CALDER

AIR

The Potential for Community College Students to Expand and Diversify University Degree Production in STEM Fields

Cheng Qian
Cory Koedel

The Potential for Community College Students to Expand and Diversify University Degree Production in STEM Fields

Cheng Qian
University of Missouri

Cory Koedel
University of Missouri/CALDER

Contents

Acknowledgments

The Potential for Community College Students to Expand and Diversify University Degree Production in STEM Fields

Cheng Qian & Cory Koedel
CALDER Working Paper No. 244-1020
October 2020

Abstract

We use rich administrative microdata from Missouri to examine the potential to expand and diversify the production of STEM degrees at universities by tapping into the population of community college students. We find that the scope for expansion is modest, even at an upper bound, because most community college students have academic qualifications that make them unlikely to succeed in a STEM field at a university. We also find there is almost no scope for community college students to improve the racial/ethnic diversity of four-year STEM degree recipients. We conclude that it will be challenging to expand and diversify STEM degree production at universities with interventions targeted toward community college students.

1. Introduction

We examine the potential for expanding and diversifying the production of university degrees in science, technology, engineering, and mathematics (STEM) fields by tapping into the pool of students who attend community colleges. Our analysis is based on a thought experiment in which we nudge—either in the traditional sense of the word (Thaler and Sunstein, 2008) or by meaningfully altering student incentives—academically-qualified community college students to enroll in universities.

Our focus on STEM fields is motivated by concerns that the United States is falling behind globally in the production of STEM human capital and this will adversely impact long-term economic prosperity (Atkinson and Ezell, 2012; National Academy of Sciences, National Academy of Engineering, & Institute of Medicine of the National Academies, 2007). Improving and expanding STEM education has been a consistent policy priority at the highest levels of government in the U.S. (National Science & Technology Council, 2018; White House, 2016) and an area of active scholarship (e.g., see Coleman, Smith and Miller, 2019). Underlying reasons for the focus on STEM education are its perceived importance for innovation and the potential for positive spillover effects of STEM-trained workers (Shambough, Nunn, and Portman, 2017; Winters, 2014).

Diversifying the STEM workforce is also an explicit policy objective (e.g., White House, 2016). Participation in STEM fields is low among women and underrepresented minorities (URMs; i.e., Black and Hispanic workers) relative to White and Asian men. An implication is that the size of the STEM workforce can be expanded by increasing participation among these groups (Anderson and Kim, 2006; Committee on Science, Engineering, and Public Policy, 2011). STEM graduates also earn significantly more than graduates in other fields, on average,

which suggests that diversifying the STEM workforce can reduce earnings inequality (Altonji, Blom, and Meghir, 2012; Fayer, Lacey, and Watson, 2017; Kinsler and Pavan, 2015). Numerous government programs operate with the goal of improving STEM diversity.[1]

Our focus on community college students is motivated by several factors. First, in efforts to expand and diversify degree production in STEM fields, the most natural alternative to community college students is students who already attend universities and either (a) tried and failed in a STEM field, or (b) made no attempt to pursue a STEM degree. While these students are an appealing group to consider in some respects—most notably, many have strong academic qualifications—a significant drawback is that they have actively decided against pursuing STEM degrees. This is important because Kirkeboen, Leuven, and Mogstad (2016) show that students choose their fields of study based on comparative advantage, which implies that altering these decisions may be undesirable. In contrast, nudging academically-qualified community college students up to the four-year-university level, then letting them self-select into STEM fields, ensures that students' inherent field-selection processes are preserved.[2]

The community college population is an also appealing group because it is diverse demographically and socioeconomically. Students who attend community colleges are more likely to come from groups that are traditionally underrepresented at universities along the dimensions of race/ethnicity and income (Deming, Goldin and Katz, 2012; Provasnik and Planty, 2008; Wang, 2013). They also have a revealed preference for the pursuit of higher education and

---

[1] Examples include the USDA's Women and Minorities in Science, Technology, Engineering and Mathematics Fields Program (WAMS) and the US Department of Education's Developing Hispanic Serving Institutions STEM and Articulation Program.

[2] Empirically, it is uncommon for students who start in non-STEM fields to switch to STEM fields during college (Stinebrickner and Stinebrickner, 2014) and Kerr et al. (2020) show that getting students to change majors is generally difficult.

most indicate university aspirations.[3] The potential benefits of infusing the STEM pipeline with

community college students have been discussed recently in Bahr et al. (2017), Evans, Chen, and

Hudes (2020), Hagedorn and Purnamasari (2012), and Terenzini et al. (2014). Bahr et al. (2017)

claim "the potential role of community colleges in the production of STEM degrees and

professionals is undeniably important, as is the role of community colleges in providing access to

STEM pathways for historically disadvantaged groups" (p. 433). Similarly, Evans, Chen and

Hudes (2020) argue that community colleges "can act as a bridge between local high schools, 4-

year institutions, and the STEM workforce" (p. 247).

Noting these appealing aspects of the community college population, there are challenges

associated with intervening with these students. First, success rates of transfer students from

community colleges in STEM are low (Wang, 2015) and policies that ease transfer requirements

have been shown to have little effect on outcomes at universities (Baker, 2016; Gross and

Goldhaber, 2009; Roksa and Keith, 2008).[4] This suggests that boosting the STEM pipeline

through transfer policies is unlikely to be successful. For this reason, the hypothetical policy we

consider aims to circumvent community colleges entirely by re-routing academically-qualified

students directly to universities.[5]

Another consideration is whether enough community college students have the academic

preparation necessary to succeed in university STEM programs. Hoxby and Avery (2013) find

that there are many academically-qualified, low-income students who are undermatched to

---

[3] Using data from the Beginning Postsecondary Students Longitudinal Study (BPS), Deming, Goldin and Katz (2012) and Horn and Skomsvold (2012) report that about 80 percent of first-time community college students self-report their education goals as a bachelor's degree or higher.

[4] We are not aware of any studies that have looked at the effect of transfer policies on STEM outcomes specifically, but the general lack of efficacy evidence for these types of policies suggests that any STEM specific policies in a similar vein will face similar challenges.

[5] Although we do not find any policies that specifically try to re-route academically-qualified students directly to universities, evidence from Carrell and Sacerdote (2013) and Hyman (2020) suggest that some interventions aiming to increase college attendings have more impact on enrolling in four-year colleges.

postsecondary institutions, including community colleges. These students would be prime candidates to move up to more rigorous academic programs at universities in STEM fields. However, Chetty et al. (2020) find that the Hoxby and Avery numbers are likely inflated and there are many fewer of these students. Our study is an indirect, empirical test of sorts for these competing numbers—undergirding our analysis is the question of whether there are enough academically-qualified students who attend community colleges to meaningfully boost the production of STEM degrees at universities, if we could shift their enrollment.

Our research design is an exercise in predictive modeling. We use rich administrative data provided by the Missouri Department of Higher Education and Workforce Development (DHEWD) to design and evaluate a hypothetical policy that can be thought of as a perfectly-effective "nudge" intervention; i.e., in which all nudged individuals respond as intended. The nudge shifts initial community college enrollees to attend universities instead. We start by identifying the subpopulation of community college students likely to succeed in a university STEM program based on observable information. To do this, we use a flexible logistic regression to estimate the likelihood of STEM degree attainment among university students, then apply the parameters out of sample to community college students. We label a community college student with academic qualifications that imply a (relatively) high likelihood of STEM degree completion at a university as "STEM qualified." Assuming we could nudge all of these students to attend universities, we predict their individual likelihoods of completing STEM degrees and produce summary predictions of total STEM degrees produced. We also examine the diversity of students who we predict would earn degrees.

Our predictions assume that initial entrants into community colleges are just as likely to succeed at universities as their observationally-similar peers who start at universities, which is

unlikely. As noted above, we also assume that every student we "nudge" changes his or her behavior, which is well outside of the bounds of what can be expected from a plausible real-world intervention (e.g., see Bird et al., 2019; Castleman and Page, 2016; DellaVigna and Linos, 2020; Oreopoulos and Petronijevic, 2019).[6] Primarily for these two reasons, our estimates reflect upper bounds—probably high upper bounds—on the potential for similarly-spirited, real-world policies to affect the production of STEM degrees at universities. We also perform analyses to get more realistic estimates by (1) parameterizing and removing bias from unobserved selection into two-year and four-year colleges, and (2) parameterizing more realistic behavioral changes in response to our hypothetical intervention.

The most closely-related literature to our work includes studies of policy changes and other interventions targeted toward community college students that encourage them to either transfer to or directly enroll in a university (Castleman and Page, 2016; Marx and Turner, 2019; Hyman, 2020; Gurantz et al., forthcoming). We note two major differences between our work and these previous studies. The first difference is that we focus specifically on how community college students can impact *STEM* degree production at universities, motivated by policy interest in the STEM workforce. The second difference is that we do not study a specific policy intervention, but rather focus on the question of whether there is the *potential* for the community college population to be tapped in this way. We view our contribution as a predecessor to policy-intervention studies. We ask whether there is the scope for such policies to be effective, and if so, at what scale, which can inform future efforts to expand and diversify the STEM pipeline.

Our upper-bound results suggest modest potential to expand the production of four-year STEM degrees by tapping into the pool of community college students. The expansion effect is

---

[6] This is the true whether with respect to a textbook nudge as traditionally defined, or a more substantial and costly intervention.

modest, even at an upper bound, because the vast majority of community college students are not academically prepared to succeed in a STEM program at a university. While it is not surprising that few community college students are academically prepared for STEM programs at universities, we show that the magnitude of the drop-off is substantial from the full population of community college students to those who have an appropriate level of academic preparation. When we add more realistic features to our analysis to get away from the upper bound, the expansion potential of policies targeted toward students who attend community colleges shrinks rapidly.

We also find that there is no scope for the community college pipeline to improve the racial/ethnic diversity of four-year STEM degree recipients. Although the community college population on the whole is more diverse than the university population, our analysis reveals that most URM students attending community colleges are not academically prepared to succeed in STEM at universities. Thus, while at a cursory glance the community college population seems like an appealing source to diversify the university-trained STEM workforce, our analysis suggests efforts in this regard will likely fall flat.

