, Aurelie Ouss, Raj Chetty, Sally Hudson, Susan

Dynarski and seminar participants at Harvard, MIT, Iowa State University, and University of Michigan for helpful comments. Excellent research assistance was provided by Helen Ho, Iris Song, and Luis Gil. Finally, I thank everyone at the host institution, and Steve Minicucci, for their support and for granting access to their data.

REFERENCES

Altonji, Joseph G., Erica Blom, and Costas Meghir. 2012. "Heterogeneity in Human Capital Investments: High School Curriculum, College Major, and Careers." *Annual Review of Economics*, 4: 185–223.

Altonji, Joseph G., Peter Arcidiacono, and Arnaud Maurel. 2015. "The Analysis of Field Choice in College and Graduate School: Determinants and Wage Effects." National Bureau of Economic Research Working Paper 21655.

Altonji, Joseph G., Todd E. Elder, and Christopher R. Taber. 2005. "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools." *Journal of Political Economy*, 113: 151–184.

Anderson, Eugene, and Dongbin Kim. 2006. "Increasing the Success of Minority Students in Science and Technology." American Council on Education, Washington, DC.

Andrews, Rodney J., Scott A. Imberman, and Michael F. Lovenheim. 2016. "Recruiting and Supporting Low-Income, High-Achieving Students at a Flagship University." National Bureau of Economic Research Working Paper 22260.

Angrist, Joshua, and Jinyong Hahn. 2004. "When to Control for Covariates? Panel Asymptotic Results for Estimates of Treatment Effects." *The Review of Economics and Statistics*, 86: 58–72.

Angrist, Joshua D., and Jrn-Steffen Pischke. 2009. *Mostly Harmless Econometrics.* Princeton, NJ:Princeton University Press.

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

Arcidiacono, Peter, Esteban M. Aucejo, and V. Joseph Hotz. 2016. "University Differences in the Graduation of Minorities in STEM Fields: Evidence from California." *American Economics Review*, 106(3): 525–562.

Autor, David H., and David Scarborough. 2008. "Does Job Testing Harm Minority Workers? Evidence from Retail Establishments." *Quarterly Journal of Economics*, 123: 219–277.

Avery, Christopher. 2010. "The Effect of College Counseling on High-Achieving, Low-Income Students." National Bureau of Economic Research Working Paper 16359.

Avery, Christopher, and Thomas J. Kane. 2004. "Student Perceptions of College Opportunities: The Boston COACH Program." In *College Choices: The Economics of Where to Go, When to Go, and How to Pay For It.* , ed. Caroline Hoxby, 355–94. Chicago:University of Chicago Press.

Avery, Christopher, Mark Glickman, Caroline Hoxby, and Andrew Metrick. 2013. "A Revealed Preference Ranking of U.S. Colleges and Universities." *Quarterly Journal of Economics*, 1: 425–467.

Becker, Charles M., Cecilia Elena Rouse, and Mingyu Chen. 2016. "Can a Summer Make a Difference? The Impact of the American Economic Association Summer Program on Minority Student Outcomes." *Economics of Education Review*, 53: 46–71.

Bettinger, Eric, Bridget Terry Long, Philip Oreopoulos, and Lisa Sanbonmatsu. 2012. "The Role of Application Assistance and Information in

College Decisions: Results from the H&R Block FAFSA Experiment." *The Quarterly Journal of Economics*, 3(127).

Black, Dan, and Jeffrey Smith. 2004. "How Robust is the Evidence on the Effects of College Quality? Evidence from Matching." *Journal of Econometrics*, 121: 99–124.

Carrell, Scott, and Bruce Sacerdote. 2013. "Why Do College Going Interventions Work?" National Bureau of Economic Research Working Paper 19031.

Chetty, Raj, John N. Friedman, Emmanuel Saez, Nicholas Turner, and Danny Yagan. 2017. "Mobility Report Cards: The Role of Colleges in Intergenerational Mobility." Unpublished manuscript.

Cohodes, Sarah, and Joshua Goodman. 2014. "Merit Aid, College Quality and College Completion: Massachusetts Adams Scholarship as an In-Kind Subsidy." *American Economic Journal: Applied Economics*, 6(4): 251–285.

Cook, Thomas, William R. Shadish, and Vivian C. Wong. 2008. "Three Conditions Under Which Experiments and Observational Studies Produce Comparable Causal Estimates: New Findings from Within-Study Comparisons." *Journal of Policy Analysis and Management*, 27(4): 724–750.

Dale, Stacy, and Alan B. Krueger. 2002. "Estimating the Payoff to Attending a More Selective College: An Application of Selection on Observables and Unobservables." *Quarterly Journal of Economics*, 117(4): 1491–1528.

Dale, Stacy, and Alan B. Krueger. 2014. "Estimating the Effects of College Characteristics over the Career Using Administrative Earnings Data." *Journal of Human Resources*, 49(2): 323–358.

Diaz, Juan Jose, and Sudhanshu Handa. 2006. "An Assessment of Propensity Score Matching as a Nonexperimental Impact Estimator Evidence from Mexico's PROGRESA Program." *The Journal of Human Resources*, 41(2): 319–345.

Gemini, Ahu, and Matthew Wiswall. 2014. “Evolution of Gender Differences in Post-Secondary Human Capital Investments: College Majors.” *International Economic Review*, 55(1): 23–56.

Griffith, Amanda L. 2010. “Persistence of Women and Minorities in STEM Field Majors: Is it the School that Matters?” *Economics of Education Review*, 29(6): 911–22.

Hastings, Justine S., Christopher A. Neilson, and Seth D. Zimmerman. 2013. “Are Some Degrees Worth More Than Others? Evidence from College Admission Cutoffs in Chile.” National Bureau of Economic Research Working Paper 19241.

Hirano, Keisuke, Guido W. Imbens, and Geert Ridder. 2003. “Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score.” *Econometrica*, 71: 1161–1189.

Hoekstra, Mark. 2009. “The Effect of Attending the Flagship State University on Earnings: A Discontinuity-Based Approach.” *The Review of Economics and Statistics*, 91(4): 717–724.

Hoxby, Caroline, and Christopher Avery. 2013. “The Missing “One-Offs”: The Hidden Supply of High-Achieving, Low-Income Students.” *Brookings Papers on Economic Activity*, 1–65.