A broad takeaway from our study is that it will be challenging to expand and diversify the pool of university-trained STEM workers with interventions targeted toward students who would otherwise plan to attend community college. This finding contributes to what is emerging as a common theme of research examining the determinants of postsecondary outcomes: interventions at the postsecondary level are too late (Cameron and Heckman, 2001; Arcidiacono and Koedel, 2014; Stinebrickner and Stinebrickner, 2014).

2. Data

We use administrative microdata from the Missouri DHEWD for our analysis. The data

contain student background characteristics (race, gender, age, high school attended, etc.), pre-entry academic qualifications (high school class percentile rank, ACT test scores), and in-college outcomes (majors, credits, GPA, and graduation). We restrict our analytic sample to first-time, full-time, state-resident students who entered the public college system—which includes 13 universities and 14 community colleges—between 2006 and 2010 as college freshman. We track students for up to six years after initial entry into the system to determine whether they graduate with a four-year degree from any public college in Missouri.[7]

Table 1 shows summary statistics for all university and community college students, and for various restricted subsamples that lead to our primary analytic sample. We review our data restrictions briefly here and examine the sensitivity of our findings to relaxing them below. The first restriction, imposed going from columns (1)-(2) and (4)-(5) in Table 1, focuses the analysis on in-state students only. This is due to data limitations for out-of-state students, especially at two-year colleges.[8] The more substantive restrictions occur moving from columns (2)-(3) and (5)-(6), where we drop students from the sample who are (a) older than 20 upon entry as a first-time college student, (b) enrolled part-time at entry, which we define as attempting fewer than 12 credits, (c) missing math or English ACT scores, and/or (d) are from a very small high school (i.e., that sent five or fewer students to a public university during the period covered by our data panel) or the high school attended is missing.

---

[7] Our data are comprehensive for public colleges and universities statewide but do not cover private or out-of-state institutions. This limits the generalizability of our analysis—most directly in that our results cannot speak to the potential for expanding the STEM pipeline by redirecting community college students to private universities or to universities outside of the state. We view this as a modest limitation in the context of our thought experiment because if we were to nudge community college students to universities at scale, existing transfer patterns suggest that they would be more likely to gravitate toward public in-state universities (Shapiro et al., 2017).

[8] The key data issue is that especially for students who attend two-year colleges, out-of-state students often have missing ACT scores even when they took the test. Some ACT scores are reported directly by institutions, but we also have access to scores for all ACT test takers in Missouri. We also use information about students' individual high schools to predict their success in college and out-of-state students attend high schools that are typically sparsely attended by Missouri college-goers, which creates analytic problems in our empirical models.

The age restriction is to focus our thought experiment on the population most likely to be susceptible to an intervention that shifts the sector of enrollment. The ACT and full-time-enrollment restrictions are because we treat the steps of taking the ACT prior to college enrollment, and enrolling full time, as indicators of stronger interest and ability to pursue a higher-level degree among community college students (virtually all four-year college entrants have ACT scores and enroll full time). We drop students from small high schools because we use the high school attended in our prediction models and small schools are problematic empirically.[9] Additional information about data construction, including more information about how the sample changes and summary statistics as each data restriction is enforced, is in Appendix Tables A1-A3. Again, we examine the substantive implications for our findings to relaxing these data restrictions below.

Students' class percentile ranks are important predictors of STEM success and Table 1 shows that class percentile ranks are particularly likely to be missing among community college students. This is because of inconsistent institution-level reporting in the DHEWD data, which derives partly from the fact that community colleges' open enrollment policies do not require them to collect data on academic qualifications.[10] For students with missing class percentile ranks, we use linear regression to impute their ranks based on their demographic information, math and English ACT scores, and high schools of attendance. A feature of imputing via linear regression is that the imputed values are shrunken toward the mean, which is problematic because it is students in the upper tail of the distribution of academic qualifications among community college students who are most likely to succeed in STEM fields at universities. To address this issue, we inflate the variance of the imputed values *ex post* at each college level

---

[9] This restriction has no substantive implications for our analysis because it affects very few students.
[10] In contrast, we have data on ACT scores for all students who took the test in Missouri, as noted above.

(two-year and four-year) to match the variance of observed percentile ranks at the same level. We also examine the sensitivity of our findings to the variance inflation procedure in the robustness section.

The racial/ethnic diversity information in Table 1 previews our finding regarding the potential for community college students to impact the racial/ethnic diversity in STEM fields at universities. First, while the full community college population has a higher proportion of Black students than the university population (given Missouri demographics, the proportion Black is the most relevant consideration for diversity)—0.14 versus 0.12—the gap in Black representation is not large. Moreover, the Black share among community college students falls to just 0.10 after we restrict the sample to in-state, full-time students (with the latter restriction being most impactful), and then to just 0.08 in the final sample after we add the additional restrictions for age and taking the ACT. In the analytic sample, the Black share among community college students is below the Black share among university students.

We use the Classification of Instructional Programs (CIP) to identify majors in STEM fields, then divide university students into groups of STEM and non-STEM entrants based on their majors at entry. Following Darolia et al. (2020), we use the NSF definition of STEM fields, which includes majors in mathematics, natural sciences, engineering, computer and information sciences, and selected technical subfields within the social and behavioral sciences.

Table 2 shows the summary statistics for STEM and non-STEM university entrants. We also include summary statistics for STEM completers. Compared to non-STEM entrants, STEM entrants have higher ACT math scores (25.58 vs 22.16), higher ACT English scores (25.19 vs 23.27), and higher percentile ranks in high school (0.77 vs 0.69). Compared to STEM entrants, students who successfully earn a STEM degree possess even stronger academic qualifications.

Consistent with previous research, there is a significantly lower percentage of female students in STEM fields (Chen, 2009) and a slightly lower percentage of URM students (Hill, 2017). But whereas the percentage of female students is the same among STEM completers and STEM entrants, there is a significant drop in the percentage of Black students in STEM fields from entrants to completers (also see Arcidiacono, Aucejo, and Spenner, 2012). Unsurprisingly, both STEM entrants and non-STEM entrants at universities possess substantially stronger academic qualifications than their community college counterparts.

The bottom row of Table 2 shows that 44 percent of STEM entrants graduate with a STEM degree in 6 years, but just 4 percent non-STEM entrants transfer to STEM fields and graduate with a STEM degree. This highlights the importance of the initially-declared major in determining the production of four-year STEM degrees.

3. Methodology

Our goal is to determine the potential for the STEM pipeline at universities to be expanded and diversified by tapping into the population of students who attend community colleges. We situate our investigation in the context of a behavioral intervention that shifts student enrollment from community colleges to universities. For most of our analysis we assume that all community college students who we choose to intervene with respond as intended by enrolling in a university. Based on the existing nudge literature, this is well outside the bounds of what is plausible. It is also implausible for an intervention with forceful incentives—i.e., an intervention that is more than a nudge. This feature of our study contributes to the interpretation of our estimates as giving upper bounds.

The first step in our process is to identify who to target among the community college population for our hypothetical intervention. If the objective function is simply to maximize the

number of STEM degrees produced, then the optimal policy would be to target all community college students. However, it would be costly and undesirable to shift students to universities who are underprepared or uninterested in STEM fields given their low likelihoods of success. Therefore, we focus only on students whose observable characteristics suggest they are reasonably likely to succeed in a STEM field. We refer to these students as "STEM qualified," noting that this term broadly reflects academic preparation and interest in STEM fields.

We propose a simple data-driven framework to identify "STEM-qualified" community college students. We begin by estimating the following empirical model to predict STEM degree completion among four-year college students within six years of initial enrollment, which we specify as a logistic regression:

$$Y_{ijt}^* = \mathbf{X_i}\boldsymbol{\beta_1} + \gamma_j + \delta_t + \varepsilon_{ijt} \tag{1}$$

In equation (1), Y_{ijt}^* is the latent utility of completing a STEM degree within six years, versus not completing a STEM degree, for student i from high school j who first enrolled in one of Missouri's 13 four-year public universities in year t. Students who complete a STEM degree within six years—i.e., $Y_{ijt} = 1$—have latent utility above zero. $\mathbf{X_i}$ is a vector of control variables including student' ACT math and English scores, high school percentile ranks, racial/ethnic and gender designations, and interaction terms between race/ gender and ACT scores/ rank. γ_j is a fixed effect for high school j and δ_t is a fixed effect for year t. ϵ_{ijt} is the error term.

The fitted values from equation (1), $\hat{P}_{ijt} = Pr(Y_{ijt} = 1|\mathbf{X_i}, \gamma_j, \delta_t)$, indicate the predicted likelihood of completing a STEM degree conditional on pre-entry student characteristics and qualifications among four-year-university entrants. The next step in our process is to apply the parameter estimates from equation (1) to the profiles of community college students. This generates predicted values $\hat{P}_{ijt}^{cc}$, where the superscript cc denotes that the value is an out-of-

sample prediction for community college student i. $\hat{P}_{ijt}^{cc}$ is the likelihood that student i would complete a four-year degree in a STEM field if the student initially enrolled in a university instead of a community college, and if there were no unobserved differences between observationally similar students who differ by initial enrollment sector.

We identify the subpopulation of community college students who are "STEM qualified" based on the distribution of $\hat{P}_{ijt}$ among university STEM entrants. Specifically, in our preferred set up, we identify student i as "STEM qualified" if $\hat{P}_{ijt}^{cc} > \tilde{\tilde{P}}$, where $\tilde{\tilde{P}}$ is the median predicted likelihood of STEM success among initial STEM entrants in the university sample. Again, although we refer students with $\hat{P}_{ijt}^{cc} > \tilde{\tilde{P}}$ as "STEM qualified," the predicted values are more-precisely described as embodying two factors that determine STEM success: academic preparation and interest in pursuing a STEM degree. In our data, $\tilde{\tilde{P}} \approx 0.17$.

Whether the median value $\tilde{\tilde{P}}$ is an appropriate threshold for identifying STEM-qualified students is a normative question. On the one hand, we want to choose a threshold that is high enough that affected students would have a reasonable likelihood of success in STEM fields. On the other hand, total STEM degree production is at least weakly increasing as our threshold for STEM-qualified declines. We use the median success rate among observed four-year college STEM entrants as our primary threshold because it is an intuitive, data-driven anchor for this value. We consider the sensitivity of our estimates to modifications of the threshold below.