Hoxby, Caroline, and Sarah Turner. 2013. “Expanding College Opportunities for High-Achieving, Low Income Students.” Stanford Institute for Economic Policy Research Discussion Paper 12-014.

Kirkebøen, Lars J., Edwin Leuven, and Magne Mogstad. 2015. “Field of Study, Earning, and Self-selection.” Statistics Norway Research Department Discussion Paper 794.

Oster, Emily. 2013. “Unobservable Selection and Coefficient Stability: Theory and Evidence.” National Bureau of Economic Research Working Paper 19054.

Pallais, Amanda. 2009. "Taking a Chance on College: Is the Tennessee Education Lottery Scholarship Program a Winner?" *Journal of Human Resources*, 44: 199–222.

Price, Gregory N. 2005. "The Causal Effects of Participation in the American Economic Association Summer Minority Program." *Southern Economic Journal*, 72(1): 78–97.

Robles, Silvia C. 2016. "Essays on Access to Higher Education." PhD diss. Harvard University.

Rosenbaum, Paul R., and Donald B. Rubin. 1983. "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika*, 70: 41–55.

Smith, Jonathan, Matea Pender, and Jessica Howell. 2013. "The Full Extent of Academic Undermatch." *Economics of Education Review*, 32: 247–261.

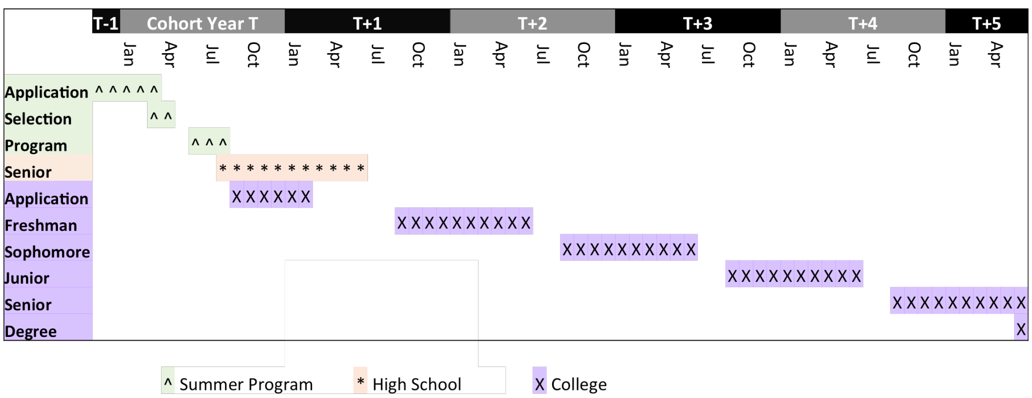

Figure 1. Typical progression of a student in cohort year T. Source: Author's calculations.

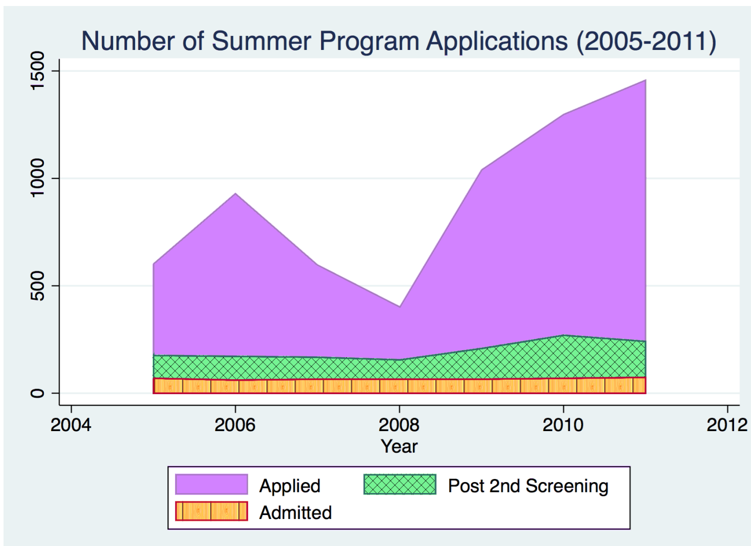

Figure 2. Count of the number of students who applied, made it past the second screen of admission, and were accepted to the summer program program by year. Source: Summer program applications and author's calculations.

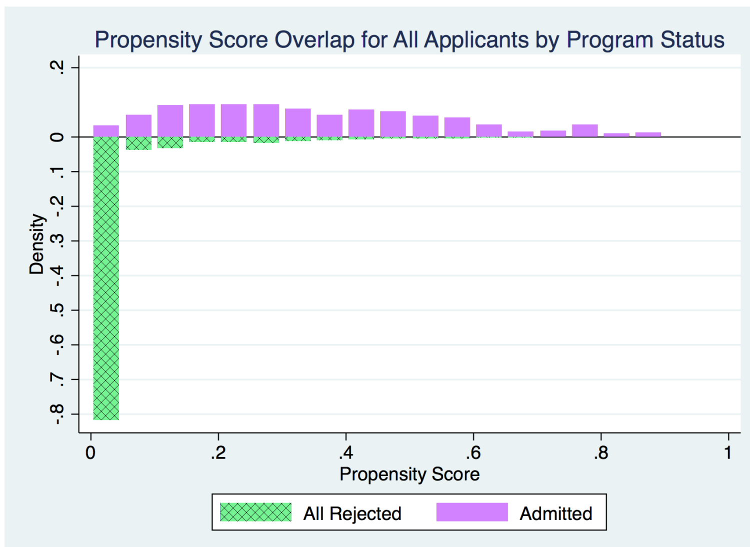

Figure 3. Mirror images of the histograms of the propensity scores estimated using the full sample of summer program applicants between 2005 and 2011. Propensity scores for admitted students are shown in pink while those for rejected applicants are shown in blue below the horizontal axis. Source: Summer program applications and author's calculations.