The policies we have in mind are of the sort that alter behavior—either by a textbook nudge or a nudge combined with stronger incentives—such that the community college students we identify as "STEM qualified" instead choose to enroll in universities. The most-commonly studied interventions of this type in recent research are nudges that encourage students to make different college and major choices (Bird et al., 2019; Castleman and Page, 2016). Although the

literature is clear that the efficacy of these types of behavioral interventions is limited, we

assume that all students who we intervene with will change their enrollment behavior. Under this

assumption, and continuing to assume these students would succeed in STEM at the same rates

as their observationally similar counterparts who initially enrolled in universities, the total

number of predicted four-year STEM degrees produced among STEM-qualified community

college students is given by:

$$\theta_{STEM}^{cc} = \sum_{i=1}^{N_S^{cc}} \hat{P}_i^{cc} \qquad\qquad (2)$$

Finally, noting the interpretive caveats given above, it is also straightforward to modify

equation (2) to get numbers for specific demographic groups to inform the diversity question

(i.e., we can redefine the summations to be over targeted groups only). To obtain error bands for

our estimates of STEM degrees produced that account for error throughout the process described

in this section, we bootstrap the entire procedure 500 times and report 95-percent empirical

confidence intervals based on the bootstrapped values.

4. Primary Results

Table 3 shows the raw logit coefficients and bootstrapped 95% confidence intervals for

equation (1), estimated on the university sample and using our preferred specification. The

results are intuitive and consistent with past research showing that students who succeed in

STEM fields are positively selected (Arcidiacono and Koedel, 2014; Arcidiacono et al., 2016).

Among the pre-entry academic qualifications in our data, the class percentile rank is by far the

strongest predictor of STEM success. The ACT math score is also a significant predictor of

STEM success.[11] In terms of demographics, the familiar gender difference in STEM success is

---

[11] The coefficient on the ACT English score is negative, which is perhaps unintuitive, but this reflects the conditional relationship only—if the ACT English score is included without any other controls, the coefficient is positive (results omitted for brevity).

clearly present in our data (also see Kahn and Ginther, 2017). Similarly, the well-documented lack of differences in success by race/ethnicity after conditioning on pre-entry academic qualifications, with the exception of Asian students, is also present (Griffith, 2010; Sass, 2015).

For all students in both the two-year and four-year samples, we use the median predicted likelihood of STEM success among initial STEM entrants as the threshold to identify "STEM qualified" students. Note that non-STEM entrants at universities can also be STEM-qualified by this definition, even if they prefer a non-STEM major. Summary statistics for these groups are shown in Table 4. Overall, about a fourth of all university entrants have academic qualifications that align with our definition of STEM qualified. These students have stronger academic qualifications along all dimensions than their non-STEM-qualified counterparts. The magnitudes of the differences in core qualifications (ACT scores and percentile ranks) are large, ranging from 0.8 – 1.4 standard deviations.

STEM-qualified students are also more likely to be male and more likely to be white or Asian than non-qualified students, but for different reasons. The gender gap reflects the strong negative coefficient for female students in Table 3. In contrast, the low representation of Black and Hispanic students is not driven by conditional differences in the likelihood of succeeding in STEM by student group—this can be seen by the insignificant coefficients on the racial/ethnic indicators for these groups in Table 3. Instead, the racial/ethnic differences emerge due to differences in pre-entry academic qualifications, which are much lower on average for Black students in particular (this result is also consistent with previous research—e.g., see Arcidiacono and Koedel, 2014; Arcidiacono et al., 2016; Bahr et al., 2017).

Unsurprisingly, the fraction of community college students whose pre-entry characteristics and qualifications are sufficient to meet our definition of STEM-qualified is much

14

smaller than the fraction of four-year students, at around 7.4 percent. This can be seen in the bottom row of Table 4. Moreover, among the STEM-qualified group in community colleges, their academic qualifications are clearly below those of their STEM-qualified peers at four-year institutions. This reflects the fact that the distribution of academic readiness at community colleges, as measured by the observable information we have, is to the left of the distribution of academic readiness among four-year college students. The implication is that among students above the STEM-qualified threshold, those at community colleges are closer to the threshold value, on average, than their four-year-college counterparts. A notable result in Table 4 is that the STEM-qualified population of community college students does not include a larger proportion of Black students than the STEM-qualified population at universities.

In the first column of Table 5 we show our estimates for STEM degree production among the community college sample from equation (2). We report the total number of four-year STEM degrees produced among STEM-qualified community college students and the characteristics of completers. Recall that we bootstrap our entire procedure 500 times and the results in Table 5 are the average outcome values across the 500 bootstrap replications, with 95-percent confidence intervals reported in parentheses. The second column replicates descriptive statistics for STEM completers among university entrants from Table 2 for comparison.

We focus first on our findings regarding the potential to expand the STEM pipeline, then turn to diversity. In total, recall that in Table 4 we move 3,209 STEM-qualified community college students to universities. Of these, our model based on observables predicts that 869 would complete a STEM degree within 6 years (with an empirical 95 percent confidence interval of 778-965 students). The number of STEM degrees produced among four-year entrants over our

sample period was 9060 students (column 2), meaning that our estimate of 869 degrees corresponds to an increase in production of 9.6 percent.

This is a non-negligible increase, although a reasonable interpretation is that it is actually quite small given the upper-bound assumptions built into our analysis. Moreover, the total number of STEM degrees conferred among the nudged sample overstates the number of new STEM degrees produced because in the absence of our hypothetical intervention, some STEM-qualified community college students would still complete university STEM degrees (i.e., via transfer). Among the 3,209 students we would hypothetically intervene with, we observe 289 actually transferred to a university and obtained a STEM degree within 6 years. Thus, our upper-bound estimate of the net increase in degrees produced is 580 (869-289), or about 6.4 percent relative to observed STEM degree production.

Turning to the characteristics of the new STEM completers, column (1) shows that their academic qualifications on average are similar to but slightly below their counterparts who start at universities: their average ACT math scores are 25.76, versus 26.63 for university entrants, and their average high school percentile ranks are 81 versus 82. There is no diversity improvement by race-ethnicity among STEM degree recipients in the community college sample relative to the four-year sample. In fact, the fraction of nudged community college students who complete a STEM degree and are Black (0.01) is substantially lower than the fraction of university students who complete a STEM degree and are Black (0.04). The result is driven by the low share of Black students at community colleges who meet our definition of STEM qualified (Table 4 shows that just 2 percent of STEM-qualified community college students are Black and the 95% confidence interval is just 1-2 percent).

The new STEM graduates from community colleges are also even more male-dominated than their four-year counterparts. This result derives from the fact that female community college students do not outperform their male peers academically to the same degree that female university students outperform their male peers. Put another way, female community college students are more negatively selected in terms of academic qualifications, in their gender-specific distribution, than their male counterparts.

5. Robustness

In this section we explore the basic robustness of our findings. First, to get a better understanding of the plausibility of our out-of-sample predictions for community college students under the maintained assumption of selection on observables, we document the predictive validity of our models in and out of sample among four-year college students (for whom true outcomes are observed, which is required to test predictive validity). To facilitate in-sample and out-of-sample comparisons, we use 80% of the data to comprise the "training dataset" and the remaining 20% to test predictive validity. We follow the same procedure described in the methodology section with this new data split: we use the training dataset to estimate equation (1), then apply the estimated parameters to the prediction dataset, which in this case is the 20% holdout sample of four-year entrants.

In-sample and out-of-sample prediction accuracy are shown in Appendix Table A5. Columns (1)-(2) show the in-sample comparison of true outcomes versus predicted values, and Columns (3)-(4) give the out-of-sample comparison. For the in-sample comparison, unsurprisingly, we see no difference between predicted and true-outcome values. For the out-of-sample comparison the actual and predicted values are also nearly identical. This basic test confirms that the prediction model is effective when applied to university students.

Next, we examine the robustness of our findings to modifying our procedure for imputing missing high school percentile ranks. As noted above, we inflate the variance of the imputed percentile rank values in order to offset the shrinkage inherent to the imputation process in our primary analysis. In Table 6, we explore the implications of using the shrunken class-rank values directly without the variance inflation. Following on the discussion above, a straightforward prediction is that using the shrunken values directly will reduce the number of community college students identified as STEM-qualified by reducing the prevalence of students with imputed values in the tails of the class-rank distribution. The bottom row of Table 6 shows that this is indeed the case—we predict a gross increase in STEM degree recipients of just 789 students in Table 6 versus 869 students in Table 5.

6. Minor Extensions

Next we extend our analysis along several minor dimensions. We describe the extensions in this section as "minor" because their substantive implications for our key findings regarding STEM expansion and diversification are modest.

6.1   *Demographic Predictors of STEM Success*

In our preferred specification of equation (1) we include race/ethnicity and gender indicators, and interactions between these indicators and ACT scores and percentile ranks, to improve the accuracy of our predictions. To test the implications of this for our diversity findings, in Table 7 we replicate our entire procedure after (a) dropping the race/ethnicity and gender interactions with ACT scores and percentile ranks, but keeping the race/ethnicity and gender indicators themselves, and (b) dropping all racial/ethnic and gender information from the models. The results from these scenarios are shown in columns (3)-(6) of Table 7; columns (1)-(2) replicate the results from our preferred specification for comparison.

Substantively, there are two main takeaways from Table 7. First, the racial/ethnic composition of STEM completers in the community college sample does not depend on whether we rely on racial/ethnic data in the prediction model, with the exception of Asian students, whose predicted STEM completion rate declines slightly when race-ethnicity information is omitted. This result is consistent with the pattern of estimates in Table 3, which shows that outside of Asian students, the racial/ethnic designations are not important predictors of STEM completion conditional on measured academic preparation.[12] Second, in terms of gender diversity, the female shares jump markedly in columns (5) and (6) if we remove all information about gender from the predictive model. This is because female students have generally strong academic qualifications (much stronger than for male students, on average). If we ignore their preference for non-STEM fields in our prediction models, we predict many more female students would pursue and complete STEM degrees. Again, this result is not surprising based on the estimates in Table 3, and only informative if one believes that female students who attend two-year colleges have fundamentally different preferences for STEM education than their four-year counterparts, which we view as unlikely.