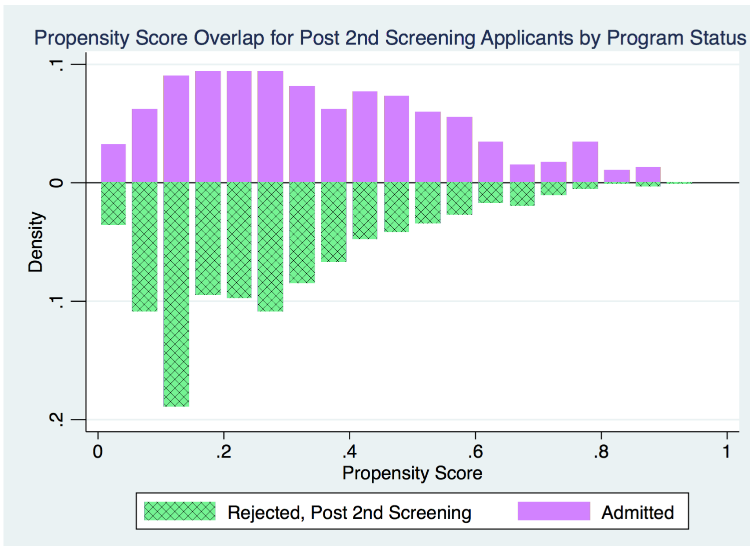

FIGURE 4. MIRROR IMAGES OF THE HISTOGRAMS OF THE PROPENSITY SCORES ESTIMATED USING THE POST-SECOND SCREENING SAMPLE OF SUMMER PROGRAM APPLICANTS BETWEEN 2005 AND 2011. PROPENSITY SCORES FOR ADMITTED STUDENTS ARE SHOWN IN PINK WHILE THOSE FOR REJECTED APPLICANTS WHO MADE IT PAST THE SECOND SCREENING DURING SELECTION ARE SHOWN IN BLUE BELOW THE HORIZONTAL AXIS. SOURCE: SUMMER PROGRAM APPLICATIONS AND AUTHOR'S CALCULATIONS.

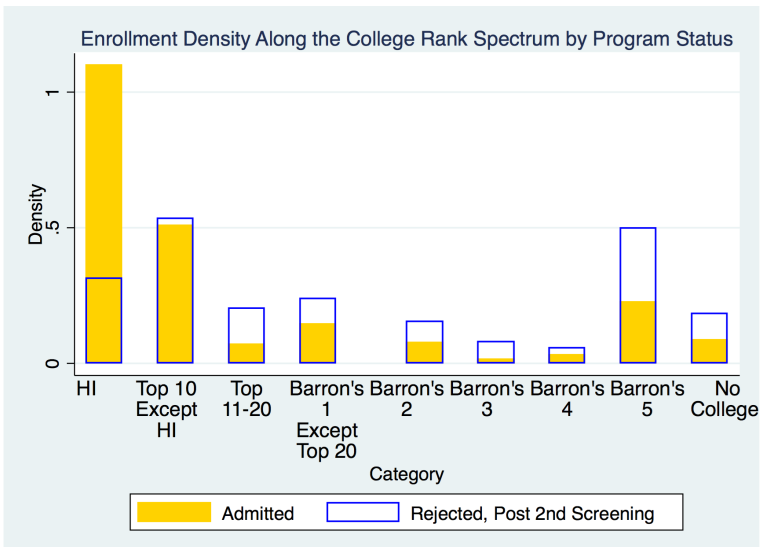

Figure 5. Histogram of the unadjusted difference in the density of summer program admits enrolled at universities by their rank. Enrolled university is defined as the first school where a student is observed during the fall semester as long as it is within two years after senior year of high school. Rankings categories are from Barron's Profiles of American Colleges, and Avery et al. (2013) for the top 20. Source: Summer program applications, NSC, peer schools, and author's calculations.

Table 1—Descriptive Statistics and Covariate Balance

	Program Admits vs. All Rejected Applicants				Program Admits vs. Rejected Applicants, Post 2nd Screening		
Academics	All (1)	Admitted (2)	Rejected (3)	Difference (4)	Rejected Post 2nd Screening (5)	Difference (6)	Propensity-Adjusted Difference (7)
PSAT Math	63.82	66.53	63.58	2.706	65.96	0.522	0.003
				(0.325)		(0.369)	(0.404)
PSAT Verbal	59.78	61.73	59.61	1.948	61.52	0.198	0.001
				(0.374)		(0.434)	(0.468)
Took Calculus	0.18	0.27	0.17	0.097	0.20	0.069	0.000
				(0.021)		(0.024)	(0.025)
# Awards	3.97	4.52	3.92	0.601	4.50	0.020	0.000
				(0.093)		(0.11)	(0.119)
Demographics							
Female	0.39	0.49	0.38	0.107	0.42	0.066	0.000
				(0.024)		(0.028)	(0.031)
Black	0.36	0.40	0.35	0.052	0.44	-0.038	0.000
				(0.024)		(0.028)	(0.03)
Hispanic	0.33	0.43	0.33	0.109	0.39	0.043	0.000
				(0.024)		(0.028)	(0.03)
Asian	0.20	0.09	0.20	-0.119	0.09	-0.002	0.000
				(0.014)		(0.016)	(0.018)
Native American	0.02	0.04	0.01	0.025	0.03	0.014	0.000
				(0.009)		(0.01)	(0.012)
White	0.04	0.03	0.04	-0.014	0.01	0.016	0.000
				(0.008)		(0.008)	(0.008)
Other Ethnicity	0.06	0.01	0.06	-0.052	0.04	-0.033	0.000
				(0.006)		(0.008)	(0.007)
Age at Program	17.07	17.06	17.08	-0.012	17.06	0.009	0.000
				(0.024)		(0.037)	(0.033)
Socioeconomic Status							
Income < $50k	0.36	0.63	0.34	0.144	0.34	0.150	0.001
				(0.018)		(0.02)	(0.02)
Free Lunch	0.31	0.59	0.29	0.147	0.33	0.127	0.001
				(0.018)		(0.02)	(0.02)
Any College Educated Parent	0.73	0.58	0.74	-0.163	0.77	-0.188	-0.001
				(0.024)		(0.027)	(0.026)
High School Characteristics							
HS % Minority		0.57			0.53	0.035	0.000
						(0.016)	(0.016)
Rural HS		0.14			0.12	0.019	0.000
						(0.018)	(0.02)
HS Peer School App. Rate		0.12			0.18	-0.044	0.000
						(0.01)	(0.01)
F statistic				30.109		6.677	0.001
p-value				0.00		0.00	1.00
Observations	6324	467	5857	6324	920	1387	1387

Note: Table shows descriptive statistics for summer program applicants between 2005 and 2011. Column (1) shows covariate means for all applicants. Column (2) shows means for applicants who were admitted to the summer program. Columns (3) and (5) show means for rejected applicants; the latter restricts the sample to those who made it past the second screening in the selection process. Columns (4) and (6) show OLS coefficients on an indicator for program admission. Column (7) shows OLS coefficients on program admission controlling for the estimated propensity score. The F-statistics and p-values are from a Wald test that the coefficients on program admission are jointly different from zero. Standard errors are robust to heteroskedasticity and displayed in parentheses.

Table 2—NSC and Peer School Follow-up Rates

	All (1)	Rejected (2)	Admitted (3)	Rejected, Post 2nd Screening (4)
All Cohorts	0.891	0.885	0.961	0.918
	(0.312)	(0.319)	(0.193)	(0.274)
2005 Cohort	0.869	0.865	0.899	0.907
	(0.338)	(0.342)	(0.304)	(0.292)
2006 Cohort	0.878	0.871	0.984	0.936
	(0.327)	(0.335)	(0.128)	(0.246)
2007 Cohort	0.886	0.874	0.984	0.854
	(0.318)	(0.332)	(0.125)	(0.354)
2008 Cohort	0.923	0.914	0.969	0.956
	(0.267)	(0.281)	(0.174)	(0.206)
2009 Cohort	0.863	0.856	0.970	0.943
	(0.344)	(0.351)	(0.173)	(0.232)
2010 Cohort	0.906	0.902	0.971	0.926
	(0.292)	(0.297)	(0.170)	(0.263)
2011 Cohort	0.907	0.904	0.959	0.904
	(0.291)	(0.295)	(0.199)	(0.295)
Observations	6324	5857	467	920

Note: Observations include applicants to the summer program between 2005 and 2011, matched to the NSC and peer schools. Follow-up is defined as whether the student is observed enrolled anywhere within two years of senior year of high school. Standard deviations are displayed in parentheses.

Table 3—Effect of Summer Program Admission on College Enrollment and Persistence

	Cohorts			
	2005-2011	2005-2011	2005-2010	2005-2009
Enrolled In	Year 1	Year 2	Year 3	Year 4
	(1)	(2)	(3)	(4)
Any College	0.026	0.025	0.036	0.039
	(0.015)	(0.018)	(0.021)	(0.025)
$\bar{Y}_0$	0.905	0.877	0.865	0.838
Any 4-Year	0.027	0.028	0.036	0.037
	(0.015)	(0.019)	(0.021)	(0.026)
$\bar{Y}_0$	0.900	0.861	0.851	0.828
4-Year Private	0.145	0.146	0.154	0.153
	(0.022)	(0.024)	(0.028)	(0.033)
$\bar{Y}_0$	0.660	0.624	0.595	0.563
4-Year Public	-0.122	-0.124	-0.120	-0.118
	(0.020)	(0.020)	(0.024)	(0.027)
$\bar{Y}_0$	0.242	0.242	0.259	0.265
Observations	1387	1387	1146	876

Note: Coefficients on program admission come from an OLS regression. The sample includes applicants between 2005 and 2011 who made it past the second wave of screening. The outcomes are indicators for whether a student was enrolled in college one, two, three, or four years after senior year of high school, for different types of colleges. $\bar{Y}_0$ is the mean for post second screening students who were rejected from the program (the comparison group). The regression results control for full list of covariates in Table A3. Standard errors are robust to heteroskedasticity and displayed in parentheses. Data are from program applications, NSC, and peer schools.

Table 4—Effect of Summer Program Admission on Enrollment by School Selectivity

	Cohorts			
	2005-2011	2005-2011	2005-2010	2005-2009
Enrolled In	Year 1	Year 2	Year 3	Year 4
	(1)	(2)	(3)	(4)
HI	0.308	0.305	0.294	0.300
	(0.027)	(0.027)	(0.029)	(0.032)
$\bar{Y}_0$	0.142	0.139	0.13	0.114
STEM Except HI	-0.041	-0.043	-0.043	-0.049
	(0.014)	(0.014)	(0.015)	(0.017)
$\bar{Y}_0$	0.089	0.087	0.086	0.087
Top 10 Except HI	-0.001	-0.004	-0.000	0.002
	(0.003)	(0.007)	(0.006)	(0.008)
$\bar{Y}_0$	0.235	0.216	0.206	0.196
Top 11 to 20	-0.058	-0.061	-0.050	-0.052
	(0.013)	(0.013)	(0.015)	(0.015)
$\bar{Y}_0$	0.086	0.087	0.08	0.071
Most Competitive,	-0.058	-0.057	-0.064	-0.070
Except Top 20	(0.020)	(0.019)	(0.020)	(0.022)
$\bar{Y}_0$	0.14	0.133	0.133	0.129
Highly Competitive	-0.064	-0.065	-0.070	-0.065
	(0.018)	(0.017)	(0.018)	(0.021)
$\bar{Y}_0$	0.141	0.136	0.141	0.145
Very-Less Competitive	-0.101	-0.094	-0.078	-0.079
	(0.020)	(0.023)	(0.027)	(0.032)
$\bar{Y}_0$	0.255	0.29	0.311	0.345
Observations	1387	1387	1146	876

Note: Table reports OLS coefficients on program admission. Sample includes (post second screening) applicants from 2005 to 2011. Outcomes are indicators for whether a student was enrolled in college one, two, three, or four years after senior year of high school, for different types of colleges. $\bar{Y}_0$ is the mean for post second screening students who were rejected from the program (the comparison group). Top 20 rankings are from Avery et al. (2013); competitive rankings are from Barron's Profile of American Colleges. Results control for covariates in Table A3. Robust standard errors are in parentheses. Students enrolled in unranked schools, or no school, are put in Barron's category 5. Data are from program applications, NSC, and peer schools.