6.2   Modifications to the Sample Restrictions

In Section 2 we describe a number of restrictions that we impose on our preferred estimation sample. In this section we examine the substantive implications of these restrictions for our findings.

First, in Table 8 we relax the credit-hour and age-based restrictions, which are set to 12 first-semester credits and age ≤ 20 in the main analysis. We consider reductions of the credit-hour constraint to 9 and then 6 credit hours to accommodate part-time students, and raise the

---

[12] That said, the evidence in Table 7 is more comprehensive because Table 3 does not show all of the interaction coefficients.

19

Substantively, there are two main takeaways from Table 7. First, the racial/ethnic composition of STEM completers in the community college sample does not depend on whether we rely on racial/ethnic data in the prediction model, with the exception of Asian students, whose predicted STEM completion rate declines slightly when race-ethnicity information is omitted. This result is consistent with the pattern of estimates in Table 3, which shows that outside of Asian students, the racial/ethnic designations are not important predictors of STEM completion conditional on measured academic preparation.[12] Second, in terms of gender diversity, the female shares jump markedly in columns (5) and (6) if we remove all information about gender from the predictive model. This is because female students have generally strong academic qualifications (much stronger than for male students, on average). If we ignore their preference for non-STEM fields in our prediction models, we predict many more female students would pursue and complete STEM degrees. Again, this result is not surprising based on the estimates in Table 3, and only informative if one believes that female students who attend two-year colleges have fundamentally different preferences for STEM education than their four-year counterparts, which we view as unlikely.

6.2   Modifications to the Sample Restrictions

In Section 2 we describe a number of restrictions that we impose on our preferred estimation sample. In this section we examine the substantive implications of these restrictions for our findings.

First, in Table 8 we relax the credit-hour and age-based restrictions, which are set to 12 first-semester credits and age ≤ 20 in the main analysis. We consider reductions of the credit-hour constraint to 9 and then 6 credit hours to accommodate part-time students, and raise the

---

[12] That said, the evidence in Table 7 is more comprehensive because Table 3 does not show all of the interaction coefficients.

19

maximum entry age to 22 and then 24. These changes results in small increases in the numbers

of students who are identified as STEM qualified, and correspondingly, small increases in the

number of new STEM degrees produced. For example, relaxing the credit-hours restriction from

12 all the way to 6 credit hours results in a total increase in STEM degree production of just 73

degrees; raising the entry age threshold from ≤ 20 to ≤ 24 adds just 14 degrees. The changes are

small because relatively few part-time or older community college students are STEM qualified

based on their academic profiles. And if anything, part-time and older community college

students are even more likely to have unobservable characteristics that make them less likely to

complete a STEM degree at a university than their full-time, younger peers.[13]

Next we revisit to our decision to drop students who did not take the ACT prior to college

enrollment. Again, the rationale for this decision is that we view the act of taking the ACT as an

observable indicator of interest in and/or aptitude for postsecondary education. However, it is

possible that some well-prepared community college students elect not to take the ACT for other

reasons and to the extent this is true, our exclusion restriction with respect to ACT scores would

lead to an understatement of the potential STEM pipeline in community colleges.

To test this, we bring community college students with missing ACT scores back into the

analysis by imputing their ACT scores using their first-semester completed credit hours and

GPAs. The imputation coefficients are obtained from a regression of each ACT score (math and

English) on first-semester completed credit hours and GPAs among community college students

with all available information. With these students added back into the sample, we replicate our

---

[13] Note that some or perhaps all of these unobservables will be unrelated to competency, but rather derived from circumstances. As just one example, older and part-time students are more likely to have more non-schooling commitments that would make it harder for them to attend universities, which might require moving and will typically have less flexible degree programs.

entire predictive procedure. Table 9 shows the results compared to the results using our main settings.[14]

Incorporating students with missing ACT scores into the community college sample leads to a substantial increase in its size—the sample increases from 43,214 students (Table 1) to 57,382 students, a 33 percent increase. However, Table 9 shows that this translates into only a very small increase in the sample of students identified as STEM qualified and who complete STEM degrees. Indeed, the number of STEM degrees produced increases by just 115 degrees relative to baseline. The reason is that most community college students have academic credentials that put them below the threshold for STEM qualified as we've defined it, and those without ACT scores are strongly negatively selected (i.e., when we impute their ACT scores, the imputed values are low). Consistent with the spirit of our initial exclusion of students who do not take the ACT, adding these students back into our sample has a negligible effect on our findings.

6.3 Sensitivity to the Nudge Threshold

Next we revisit our decision to set the intervention threshold at the median predicted likelihood of success among initial STEM entrants in the university sample. This is a reasonable but arbitrary threshold. As we lower the threshold we will identify more students as "STEM qualified," but the STEM success rate will fall because the marginally-induced students will be less academically prepared. Alternatively, as we raise the threshold the likelihood of obtaining a STEM degree conditional on being identified as "STEM qualified" will rise, but fewer students will be identified so total degree production will fall.

In Table 10 we show the sensitivity of our estimates to moving the nudge threshold between the 40th and 60th percentiles of the distribution, at 5-unit intervals. When we decrease

---

[14] Note that like with the imputed class ranks, we inflate the variance of students' imputed ACT scores to match the variance of observed scores.

the intervention threshold substantially—i.e., down to the 40th percentile—we increase the number of identified students by 57 percent to just over 5000. We also increase the number of STEM degrees produced, although not commensurately because the average student is less prepared—the number of degrees produced increases by 31 percent, or 273 degrees. If we move the threshold up to the 60th percentile we produce 245 fewer degrees in total (624 versus 869), but also nudge 1265 fewer students (1,944 versus 3,209). The conversion rate of enrollment to STEM degrees increases from 27.1 percent (869/3,209) in the baseline case to 32.1 percent (624/1,944) using the 60th-percentile threshold, but at the expense of reduced total production.

Table 10 also shows modest diversity implication of modifications to the intervention threshold. Although the implied diversity effects are substantively similar at all thresholds, at lower thresholds the population of students identified as STEM-qualified, and the population predicted to earn STEM degrees, is slightly more likely to be female and Black (relative to male and White).

Overall, Table 10 illustrates the tradeoffs as the intervention threshold varies in terms of total STEM degrees produced, the degree-conversion rate, and to a lesser extent the diversity of students who receive STEM degrees. Determining the appropriate threshold requires a normative judgement about the value of degrees produced and the cost of failed interventions (i.e., students who do not complete a STEM degree) and we do not make this judgment here. That said, in assessing the cost of failures, an important distinction exists between failure in STEM and failure to complete any degree at a university. As a point of information, supplementary predictive models—for which the results are reported in Appendix Table A6—indicate that among students who we intervene with but who fail to complete a STEM degree, just under half (49 percent) would be expected to complete a non-STEM university degree based on their observable

22

characteristics and qualifications, and the other half would fail to complete a degree within 6 years. Among these students, available evidence from Goodman, Hurwitz, and Smith (2017) and Mountjoy (2019) suggests that bachelor's degree receipt would be much higher than in the absence of our intervention; moreover, Mountjoy (2019) shows that the diversion of marginal students toward four-year colleges (or similarly, away from two-year colleges) leads to an increase in earnings.

6.4 Excluding Biology

Biology is one of the largest STEM majors in Missouri and nationally (Snyder, de Brey and Dillow, 2019). However, the field of biology differs from other STEM fields in that it is less mathematically oriented and biology degrees have lower earnings returns.[15] The lower labor-market returns imply that compared to other STEM majors, market demand for biology degrees is low. Therefore, it may be appropriate for policies designed to increase STEM degree production to focus on fields outside of biology.

In Table 11 we modify our analysis to examine STEM degree production outside of majors in biology; specifically, outside of majors under the 2-digit CIP code classification for biology. We exclude these majors from the definition of STEM fields, and replicate our entire analytic procedure. In the university sample for the cohorts we study, majors under the biology heading account for 29 percent of all STEM majors, or 2,619 degrees.

Table 11 shows that excluding biology, our hypothetical intervention is predicted to generate 705 STEM degrees, versus 869 STEM degrees in the analysis inclusive of biology. The percent increase in non-biology STEM degrees at universities is similar to, but somewhat higher than, in the base case, at 10.9 percent (705/6441). The results for racial/ethnic and gender

---

[15] For example, Webber (2016) shows that the earnings returns to biology degrees are more closely aligned with the returns to degrees in arts and humanities fields than they are with earnings in other STEM disciplines.

diversity are substantively similar to what we find in the base case in that diversity conditions worsen in the community college sample. An especially sharp decline is apparent in the female share of STEM degrees produced—with biology included, 14 percent of STEM degrees predicted to result from our intervention are female (Table 5), versus just 7 percent when we exclude biology (Table 11). This is partly because female students are less likely to enroll in and complete STEM degrees in non-biology fields, which can be seen by comparing column (3) in Table 11 to column (2) in Table 5, but this does not explain the full decline.

7. Major Extensions

In this section we assess the implications of relaxing the two major assumptions that drive the upper-bound interpretation of our findings thus far: (1) selection-on-observables into universities, and (2) the perfect efficacy of our enrollment intervention.

7.1   Selection on Unobservables

Thus far we have maintained the unrealistic assumption of selection on observables into college sector. This assumption is imbedded in our predictions because we assume that students who choose to initially enroll in community colleges, if they were shifted to start at a university instead, would perform just as well as observationally similar students who choose to enroll in universities on their own. However, the fact that the community college students did not choose a university on their own suggests that this assumption is unlikely to hold. To the extent that it is violated the expected direction of bias in our estimates of STEM degree production will be positive—i.e., we will overstate the likelihood of STEM success among our hypothetically nudged students.

Although we expect unobserved selection to be non-zero, we are not aware of any research that we can draw on to parameterize a precise value for its magnitude in our context.