TABLE 5—EFFECT OF SUMMER PROGRAM ADMISSION ON APPLICATION, ADMISSION, AND ENROLLMENT AT THE HI AND OTHER PEER SCHOOLS WITHIN TWO YEARS OF HIGH SCHOOL

	Any Peer School (1)	Ivy (2)	HI (3)	Non-Ivy Excluding HI (4)	Liberal Arts College (5)
Panel A: Application					
Program	0.037	0.033	0.318	-0.090	-0.040
	(0.012)	(0.027)	(0.023)	(0.028)	(0.019)
$\bar{Y}_0$	0.764	0.554	0.453	0.524	0.111
Adjusted R-squared	0.773	0.268	0.383	0.233	0.068
Panel B: Admission					
Program	0.128	0.031	0.338	-0.040	-0.032
	(0.023)	(0.027)	(0.026)	(0.027)	(0.016)
$\bar{Y}_0$	0.567	0.324	0.224	0.343	0.089
Adjusted R-squared	0.398	0.177	0.316	0.163	0.058
Panel C: Enrollment					
Program	0.184	-0.039	0.289	-0.058	-0.008
	(0.026)	(0.023)	(0.026)	(0.018)	(0.008)
$\bar{Y}_0$	0.467	0.184	0.135	0.128	0.021
Adjusted R-squared	0.314	0.082	0.211	0.041	0.026
Observations	1387	1387	1387	1387	1387

Note: Table reports OLS coefficients on program admission. Sample is (post second screening) applicants from 2005 to 2011. Outcomes are application, admission, and enrollment within two years of the senior year of high school. $\bar{Y}_0$ is the mean for post second screening students who were rejected from the program (the comparison group). Results control for covariates in Table A3. Standard errors are robust to heteroskedasticity and displayed in parentheses. Data are from program applications, and peer schools.

Table 6—Effect of Summer Program Admission on Likelihood of Declaring a STEM Major

	NSC 2007-2011	NSC & Peer Schools 2007-2011	NSC & Peer Schools 2005-2011	
	(1)	(2)	(3)	(4)
Program	0.077	0.102	0.091	0.052
	(0.035)	(0.033)	(0.029)	(0.024)
$\bar{Y}_0$	0.464	0.520	0.517	0.694
Missing Values Imputed	No	No	No	Yes
Observations	1041	1041	1387	1387
Adjusted R-squared	0.069	0.124	0.103	0.084

Note: Coefficients on program admission come from an OLS regression. The sample includes applicants between 2005 and 2011 who made it past the second wave of screening. The outcome is whether a student ever declared a STEM major. $\bar{Y}_0$ is the mean for post second screening students who were rejected from the program (the comparison group). The regression results control for full list of covariates in Table A3. School type fixed effects are indicators for whether a student enrolled at the HI, a technical school other than the HI, or a non-technical school. The omitted category is no college. Missing values are imputed by giving admitted summer program students the STEM rate for non-missing rejected students and rejected students the rate of non-missing admitted students. Data are from the NSC, and peer schools.

Table 7—Effect of Summer Program Admission on Graduation Rates

	Cohorts		
	2005-2009	2005-2008	2005-2007
Graduated By	Year 4	Year 5	Year 6
	(1)	(2)	(3)
Any College	0.042	0.026	0.032
	(0.035)	(0.033)	(0.036)
$\bar{Y}_0$	0.601	0.783	0.806
Any 4-Year	0.037	0.027	0.034
	(0.035)	(0.034)	(0.037)
$\bar{Y}_0$	0.601	0.778	0.799
With STEM Degree	0.085	0.091	0.088
	(0.036)	(0.041)	(0.047)
$\bar{Y}_0$	0.406	0.542	0.531
Observations	876	669	513

Note: Table reports OLS coefficients on program admission. Sample includes (post second screening) applicants from 2005 to 2009. Outcomes are graduation by year four through six after the senior year of high school. $\bar{Y}_0$ is the mean for post second screening students who were rejected from the program (the comparison group). Comparison students who are missing degree major information are imputed to have the STEM graduation rate for admitted students and observations for admitted students that are missing degree major are imputed to have the STEM graduation rate of comparison students. The regression results control for covariates in Table A3. Robust standard errors are in parentheses. Data are from program applications, NSC, and peer schools.

Appendix For Online Publication Only

A1.

Table A1—List of Participating Peer Schools

Amherst College
Barnard College
Bowdoin College
Bryn Mawr College
Carleton College
Columbia University
Cornell University
Dartmouth College
Duke University
Georgetown University
Harvard University
Johns Hopkins University
Massachusetts Institute of Technology
Mount Holyoke College
Northwestern University
Oberlin College
Pomona College
Princeton University
Rice University
Smith College
Stanford University
Swarthmore College
Trinity College
University Of Chicago
University Of Pennsylvania
University Of Rochester
Washington University in St. Louis
Wesleyan University
Williams College
Yale University

Note: Source is Consortium on Financing Higher Education.

A2. Sensitivity to Selection Bias

Table A2 shows results of bounding exercises to see whether estimates of the core effects of the summer program are robust to selection bias. Following Oster (2013), I assume a proportional selection relationship between the observed and unobserved characteristics of summer program applicants, with δ as the coefficient of proportionality (in Altonji, Elder and Taber, 2005, $\delta = 1$ implicitly). R^2 is conservatively set to one in order to calculate the level of δ necessary to drive estimates of the impact of admission to zero. Intuitively, the question asked is how much of the variation in the outcome needs to be explained by unobserved variables relative to the variation explained by observed control variables in order for the selection bias to completely negate the estimated impacts. A natural upper bound on δ to consider as a benchmark suggested by both Oster (2013) and Altonji, Elder and Taber (2005) is $\delta = 1$; that is, unobservable characteristics are as important as observed characteristics in explaining the variation in the outcomes.

Column (1) reports the coefficient of proportionality that would yield an effect size equal to zero under the assumption of R-squared equal to one. Column (2) contains lower and upper bounds of the identified set obtained by assuming a delta and a maximum R-squared equal to one on one end of the set and the estimated coefficients on the other extreme of the set.

Selection bias would attenuate the estimated impacts. However, selection bias would have to explain as much as 40% of the variation captured by the rich set of controls in this analysis to invalidate all of the core results, and for enrollment at the HI, the constant of proportionality is near one. In other words, selection bias would have to be extreme, and given the early results of an experimental follow-up, it is unlikely that selection is large enough to explain the enrollment effects at the HI.

Table A2—Estimates of the Coefficient of Proportionality Necessary to Reduce Estimated Summer Program Effects to Zero, and Bounds on the Effect for Selected Outcomes

	δ (1)	Identified Set (2)
Enrolled HI	0.949	[-0.028, 0.310]
Enrolled in Barron's Very-Less Competitive Schools	0.728	[-0.149, 0.066]
Graduated with a STEM Major within 4 Years (Cohorts 2005-2009)	0.391	[-0.159, 0.092]

Note: Column (1) reports estimated coefficients of proportionality that would cause estimates of the program effect to be zero, assuming an upper bound on R-squared equal to one. Column (2) reports lower and upper bounds on the impact of program admission assuming δ equal to one and R-squared equal to one. Estimates in columns obtained using the psacalc command in Stata from Oster (2013). One of the bounds in column (2) also corresponds to the coefficient on program admission from a regression controlling for the full list of covariates in Table A3. Graduation with a STEM major is for four-year schools. Data are from program applications, NSC, and peer schools.

A3. Propensity Score Analysis

In OLS, all control observations are used in estimating program effects, with extrapolation occurring in covariate cells that do not have overlapping support. The most weight is placed on covariate cells with the highest treatment variance (Angrist and Pischke, 2009). Propensity score methods weight control observations differently, with the most weight going to control units with higher propensity scores.

In addition, it relaxes the implicit linearity assumption which has been shown to be meaningful in examining college quality (Black and Smith, 2004), there may be efficiency gains of propensity scores in finite samples (Angrist and Hahn, 2004), and certain propensity score methods may be less sensitive to model misspecification (Hirano, Imbens and Ridder, 2003).

Table A3 reports the estimates for the propensity score obtained using a logistic regression for both the full sample and the post second screening sample. Define $D_i = 1$ if a student was offered a spot in the summer program, and $D_i = 0$ if the student was not offered admission. Let $Z_i = \sum_j \alpha_j + X_i\gamma$, the vector of covariates from the application data and cohort fixed effects. The propensity score is defined as

$$\begin{aligned} P(D_i = 1|Z_i) =& \frac{\exp(Z_i'\beta)}{1 + \exp(Z_i'\beta)} \\ =& p(Z_i). \end{aligned}$$

Many of the application characteristics are significant predictors of program admission, and leaving fields blank is predictive (often in favor of admission). After conditioning on advancing past the second screening, fewer characteristics are associated with admission. These are generally socio-demographic variables, and PSAT math scores.

The propensity scores estimated in Table A3 are used in three different models: nearest neighbor matching models with one and three neighbors, and inverse propensity score weighting. These models were chosen to broadly characterize the variation likely to be present among propensity score estimators. Nearest neighbor matching generally has less bias, while inverse propensity score weighting generally has lower variance. Nearest neighbor matching is done without replacement when using one neighbor, and with replacement when using three neighbors.

The results in Table A4 are generally not sensitive to the model in terms of the size and sign of the coefficients. For example, estimates of the program impact on enrollment in any college from models two through five (which all control for covariates) range from 1.5 to 2.6 percentage points. Headline results are presented, but results for each outcome in the paper are available from the author upon request.

Table A3—: Propensity Score Model for Admission to the Summer Program

	Odds-Ratio (1)	S.E. (2)	Odds-Ratio (3)	S.E. (4)
Academics				
PSAT Math	1.031	(0.012)	1.033	(0.012)
PSAT Verbal	1.008	(0.010)	1.009	(0.009)
Took Calculus	0.878	(0.151)	1.063	(0.182)
Took Trigonometry	0.887	(0.169)	0.781	(0.144)
Took Algebra	1.244	(0.572)	1.542	(0.713)
Took Precalculus	0.989	(0.250)	1.165	(0.281)
Took Biology	1.553	(0.646)	1.615	(0.658)
Took Physics	1.106	(0.218)	1.039	(0.194)
Took Chemistry	0.425	(0.158)	0.522	(0.177)
Took Science	1.105	(0.248)	1.113	(0.241)
Extra-curriculars				
# Awards	0.993	(0.043)	1.013	(0.045)
# Extracurricular Activities	0.998	(0.050)	1.021	(0.054)
# Summer Experiences	0.987	(0.046)	0.991	(0.045)
# Work/Volunteer Experiences	0.985	(0.046)	0.970	(0.046)
Demographics				
Age at Program	0.949	(0.047)	0.958	(0.053)
Female	1.429	(0.185)	1.404	(0.180)
Black	4.619	(2.248)	4.264	(2.001)
White	5.046	(3.052)	9.815	(6.682)
Hispanic	4.182	(2.037)	4.161	(1.949)

Table A3—: (continued)

	Odds-Ratio (1)	S.E. (2)	Odds-Ratio (3)	S.E. (4)
Native American	7.983	(5.019)	8.631	(5.083)
Asian	1.722	(0.903)	2.212	(1.131)
Socioeconomic Status				
Income < $50K	2.240	(0.576)	2.112	(0.545)
Qualify Free Lunch	1.304	(0.343)	1.327	(0.352)
Any College Educated Parent	0.341	(0.053)	0.421	(0.064)
HS % Minority	0.887	(0.243)	1.100	(0.282)
Rural HS	1.416	(0.309)	1.542	(0.326)
HS Peer School App. Rate	0.573	(0.205)	0.322	(0.122)
Blank Application Fields				
Qualify Free Lunch	1.158	(0.514)	0.798	(0.359)
Income < $50K	1.302	(0.592)	0.778	(0.359)
Math Class Data	12.493	(6.533)	1.787	(0.938)
Science Class Data	5.897	(2.572)	1.057	(0.455)
PSAT Math	14.853	(8.821)	0.947	(0.260)
PSAT Verbal	0.268	(0.155)		
# Missing Fields	0.620	(0.015)	0.994	(0.042)
HS % Minority Students	3.385	(3.332)	3.262	(3.376)
Rural HS	0.538	(0.317)	0.635	(0.374)
HS Peer School App. Rate	0.119	(0.110)	0.176	(0.170)
Year Fixed Effects				
2006 Cohort	0.409	(0.128)	0.758	(0.237)

Table A3—: (continued)

	Odds-Ratio (1)	S.E. (2)	Odds-Ratio (3)	S.E. (4)
2007 Cohort	0.844	(0.287)	1.169	(0.384)
2008 Cohort	0.455	(0.164)	0.727	(0.261)
2009 Cohort	0.286	(0.101)	0.553	(0.193)
2010 Cohort	0.104	(0.055)	0.390	(0.200)
2011 Cohort	0.124	(0.064)	0.497	(0.253)
Observations		6324		1387
Pseudo R-squared		0.4315		0.094

Note: Columns (1) and (2) show results from a propensity score model estimated on all applicants between 2005 and 2011 using a logistic regression. Columns (3) and (4) show results from a propensity score model estimated on post second screening applicants between 2005 and 2011 using a logistic regression. Odds ratios are reported in columns (1) and (3). The variable indicating missing age data was omitted due to the small number of people missing age data. Standard errors are robust to heteroskedasticity and displayed in parentheses in columns (2) and (4). Data are from program applications.