Absent this, we perform a bounding exercise to assess how varying degrees of unobserved selection would impact our findings. Our procedure follows the logic of Rosenbaum (2002). We begin by estimating the magnitude of *selection on observables* between two-year and four-year students who we identify as STEM qualified. Although we use a fixed threshold to identify STEM-qualified students, Table 4 shows that on average the community college sample is negatively selected based on observables. Again, this is because the distribution of academic qualifications at community colleges is shifted to the left of the four-year distribution. The difference in observed selection between students in the two-year and four-year samples can be summarized by the average difference in the likelihood of STEM success between the groups, represented by $(\bar{\hat{p}}_i^{cc} - \bar{\hat{p}}_i)$, where $\bar{\hat{p}}_i^{cc}$ and $\bar{p}_i$ are the average predicted STEM degree completion rates for two-year and four-year students, respectively, obtained using the parameters from equation (1). This calculation indicates that STEM-qualified community college students are 4 percentage points less likely to complete a STEM degree than their STEM-qualified peers who start at universities.

Next, we assume that selection into college sector on unobservables is in the same direction as observed selection and consider magnitudes of unobserved selection ranging from 50 to 300 percent as large as observed selection. To give a sense of the meaning of these values, at the high-end scenario with unobserved selection that is 300% as large as observed selection, we parameterize outcomes such that nudged community college students are 12 percentage points less likely to complete a STEM degree from a university than is implied by their observable student profiles alone. For this exercise we do not allow unobserved factors to affect who is nudged—a realistic policy could act only on observable information—but they will affect the degree production rate among the nudged sample.

Table 12 shows the results from the various selection-on-unobservables scenarios. The first row replicates our baseline condition with no unobserved selection. We nudge 3,209 students and 27 percent of these students are predicted to complete a STEM degree. As unobserved selection becomes more severe the conversion rate, and total degree production, decline. Under the assumption that selection on unobservables is of the same magnitude as observed selection (the 100% scenario in row 3 of Table 12), the number of STEM degrees produced in our nudged sample declines by 141 degrees, to 738. In the largest selection-on-unobservables condition we consider (at 300% of observed selection) degree production is reduced 45 percent, to just 481 STEM degrees.

We do not have the means to assess the true magnitude of unobserved selection in our data, so we can only provide the range of estimates in Table 12 to provide insight into how unobserved selection may impact our findings. If unobserved selection is assumed to be small the implications are modest; if unobserved selection is large, which cannot be ruled out given our context, it would imply a substantially reduced potential to expand four-year degree production in STEM by shifting community college enrollment.

7.2   A Less than Perfect Nudge

The other important assumption imbedded in our analysis up to this point is that all of the students who we intervene with respond as intended—that is, we can shift enrollment of community college students to universities with perfect efficacy. This is a useful assumption for thinking about the potential upper-bound effect of tapping into the community college population to increase four-year degree production in STEM, but it is not realistic.

If our intervention were a textbook nudge—which Thaler and Sunstein (2008) define as an intervention that is cheap and easy to avoid—evidence suggests that the behavioral response

would likely be very small, in the range of 0-5 percentage points (Barr and Turner, 2018; Castleman and Page, 2015; Gurantz et al, 2020; Oreopoulos and Petronijevic, 2019). The response rate could in principle be increased if the nudge became more expensive and costly to avoid, such as if tuition subsidies or stipends were offered, but even then available research suggests a modest behavioral response. For example, Deming and Walters (2017) find very modest impacts of tuition changes on enrollment, whereas they find much larger impacts of institutional spending changes. Marx and Turner (2019) find community college students who have more access to borrowing are around 4 percentage points more likely to transfer to a four-year college.

Based on these studies, in Table 13 we consider scenarios where we rescale the intervention effect to be more realistic, but we believe still optimistic, in terms of affecting community college students' enrollment decisions: a 5-percent effect and a 10-percent effect. The idea is that we would still nudge the same baseline pool of STEM-qualified students (i.e., the 3,209 students from Table 4), but only 5 or 10 percent would change their behavior and enroll in a university. For each scenario considered in Table 13, we show results for two cases: one where the students who change behavior are a random sample of the STEM-qualified group and one where the students who change are those most likely to succeed in a four-year STEM program.

Unsurprisingly, the results in Table 13 imply large reductions in degrees produced and no positive changes in diversity (moreover, gender diversity declines when the students who are most likely to succeed respond to the nudge, but this is tautological because our prediction model embodies the fact that female students are much less likely to succeed in STEM). The extensive-

margin effects in these more realistic nudge scenarios are small in absolute terms and as a percentage of STEM degrees produced at universities.

8. Conclusion

We assess the prospects for expanding and diversifying STEM degree production in universities by tapping into the population of academically-qualified community college students. Our work complements a large and growing literature that examines specific policy interventions by providing upper-bound estimates of the types of changes to the STEM pipeline that would be possible via interventions targeted toward community college students.

We find that the number of STEM degrees produced by universities can be modestly expanded by tapping into the community college student population. The exact magnitude of the change depends on assumptions and policy-design details along a variety of dimensions. In our baseline evaluation scenario, we estimate that the academically-qualified community college students who we would nudge with our hypothetical intervention would generate a gross increase of 869 four-year STEM degrees, or an increase of 9.6 percent on the number STEM degrees already produced by universities. The net increase is smaller—about 6.4 percent—because some of the students who we would hypothetically nudge go on to earn a STEM degree from a university regardless (i.e., by transferring to a university and completing a STEM degree).

Our extensive-margin estimate can be made somewhat larger by modifying aspects of our hypothetical intervention, most notably by lowering the threshold we use to identify the "STEM qualified" students who are nudged, but even a fairly large reduction in the threshold results in tempered gains in STEM degrees. More likely, though, is that our estimates are far above what could be feasibly achieved through a real policy because of the variety of upper-bound conditions we impose on the analysis. The two most significant upper-bound conditions are

28

assumptions: our baseline estimates assume (a) no selection into college sector on unobservables and (b) that we could implement our enrollment intervention with perfect compliance. Relaxing these assumptions quickly degrades the magnitude of gains in STEM degrees we could hope to produce.

Our findings for diversifying STEM degree production are even less promising. Community college students are more racial/ethnically diverse than their four-year counterparts overall. However, the fraction of non-White students who are academically qualified to succeed in four-year STEM degree programs among community colleges is lower than among four-year college students. The end result is that the diversity of individuals who are predicted to earn STEM degrees among our nudged sample of community college students is less than among university students already earning STEM degrees. We also find no scope for increasing the gender diversity of STEM degree production by tapping into the community college population. This result is partly tautological because we assume that female aversion to STEM, conditional on observable academic qualifications, is similar among two-year and four-year college students. However, the gender gap among community college students is also exacerbated because within their gender-specific distributions of academic qualifications, female students who attend community colleges are more negatively selected than their male peers. This compounds the gender gap in predicted STEM attainment among community college students.

The broad takeaway from our analysis is that policies and interventions targeted toward community college students are unlikely to alter macro-level features of STEM degree production at universities. This does not mean that interventions cannot be effective at the micro-level in terms of improving outcomes for individually-impacted students, and indeed there is evidence that at least for students at the margin of having appropriate academic qualifications,

shifts in enrollment from two-year to four-year colleges are beneficial (Goodman, Hurwitz, and Smith, 2017; Mountjoy, 2019). However, our analysis shows that the potential for intervening with community college students in a way that meaningfully impacts overall STEM degree production at universities is limited.

References

Altonji, J. , Blom, E. , & Meghir, C. (2012). Heterogeneity in human capital investments: high school curriculum, college major, and careers. Annual Review of Economics, 4 , 185– 223 .

Altonji, J. G., Elder, T. E., & Taber, C. R. (2005). Selection on observed and unobserved variables: Assessing the effectiveness of Catholic schools. *Journal of Political Economy*, *113*(1), 151–184.

Anderson, E., & Kim, D. (2006). Increasing the Success of Minority Students in Science and Technology. Washington, DC: American Council on Education

Arcidiacono, P., Aucejo, E. M., & Hotz, V. J. (2016). University differences in the graduation of minorities in STEM fields: Evidence from California. *American Economic Review*, 106(3), 525-62.

Arcidiacono, P., Aucejo, E.M., & Spenner, K. (2012). What happens after enrollment? An analysis of the time path of racial differences in GPA and major choice. *IZA Journal of Labor Economics* 1(5).

Arcidiacono, P., & Koedel, C. (2014). Race and college success: Evidence from Missouri. *American Economic Journal: Applied Economics*, *6*(3), 20–57.

Atkinson, R.D., Mayo, M., 2010. Refueling the U.S. Innovation Economy: Fresh Approaches to Science, Technology, Engineering and Mathematics (STEM) Education. *Information Technology and Innovation Foundation.*

Atkinson, R.D., & Ezell, S.J. (2012). *Innovation Economics: The Race for Global Advantage*. New Haven, CT: Yale University Press.

Bahr, P. R., Jackson, G., McNaughtan, J., Oster, M., & Gross, J. (2017). Unrealized Potential: Community College Pathways to STEM Baccalaureate Degrees. *Journal of Higher Education*, *88*(3), 430–478.

Baker, R. (2016). The effects of structured transfer pathways in community colleges. *Educational Evaluation and Policy Analysis*, *38*(4), 626-646.

Barr, A., & Turner, S. (2018). A letter and encouragement: Does information increase postsecondary enrollment of UI recipients? *American Economic Journal: Economic Policy*, *10*(3), 42–68.

Bird, K. A., Castleman, B. L., Denning, J. T., Goodman, J., Lamberton, C., & Rosinger, K. O. (2019). *Nudging at scale: Experimental evidence from FAFSA completion campaigns* (No. w26158). National Bureau of Economic Research.

Carrell, S. E., & Sacerdote, B. (2013). Late interventions matter too: The case of college coaching in New Hampshire (No. w19031). Cambridge, MA: National Bureau of Economic Research

Castleman, B. L., & Page, L. C. (2015). Summer nudging: Can personalized text messages and peer mentor outreach increase college going among low-income high school graduates? *Journal of Economic Behavior and Organization*, *115*, 144–160.

Castleman, B. L., & Page, L. C. (2016). Freshman year financial aid nudges: An experiment to increase FAFSA renewal and college persistence. *Journal of Human Resources*, *51*(2), 389–415.

Chen, X. (2009). Students Who Study Science, Technology, Engineering, and Mathematics (STEM) in Postsecondary Education. Stats in Brief. NCES 2009-161. *National Center for Education Statistics*.

Chetty, R., Friedman, J. N., Saez, E., Turner, N., & Yagan, D. (2020). Income Segregation and Intergenerational Mobility Across Colleges in the United States. *The Quarterly Journal of Economics*, *135*(3), 1567–1633.