Table A4—: Comparison of Effects of Summer Program Admission using OLS and Propensity Score Matching and Weighting Methods

	OLS		Propensity Score Matching		Inverse Probability Weighting
	No Covs	Covs	1 NN	3 NN	
Enrolled by Year 2	(1)	(2)	(3)	(4)	(5)
Any College	0.043	0.025	0.015	0.031	0.026
	(0.013)	(0.014)	(-0.019)	(-0.017)	(-0.013)
$\bar{Y}_0$	0.918	0.918	0.928	0.922	0.936
Any Four-Year	0.050	0.027	0.034	0.046	0.030
	(0.014)	(0.015)	(-0.021)	(-0.018)	(-0.014)
$\bar{Y}_0$	0.905	0.905	0.911	0.907	0.925
Four-Year Private	0.205	0.147	0.158	0.174	0.130
	(0.022)	(0.023)	(-0.031)	(-0.027)	(-0.023)
$\bar{Y}_0$	0.66	0.66	0.676	0.67	0.735
Four-Year Public	-0.156	-0.120	-0.124	-0.128	-0.100
	(0.019)	(0.02)	(-0.029)	(-0.025)	(-0.021)
$\bar{Y}_0$	0.246	0.246	0.235	0.236	0.19
Two-Year	-0.007	-0.002	-0.019	-0.015	-0.004
	(0.005)	(0.007)	(-0.009)	(-0.007)	(-0.005)
$\bar{Y}_0$	0.013	0.013	0.017	0.016	0.01
HI	0.350	0.310	0.313	0.347	0.310
	(0.026)	(0.027)	(-0.033)	(-0.029)	(-0.029)
$\bar{Y}_0$	0.138	0.138	0.151	0.139	0.178
Top 10	-0.012	-0.031	-0.006	-0.016	-0.036
Excluding HI	(0.024)	(0.025)	(-0.033)	(-0.029)	(-0.027)
$\bar{Y}_0$	0.237	0.237	0.235	0.238	0.261
Enrolled Top 20	-0.071	-0.092	-0.071	-0.084	-0.098
Excluding HI	(0.025)	(0.027)	(-0.035)	(-0.03)	(-0.029)

Table A4—: (continued)

	OLS		1 NN	3 NN	IPWRA
	(1)	(2)	(3)	(4)	(5)
$\bar{Y}_0$	0.326	0.326	0.326	0.33	0.353
Barron's Rank 1	-0.112	-0.137	-0.118	-0.133	-0.150
Excluding Top 20	(0.027)	(0.029)	(-0.038)	(-0.032)	(-0.031)
$\bar{Y}_0$	0.432	0.432	0.433	0.439	0.469
Barron's Rank 2	-0.033	-0.025	-0.030	-0.028	-0.030
	(0.012)	(0.013)	(-0.018)	(-0.014)	(-0.014)
$\bar{Y}_0$	0.067	0.067	0.066	0.066	0.064
Barron's Rank	-0.205	-0.149	-0.165	-0.186	-0.130
3 to 5	(0.023)	(0.023)	(-0.032)	(-0.028)	(-0.024)
$\bar{Y}_0$	0.363	0.363	0.35	0.357	0.288
Applied by Year 2					
Any Peer School	0.148	0.037	0.077	0.093	0.028
	(0.019)	(0.012)	(-0.022)	(-0.019)	(-0.01)
$\bar{Y}_0$	0.764	0.764	0.788	0.783	0.884
Ivy League	0.103	0.033	0.090	0.073	0.001
	(0.027)	(0.027)	(-0.038)	(-0.032)	(-0.028)
$\bar{Y}_0$	0.554	0.554	0.559	0.565	0.657
HI	0.399	0.318	0.345	0.375	0.293
	(0.023)	(0.023)	(-0.036)	(-0.03)	(-0.026)
$\bar{Y}_0$	0.453	0.453	0.472	0.461	0.559
Non-Ivy	-0.014	-0.090	-0.041	-0.034	-0.099
Excluding HI	(0.028)	(0.028)	(-0.038)	(-0.032)	(-0.03)
$\bar{Y}_0$	0.524	0.524	0.533	0.53	0.609
Liberal Arts	-0.019	-0.040	-0.043	-0.056	-0.048
	(0.017)	(0.019)	(-0.027)	(-0.023)	(-0.021)
$\bar{Y}_0$	0.111	0.111	0.119	0.123	0.141

Table A4—: (continued)

	OLS		1 NN	3 NN	IPWRA
	(1)	(2)	(3)	(4)	(5)
Admitted by Year 2					
Any Peer School	0.216	0.128	0.161	0.183	0.119
	(0.025)	(0.023)	(-0.033)	(-0.028)	(-0.025)
$\bar{Y}_0$	0.567	0.567	0.586	0.579	0.665
Ivy League	0.085	0.031	0.049	0.061	0.024
	(0.028)	(0.027)	(-0.038)	(-0.032)	(-0.03)
$\bar{Y}_0$	0.324	0.324	0.336	0.332	0.385
HI	0.393	0.338	0.373	0.401	0.324
	(0.026)	(0.026)	(-0.034)	(-0.029)	(-0.029)
$\bar{Y}_0$	0.224	0.224	0.231	0.221	0.293
Non-Ivy,	0.001	-0.04	-0.017	-0.009	-0.038
Excluding HI	(0.027)	(0.027)	(-0.037)	(-0.032)	(-0.03)
$\bar{Y}_0$	0.343	0.343	0.35	0.347	0.383
Liberal Arts	-0.018	-0.032	-0.032	-0.038	-0.036
	(0.015)	(0.016)	(-0.024)	(-0.02)	(-0.019)
$\bar{Y}_0$	0.089	0.089	0.094	0.096	0.106
Declared STEM Major, 2005-2011 Cohorts					
NSC and Peer	0.136	0.091	0.064	0.107	0.091
Schools	(0.028)	(0.029)	(-0.036)	(-0.033)	(-0.031)
$\bar{Y}_0$	0.517	0.517	0.541	0.527	0.562
NSC and Peer	0.060	0.052	0.022	0.045	0.045
with Missing Imputed	(0.023)	(0.024)	(-0.032)	(-0.028)	(-0.026)
$\bar{Y}_0$	0.694	0.694	0.707	0.699	0.709
Graduation					
2005-2009 Cohorts					
Graduated in 4 Years	0.087	0.052	0.105	0.105	0.071