Coleman, M.S., Smith, T.L., & Miller, E.R. (2019). Catalysts for achieving sustained improvements in the quality of undergraduate STEM education. *Daedalus, the Journal of the American Academy of Arts & Sciences* 148(4), 29-46.

Committee on Prospering in the Global Economy of the 21st Century (2007). *Rising above the gathering storm: Energizing and employing America for a brighter economic future*. Washington DC: The National Academies Press.

Committee on Science, Engineering, and Public Policy (2011). Expanding underrepresented minority participation: America's science and technology talent at the crossroads. Washington, DC: The National Academies Press.

Darolia, R., Koedel, C., Main, J., Ndashimye, F., & Yan, J. (2020). High School Course Access and Postsecondary STEM Enrollment and Attainment. *Educational Evaluation and Policy Analysis.* 42(1), 22-45.

DellaVigna, S., & Linos, E. (2020). *RCTs to Scale: Comprehensive Evidence from Two Nudge Units*. Working Paper, UC Berkeley.

Deming, D.J., Goldin, C., & Katz, L.F. (2012). The for-profit postsecondary school sector: Nimble critters or agile predators? *Journal of Economic Perspectives* 26(1), 139-164.

Deming, D.J., & Walters, C.R. (2017). The impact of price caps and spending cuts on U.S postsecondary attainment. NBER Working Paper No. 23736.

Evans, C. A., Chen, R., & Hudes, R. P. (2020). Understanding Determinants for STEM Major Choice Among Students Beginning Community College. *Community College Review*, 48(3), 227-251.

Fayer, S., Lacey, A., & Watson, A. (2017). *STEM occupations: Past, present, and future*. Washington, DC: U.S. Bureau of Labor Statistics.

Goodman, J., Hurwitz, M., & Smith, J. (2017). Access to four-year public college and degree completion. *Journal of Labor Economics* 35(3), 829-867.

Gross, B., & Goldhaber, D. (2009). Community College Transfer and Articulation Policies: Looking Beneath the Surface. CRPE Working Paper 2009-1. *Center on Reinventing Public Education*.

Gurantz, O., Howell, J., Hurwitz, M., Larson, C., Pender, M., & White, B. (forthcoming). A national-level informational experiment to promote enrollment in selective colleges. *Journal of Policy Analysis and Management.*

Griffith, A. L. (2010). Persistence of women and minorities in STEM field majors: Is it the school that matters? *Economics of Education Review*, *29*(6), 911–922.

Hagedorn, L. S., & Purnamasari, A. V. (2012). A Realistic Look at STEM and the Role of Community Colleges. *Community College Review*, *40*(2), 145–164.

Hill, A. J. (2017). State affirmative action bans and STEM degree completions. *Economics of Education Review*, *57*, 31–40.

Horn, L., & Skomsvold, P. (2012). *Community College Student Outcomes: 1994-2009. National Center for Education Statistics*.

Hoxby, C., & Avery, C. (2013). The missing "one-offs": The hidden supply of high-achieving, low-income students. *Brookings Papers on Economic Activity*, *65*(SPRING 2013), 1–50.

Hyman, J. (2020). Can Light-Touch College-Going Interventions Make a Difference? Evidence from a Statewide Experiment in Michigan. *Journal of Policy Analysis and Management*, *39*(1), 159–190.

Kahn, S., & Ginther, D. (2017). *Women and STEM* (No. w23525). National Bureau of Economic Research.

Kerr, S. P., Pekkarinen, T., Sarvimäki, M., & Uusitalo, R. (2020). Post-secondary education and information on labor market prospects: A randomized field experiment. *Labour Economics*, *66*, 101888.

Kinsler, J., & Pavan, R. (2015). The specificity of general human capital: Evidence from college major choice. *Journal of Labor Economics*, *33*(4), 933–972.

Kirkeboen, L. J., Leuven, E., & Mogstad, M. (2016). Field of study, earnings, and self-selection. *The Quarterly Journal of Economics*, *131*(3), 1057-1111.

Mountjoy, J. (2019). Community colleges and upward mobility. Available at SSRN 3373801.

Marx, B. M., & Turner, L. J. (2019). Student loan nudges: Experimental evidence on borrowing and educational attainment. *American Economic Journal: Economic Policy 11*(2), 108–141.

National Academy of Sciences, National Academy of Engineering, & Institute of Medicine of the National Academies. (2007). *Rising Above the Gathering Storm: Energizing and Employing America for a Brighter Economic Future*. Washington DC: National Academies Press.

National Research Council (2011). *Successful K-12 STEM education: Identifying effective approaches in science, technology, engineering, and mathematics*. Washington, DC: The National Academies Press.

National Science & Technology Council (2018). Charting a Course for Success: America's Strategy for STEM Education. Washington DC: Office of the President of the United States.

Oreopoulos, P., & Petronijevic, U. (2019). *The remarkable unresponsiveness of college students to nudging and what we can learn from it* (No. w26059). National Bureau of Economic Research.

Page, L. C., Castleman, B., & Sahadewo, G. A. (2016). More than Dollars for Scholars: The Impact of the Dell Scholars Program on College Access, Persistence and Degree Attainment. *SSRN Electronic Journal*, 1–75.

Page, L. C., Kehoe, S. S., Castleman, B. L., & Sahadewo, G. A. (2019). More than Dollars for Scholars The Impact of the Dell Scholars Program on College Access, Persistence, and Degree Attainment. *Journal of Human Resources*, *54*(3), 683-725.

Provasnik, S., & Planty, M. (2008). Community Colleges: Special Supplement to The Condition of Education 2008. Statistical Analysis Report. NCES 2008-033. *National Center for Education Statistics*.

Roksa, J., & Keith, B. (2008). Credits, time, and attainment: Articulation policies and success after transfer. *Educational Evaluation and Policy Analysis*, *30*(3), 236-254.

Rosenbaum, P. R. (2002). Overt bias in observational studies. In *Observational studies* (pp. 71-104). Springer, New York, NY.

Sass, T. R. (2015). Understanding the STEM Pipeline. Working Paper 125. *National Center for Analysis of Longitudinal Data in Education Research (CALDER)*.

Shambough, J. Nunn, R., & Portman, B. (2017). Eleven Facts about Innovation and Patents. Policy Report. Washington DC: The Hamilton Project at Brookings.

Shapiro, D., Dundar, A., Huie, F., Wakhungu, P. K., Yuan, X., Nathan, A., & Hwang, Y. (2017). Tracking Transfer: Measures of Effectiveness in Helping Community College Students to Complete Bachelor's Degrees.(Signature Report No. 13). *National Student Clearinghouse*.

Snyder, T.D., de Brey, C., & Dillow, S.A. (2019) *Digest of Education Statistics 2018*. Washington, DC: National Center for Education Statistics.

Stinebrickner, R. & Stinebrickner, T.R. (2014). A major in science? Initial beliefs and final outcomes for colleg major and dropout. *Review of Economic Studies* 81(1), 426-472.

Terenzini, P., Lattuca, L., Ro, H. K., & Knight, D. (2014). America's Overlooked Engineers: Community Colleges and Diversity in Engineering Education.

Thaler, R. H., & Sunstein, C.R. (2008). *Nudge: Improving Decisions About Health, Wealth, and Happiness*. Yale University Press.

Wang, X. (2013). Modeling Entrance into STEM Fields of Study Among Students Beginning at Community Colleges and Four-Year Institutions. *Research in Higher Education*, *54*(6), 664–692.

Wang, X. (2015). Pathway to a baccalaureate in STEM fields: Are community colleges a viable route and does early STEM momentum matter? *Educational Evaluation and Policy Analysis*, *37*(3), 376–393.

Webber, D. A. (2016). Are college costs worth it? How ability, major, and debt affect the returns to schooling. *Economics of Education Review*, *53*, 296–310.

Westrick, P., Radunzel, J., & Bassiri, D. (2018). Leaning into STEM: Predicting STEM Major Choice Using Measures of Academic Tilt and Measured Interest Tilt. ACT Research & Policy. Technical Brief. *ACT, Inc.*, (October), 1–16.

White House (2016). STEM for all. [Blog post]. Retrieved from https://obamawhitehouse.archives.gov/blog/2016/02/11/stem-all

Winters, J. V. (2014). STEM graduates, human capital externalities, and wages in the US. *Regional Science and Urban Economics*, *48*, 190-198.

Table 1: Summary statistics for two-year and four-year college entrants overall and for key subsamples.

	Four Year University			Community College		
	(1)	(2)	(3)	(4)	(5)	(6)
	All	In state	Analytic Sample	All	In state	Analytic Sample
ACT math	22.84	22.7	22.89	18.84	18.84	19.04
	(4.88)	(4.87)	(4.78)	(3.85)	(3.85)	(3.78)
ACT English	23.62	23.49	23.68	18.83	18.83	18.99
	(5.45)	(5.46)	(5.34)	(4.95)	(4.95)	(4.78)
High school percentile rank/100	0.69	0.69	0.71	0.49	0.49	0.56
	(0.23)	(0.23)	(0.22)	(0.25)	(0.25)	(0.23)
High school percentile rank missing indicator	0.16	0.16	0.12	0.55	0.55	0.35
	(0.37)	(0.36)	(0.32)	(0.5)	(0.5)	(0.48)
Female	0.55	0.55	0.55	0.54	0.54	0.54
	(0.50)	(0.5)	(0.50)	(0.50)	(0.5)	(0.50)
White	0.77	0.79	0.81	0.71	0.71	0.79
	(0.42)	(0.41)	(0.39)	(0.45)	(0.45)	(0.41)
Black	0.12	0.12	0.10	0.14	0.13	0.08
	(0.32)	(0.32)	(0.30)	(0.34)	(0.34)	(0.27)
Hispanic	0.02	0.02	0.02	0.03	0.03	0.02
	(0.15)	(0.15)	(0.14)	(0.17)	(0.16)	(0.15)
Asian	0.02	0.02	0.02	0.01	0.01	0.01
	(0.14)	(0.14)	(0.14)	(0.12)	(0.12)	(0.11)
Other Race	0.03	0.02	0.01	0.03	0.03	0.02
	(0.17)	(0.14)	(0.11)	(0.16)	(0.16)	(0.14)
Race missing unknown	0.04	0.04	0.04	0.09	0.09	0.07
	(0.2)	(0.19)	(0.19)	(0.28)	(0.28)	(0.26)
Number of observations	97749	83263	70737	110695	108198	43214

Notes: Table shows means and standard deviations (in parenthesis) for university students and community college students. See the text and Appendix table A1 for details about the construction of the analytic sample.