Table A4—: (continued)

	OLS		1 NN	3 NN	IPWRA
	(1)	(2)	(3)	(4)	(5)
from a 4-Year	(0.033)	(0.034)	(-0.049)	(-0.04)	(-0.036)
$\bar{Y}_0$	0.593	0.593	0.587	0.587	0.609
Graduated w/ STEM Deg.	0.128	0.092	0.116	0.129	0.100
in 4 Years	(0.035)	(0.036)	(-0.048)	(-0.041)	(-0.038)
$\bar{Y}_0$	0.405	0.405	0.409	0.404	0.432
2005-2008 Cohorts					
Graduated in 5 years	0.057	0.030	0.066	0.086	0.056
from a 4-Year	(0.032)	(0.035)	(-0.049)	(-0.042)	(-0.039)
$\bar{Y}_0$	0.754	0.754	0.75	0.742	0.755
Graduated w/ STEM Deg.	0.113	0.085	0.068	0.108	0.094
in 5 Years	(0.039)	(0.042)	(-0.054)	(-0.047)	(-0.046)
$\bar{Y}_0$	0.524	0.524	0.542	0.526	0.543
2005-2007 Cohorts					
Graduated in 6 years	0.060	0.02	0.082	0.05	0.045
from a 4-Year	(0.036)	(0.04)	(-0.052)	(-0.042)	(-0.044)
$\bar{Y}_0$	0.765	0.765	0.756	0.769	0.779
Graduated w/ STEM Deg.	0.098	0.063	0.066	0.063	0.056
in 6 Years	(0.044)	(0.048)	(-0.061)	(-0.052)	(-0.051)
$\bar{Y}_0$	0.507	0.507	0.519	0.521	0.549

Note: Columns (1) and (2) show results for regression on an indicator for program admission with no covariates and controlling for observables, respectively. Columns (3) and (4) show results using nearest neighbor matching with one (without replacement) and three neighbors (with replacement), respectively. Column (5) shows results using inverse probability weighting adjusted estimates as in Hirano, Imbens and Ridder (2003). Standard errors are in parentheses and are robust to heteroskedasticity in columns (1) and (2), and to the fact that the propensity score is estimated in columns (3) through (5). Standard deviations are in brackets. Data are from program applications, NSC, and peer schools.

Table A5—: Predicting Missing Majors in the NSC, Conditional on Matching

		S.E.		S.E.
	(1)	(2)	(3)	(4)
Admitted to Program	-0.050	(0.026)	0.002	(0.028)
Year Fixed Effects	No		Yes	
Female			-0.034	(0.026)
Missing Gender			0.710	(0.067)
# Awards			-0.006	(0.009)
# Other Activities			0.006	(0.010)
# Summer Activities			0.008	(0.009)
# Work Activities			0.003	(0.009)
PSAT Math			-0.006	(0.002)
Missing PSAT Math			-0.115	(0.236)
PSAT Verbal			0.001	(0.002)
Missing PSAT Verbal			0.063	(0.232)
Took Calculus			0.007	(0.031)
Took Trigonometry			-0.037	(0.032)
Took Algebra			-0.069	(0.084)
Took Precalculus			-0.049	(0.045)
Took Biology			0.110	(0.064)
Took Physics			0.029	(0.035)
Took Chemistry			0.016	(0.060)
Took Science			-0.023	(0.040)
Missing Math Class Data			-0.015	(0.103)
Missing Science Class Data			-0.007	(0.077)
Age at Program			-0.002	(0.030)
Any College Educated Parent			0.029	(0.031)
Income < \$50k			-0.033	(0.054)
Missing Income Indicator			0.064	(0.077)
Quality for Free Lunch			0.004	(0.055)
Missing Lunch Indicator			-0.022	(0.073)
Black			-0.070	(0.077)
White			-0.241	(0.101)
Hispanic			-0.142	(0.077)
Native American			-0.055	(0.106)
Asian			-0.079	(0.085)

Table A5—: (continued)

	(1)	Std. Error (2)	(3)	Std. Error (4)
# Missing Fields			0.009	(0.007)
HS % Minority Students			-0.024	(0.053)
Missing HS % Minority			0.221	(0.216)
Rural HS			-0.059	(0.045)
Missing Rural HS			0.081	(0.105)
HS Peer School App. Rate			-0.100	(0.080)
Missing HS App. Rate			-0.250	(0.227)
Constant	0.266	(0.016)	0.395	(0.566)
Observations	1152		1152	
Adjusted R-squared	0.002		0.100	

Note: Column (1) reports the results of predicting the likelihood of missing a major in the NSC using admission to the summer program. Column (3) reports the results of predicting the likelihood of missing a major in the NSC using admission to the program and other observable characteristics for students in the NSC. Even columns report heteroskedasticity robust standard errors for each regression. Data are from program applications, NSC, and peer schools.

Table A6—Percent of Students Missing Data on Majors by Treatment and School

	Missing Data in NSC			Missing Data in NSC and Peer Schools		
	All (1)	Admitted (2)	Rejected (3)	All (4)	Admitted (5)	Rejected (6)
HI	21%	25%	13%	1%	1%	2%
Observations	355	228	127	355	228	127
Tech, except HI	25%	26%	24%	25%	26%	24%
Observations	101	19	82	101	19	82
Non-tech	39%	42%	39%	29%	31%	28%
Observations	838	202	636	838	202	636
All schools	33%	33%	33%	21%	15%	24%
Observations	1294	449	845	1294	449	845

Note: The first three columns show the percentage of students who attended college and are missing degree or declared major data in the NSC. The last three columns show the percentage of students who attended college and are missing degree major or declared major data in the NSC and peer schools. The sample is restricted to applicants who made it past the second screening. Data are from the NSC and peer schools listed in Appendix Table A1.