Table 2: Summary statistics for four-year college entrants in the analytic sample by STEM entry and exit conditions.

	Four Year University			
	(1)	(2)	(3)	(4)
	Analytic Sample	STEM entrants	non-STEM entrants	STEM completers
ACT math	22.89	25.58	22.16	26.63
	(4.78)	(4.76)	(4.52)	(4.54)
ACT English	23.68	25.19	23.27	26.05
	(5.34)	(5.13)	(5.32)	(5.01)
High school percentile rank/100	0.71	0.77	0.69	0.82
	(0.22)	(0.2)	(0.23)	(0.17)
High school percentile rank missing indicator	0.12	0.10	0.12	0.12
	(0.32)	(0.30)	(0.32)	(0.32)
Female	0.55	0.36	0.60	0.36
	(0.5)	(0.48)	(0.49)	(0.48)
White	0.81	0.82	0.80	0.86
	(0.39)	(0.39)	(0.4)	(0.35)
Black	0.10	0.08	0.11	0.04
	(0.3)	(0.27)	(0.31)	(0.20)
Hispanic	0.02	0.02	0.02	0.02
	(0.14)	(0.15)	(0.14)	(0.13)
Asian	0.02	0.03	0.02	0.03
	(0.14)	(0.17)	(0.13)	(0.18)
Other Race	0.01	0.02	0.01	0.01
	(0.11)	(0.12)	(0.11)	(0.11)
Race missing unknown	0.04	0.04	0.04	0.04
	(0.19)	(0.19)	(0.19)	(0.19)
Graduate with STEM in 6 years	0.13	0.44	0.04	1.0
	(0.33)	(0.5)	(0.2)	(0)
Number of observations	70737	15125	55612	9060

Notes: Table shows means and standard deviations (in parenthesis).

Table 3: Results from predictive logistic regression of STEM degree completion among four-year college entrants.

	Graduate with STEM
ACT math	0.163***
	[0.153,0.173]
ACT English	-0.041***
	[-0.05,-0.032]
Female	-1.253***
	[-1.599,-0.900]
Asian	2.467***
	[1.546,3.331]
Black	-0.356
	[-1.191,0.335]
Hispanic	0.609
	[-0.709,1.661]
Other Race	1.071
	[-0.426,2.488]
Race Missing Unknown	0.667
	[-0.211,1.438]
High school percentile rank (/100)	2.783***
	[2.589,2.995]
High school percentile rank missing indicator	0.174**
	[0.035,0.318]
Number of observations	68798
	[68432,69114]
High School FE	X
Cohort FE	X
ACT*Race/Ethnicity interactions	X
ACT*Gender interactions	X
Percentile Rank*Race/Ethnicity	X

Notes: The regression output corresponds to equation (1) in the main text. Bootstrapped mean estimates and 95 percent confidence intervals are reported. ***$p<0.01$, **$p<0.05$, *$p<0.10$.

Table 4: Summary statistics by STEM qualified status at two-year and four-year colleges.

	Four Year University		Community College	
	(1)	(2)	(1)	(2)
	STEM qualified	Not STEM qualified	STEM qualified	Not STEM qualified
ACT math	28.01	21.08	25.26	18.55
	[27.9,28.11]	[21.02,21.13]	[24.98,25.54]	[18.5,18.59]
ACT English	26.95	22.52	22.42	18.71
	[26.82,27.09]	[22.45,22.58]	[22.11,22.74]	[18.66,18.77]
High school percentile rank (/100)	0.86	0.66	0.79	0.53
	[0.86,0.87]	[0.65,0.66]	[0.77,0.8]	[0.52,0.53]
Female	0.28	0.64	0.16	0.57
	[0.27,0.3]	[0.64,0.65]	[0.13,0.19]	[0.57,0.58]
White	0.87	0.79	0.81	0.79
	[0.86,0.87]	[0.78,0.79]	[0.77,0.85]	[0.78,0.79]
Black	0.03	0.13	0.02	0.09
	[0.02,0.03]	[0.13,0.13]	[0.01,0.02]	[0.08,0.09]
Hispanic	0.02	0.02	0.02	0.02
	[0.01,0.02]	[0.02,0.02]	[0.01,0.03]	[0.02,0.02]
Asian	0.04	0.01	0.04	0.01
	[0.04,0.05]	[0.01,0.01]	[0.02,0.05]	[0.01,0.01]
Other Race	0.01	0.01	0.02	0.02
	[0.01,0.02]	[0.01,0.02]	[0.01,0.04]	[0.02,0.02]
Race missing unknown	0.04	0.04	0.09	0.07
	[0.03,0.04]	[0.03,0.04]	[0.06,0.13]	[0.07,0.08]
Number of students	18509	52228	3209	40005
	[18088,18975]	[51762,52649]	[2907,3520]	[39694,40307]

Notes: Table shows means and 95 percent bootstrapped confidence intervals (500 repetitions) for university students and community college students by STEM qualified status.

Table 5: Summary statistics for community college students who are predicted to complete STEM degrees at four-year colleges compared to observed STEM completers at four-year colleges.

	(1) Graduate with STEM	(2) STEM completers at universities (from Table 2)
Avg ACT math	25.76	26.63
	[25.45,26.07]	[26.53,26.71]
Avg ACT English	22.64	26.05
	[22.3,22.96]	[25.95,26.14]
Avg HS percentile rank (/100)	0.81	0.82
	[0.79,0.82]	[0.82,0.83]
Share Female	0.14	0.36
	[0.11,0.17]	[0.35,0.37]
Share White	0.81	0.86
	[0.77,0.85]	[0.85,0.86]
Share Black	0.01	0.04
	[0.01,0.02]	[0.04,0.05]
Share Hispanic	0.02	0.02
	[0.01,0.03]	[0.02,0.02]
Share Asian	0.04	0.03
	[0.02,0.05]	[0.03,0.04]
Share Other Race	0.02	0.01
	[0.01,0.04]	[0.01,0.02]
Share Race missing unknown	0.1	0.04
	[0.06,0.14]	[0.03,0.04]
Number of STEM degrees (gross)	869	9060
	[778,965]	[8882,9212]
Number of STEM degrees (net)	580	--
	[498,668]	

Notes: Table reports averages and 95 percent confidence intervals over 500 bootstrap repetitions for nudged community college students. Column (2) reports means and standard deviations for actual STEM completers among initial university entrants. The net number of STEM degrees equals subtracting the number of STEM-qualified community college students observed transferring into university and obtaining a STEM degree within 6 years directly in the data from the number of gross STEM degrees.

Table 6: Robustness of findings to selecting STEM-qualified community college students without the variance inflation adjustment to the imputed high-school class percentile ranks.

	Main Settings		Without imputed-HS Rank variance inflation	
	(1)	(2)	(3)	(4)
	STEM qualified	Graduate with STEM	STEM qualified	Graduate with STEM
Avg ACT math	25.26	25.76	25.46	25.94
	[24.98,25.54]	[25.45,26.07]	[25.2,25.73]	[25.64,26.24]
Avg ACT English	22.42	22.64	22.52	22.7
	[22.11,22.74]	[22.3,22.95]	[22.26,22.79]	[22.41,22.99]
Avg HS percentile rank (/100)	0.79	0.81	0.76	0.78
	[0.77,0.8]	[0.79,0.82]	[0.75,0.78]	[0.76,0.79]
Share Female	0.16	0.14	0.15	0.13
	[0.13,0.19]	[0.11,0.17]	[0.13,0.18]	[0.11,0.16]
Share White	0.81	0.81	0.81	0.81
	[0.77,0.85]	[0.77,0.85]	[0.77,0.85]	[0.77,0.85]
Share Black	0.02	0.01	0.01	0.01
	[0.01,0.02]	[0.01,0.02]	[0.01,0.02]	[0.01,0.02]
Share Hispanic	0.02	0.02	0.02	0.02
	[0.01,0.03]	[0.01,0.03]	[0.01,0.03]	[0.01,0.03]
Share Asian	0.04	0.04	0.04	0.04
	[0.02,0.05]	[0.02,0.05]	[0.03,0.05]	[0.03,0.05]
Share Other Race	0.02	0.02	0.02	0.03
	[0.01,0.04]	[0.01,0.04]	[0.01,0.04]	[0.01,0.04]
Share Race missing unknown	0.09	0.1	0.09	0.1
	[0.06,0.13]	[0.06,0.14]	[0.06,0.13]	[0.06,0.14]
Number of students or degrees (gross)	3209	869	2940	789
	[2907,3520]	[778,965]	[2670,3223]	[712,869]

Notes: Table reports averages and 95 percent confidence intervals over 500 bootstrap repetitions for STEM-qualified community college students when we inflate the variance of imputed high school class percentile ranks in column (1) and (2), and do not inflate the variance of imputed high school percentile ranks in column (3) and (4).

Table 7: Robustness of findings to dropping race-gender indicators and/or race-gender indicator interactions in the model that predicts STEM four-year degree completion.

	Main Settings		No Race-Gender Interaction Terms		No Race-Gender Indicators or Interactions	
	(1)	(2)	(3)	(4)	(5)	(6)
	STEM qualified	Graduate with STEM	STEM qualified	Graduate with STEM	STEM qualified	Graduate with STEM
Avg ACT math	25.26	25.76	25.46	25.98	25.86	26.37
	[24.98,25.54]	[25.45,26.07]	[25.2,25.7]	[25.71,26.23]	[25.59,26.11]	[26.07,26.63]
Avg ACT English	22.42	22.64	22.74	22.99	22.94	23.16
	[22.11,22.74]	[22.3,22.95]	[22.52,22.98]	[22.75,23.23]	[22.71,23.17]	[22.92,23.41]
Avg HS percentile rank (/100)	0.79	0.81	0.80	0.82	0.81	0.83
	[0.77,0.8]	[0.79,0.82]	[0.79,0.81]	[0.81,0.83]	[0.8,0.82]	[0.81,0.84]
Share Female	0.16	0.14	0.16	0.14	0.41	0.39
	[0.13,0.19]	[0.11,0.17]	[0.14,0.19]	[0.12,0.17]	[0.39,0.42]	[0.37,0.41]
Share White	0.81	0.81	0.85	0.84	0.87	0.87
	[0.77,0.85]	[0.77,0.85]	[0.82,0.87]	[0.82,0.86]	[0.86,0.88]	[0.86,0.88]
Share Black	0.02	0.01	0.01	0.01	0.01	0.01
	[0.01,0.02]	[0.01,0.02]	[0.01,0.02]	[0.01,0.02]	[0.01,0.02]	[0.01,0.01]
Share Hispanic	0.02	0.02	0.02	0.01	0.01	0.01
	[0.01,0.03]	[0.01,0.03]	[0.01,0.02]	[0.01,0.02]	[0.01,0.02]	[0.01,0.02]
Share Asian	0.04	0.04	0.03	0.03	0.02	0.02
	[0.02,0.05]	[0.02,0.05]	[0.02,0.03]	[0.02,0.04]	[0.02,0.02]	[0.02,0.02]
Share Other Race	0.02	0.02	0.02	0.02	0.02	0.02
	[0.01,0.04]	[0.01,0.04]	[0.01,0.03]	[0.01,0.03]	[0.02,0.02]	[0.02,0.02]
Share Race missing unknown	0.09	0.1	0.08	0.08	0.07	0.07
	[0.06,0.13]	[0.06,0.14]	[0.06,0.1]	[0.06,0.1]	[0.06,0.07]	[0.06,0.07]
Number of students or degrees(gross)	3209	869	3080	827	3105	777
	[2907,3520]	[778,965]	[2814,3361]	[754,908]	[2844,3432]	[713,859]

Notes: table reports averages and 95 percent confidence intervals of 500 bootstrap predictions

.

Table 8: Robustness of findings to using more inclusive pools of two-year college students by relaxing the full-time student and age restrictions.

	Initial Credit Hours >=9		Initial Credit Hours >=6		Age<=22		Age<=24	
	(1) STEM qualified	(2) Graduate with STEM	(3) STEM qualified	(4) Graduate with STEM	(5) STEM qualified	(6) Graduate with STEM	(7) STEM qualified	(8) Graduate with STEM
Avg ACT math	25.24	25.74	25.22	25.74	25.27	25.77	25.28	25.79
	[24.92,25.57]	[25.39,26.08]	[24.93,25.51]	[25.4,26.04]	[24.96,25.55]	[25.47,26.08]	[24.99,25.56]	[25.48,26.09]
Avg ACT English	22.42	22.64	22.43	22.66	22.41	22.61	22.43	22.63
	[22.1,22.72]	[22.3,22.96]	[22.13,22.75]	[22.35,22.99]	[22.11,22.73]	[22.31,22.97]	[22.13,22.74]	[22.3,22.96]
Avg HS percentile rank (/100)	0.79	0.8	0.79	0.8	0.79	0.81	0.79	0.81
	[0.77,0.8]	[0.79,0.82]	[0.77,0.8]	[0.79,0.82]	[0.77,0.8]	[0.79,0.82]	[0.77,0.8]	[0.79,0.82]
Share Female	0.16	0.14	0.16	0.14	0.15	0.13	0.15	0.14
	[0.13,0.18]	[0.11,0.17]	[0.13,0.19]	[0.12,0.17]	[0.12,0.18]	[0.11,0.16]	[0.13,0.18]	[0.11,0.16]
Share White	0.81	0.81	0.81	0.81	0.81	0.81	0.81	0.81
	[0.77,0.85]	[0.77,0.85]	[0.76,0.84]	[0.75,0.84]	[0.77,0.85]	[0.77,0.85]	[0.77,0.85]	[0.77,0.85]
Share Black	0.02	0.01	0.02	0.01	0.02	0.02	0.02	0.02
	[0.01,0.02]	[0.01,0.02]	[0.01,0.02]	[0.01,0.02]	[0.01,0.02]	[0.01,0.02]	[0.01,0.02]	[0.01,0.02]
Share Hispanic	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02
	[0.01,0.03]	[0.01,0.03]	[0.01,0.03]	[0.01,0.03]	[0.01,0.03]	[0.01,0.03]	[0.01,0.03]	[0.01,0.03]
Share Asian	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04
	[0.02,0.06]	[0.02,0.06]	[0.02,0.06]	[0.02,0.06]	[0.02,0.05]	[0.02,0.05]	[0.02,0.05]	[0.02,0.05]
Share Other Race	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02
	[0.01,0.04]	[0.01,0.04]	[0.01,0.04]	[0.01,0.04]	[0.01,0.04]	[0.01,0.04]	[0.01,0.04]	[0.01,0.04]
Share Race missing unknown	0.09	0.1	0.1	0.1	0.09	0.1	0.09	0.1
	[0.06,0.13]	[0.06,0.14]	[0.06,0.13]	[0.06,0.14]	[0.06,0.13]	[0.06,0.14]	[0.06,0.13]	[0.06,0.13]
Number of students or degrees (gross)	3378	907	3509	942	3261	882	3271	883
	[3041,3697]	[807,1007]	[3164,3892]	[840,1049]	[2940,3559]	[793,973]	[2961,3578]	[798,973]
Initial population considered	46493		48841		44001		44335	

Notes: table reports averages and 95 percent confidence intervals of 500 bootstrap repetitions under different under different sample restrictions: Minimum registered credit hours >=9, Minimum credit hours >=6, Maximum age in freshman year <=22, Maximum age in freshman year <=24.

Table 9: Robustness of findings to removing the ACT-taking requirement among two-year-college students to be nudged.

	Main Settings		Recover Missing ACT Scores	
	(1)	(2)	(3)	(4)
	STEM qualified	Graduate with STEM	STEM qualified	Graduate with STEM
Avg ACT math	25.26	25.76	25.06	25.53
	[24.98,25.54]	[25.45,26.07]	[24.75,25.37]	[25.19,25.85]
Avg ACT English	22.42	22.64	21.6	21.79
	[22.11,22.74]	[22.3,22.95]	[21.24,21.93]	[21.39,22.2]
Avg HS percentile rank (/100)	0.79	0.81	0.77	0.79
	[0.77,0.8]	[0.79,0.82]	[0.75,0.79]	[0.77,0.8]
Share Female	0.16	0.14	0.14	0.13
	[0.13,0.19]	[0.11,0.17]	[0.12,0.17]	[0.1,0.16]
Share White	0.81	0.81	0.8	0.8
	[0.77,0.85]	[0.77,0.85]	[0.75,0.84]	[0.74,0.84]
Share Black	0.02	0.01	0.02	0.01
	[0.01,0.02]	[0.01,0.02]	[0.01,0.02]	[0.01,0.02]
Share Hispanic	0.02	0.02	0.02	0.02
	[0.01,0.03]	[0.01,0.03]	[0.01,0.04]	[0.01,0.04]
Share Asian	0.04	0.04	0.04	0.04
	[0.02,0.05]	[0.02,0.05]	[0.02,0.06]	[0.02,0.06]
Share Other Race	0.02	0.02	0.03	0.03
	[0.01,0.04]	[0.01,0.04]	[0.01,0.05]	[0.01,0.06]
Share Race missing unknown	0.09	0.10	0.10	0.10
	[0.06,0.13]	[0.06,0.14]	[0.07,0.15]	[0.07,0.16]
Number of students or degrees (gross)	3209	869	3680	984
	[2907,3520]	[778,965]	[3263,4077]	[870,1097]
Initial population considered	43214		57382	

Notes: Table reports averages and 95 percent confidence intervals over 500 bootstrap repetitions after we recover missing ACT test scores for community college students. We use students' completed credit hours and GPAs during the first semester of community college to impute missing ACT test scores.

Table 10: Findings using different nudge thresholds for identifying STEM-qualified two-year college students based on the percentile of the distribution among four-year STEM entrants (the baseline case is at the 50th percentile).

	40th Percentile		45th Percentile		55th Percentile		60th Percentile	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	STEM qualified	Graduate with STEM	STEM qualified	Graduate with STEM	STEM qualified	Graduate with STEM	STEM qualified	Graduate with STEM
Avg ACT math	24.53	25.16	24.9	25.46	25.63	26.08	26	26.4
	[24.27,24.78]	[24.89,25.42]	[24.61,25.16]	[25.17,25.74]	[25.3,25.94]	[25.73,26.4]	[25.64,26.34]	[26.01,26.76]
Avg ACT English	22.1	22.38	22.27	22.51	22.58	22.77	22.74	22.91
	[21.83,22.35]	[22.09,22.63]	[21.98,22.54]	[22.2,22.78]	[22.24,22.91]	[22.41,23.13]	[22.38,23.09]	[22.54,23.27]
Avg HS percentile rank (/100)	0.76	0.79	0.78	0.8	0.8	0.82	0.82	0.83
	[0.75,0.78]	[0.77,0.8]	[0.76,0.79]	[0.78,0.81]	[0.79,0.82]	[0.8,0.83]	[0.8,0.83]	[0.81,0.84]
Share Female	0.2	0.17	0.17	0.15	0.14	0.12	0.12	0.11
	[0.17,0.22]	[0.14,0.2]	[0.15,0.21]	[0.13,0.18]	[0.11,0.17]	[0.1,0.15]	[0.09,0.15]	[0.08,0.14]
Share White	0.81	0.81	0.81	0.81	0.81	0.81	0.81	0.81
	[0.78,0.85]	[0.77,0.85]	[0.77,0.85]	[0.77,0.85]	[0.77,0.85]	[0.76,0.85]	[0.77,0.86]	[0.76,0.86]
Share Black	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01
	[0.01,0.02]	[0.01,0.02]	[0.01,0.02]	[0.01,0.02]	[0.01,0.02]	[0.01,0.02]	[0.01,0.02]	[0.01,0.02]
Share Hispanic	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02
	[0.01,0.03]	[0.01,0.03]	[0.01,0.03]	[0.01,0.03]	[0.01,0.03]	[0.01,0.03]	[0.01,0.04]	[0.01,0.04]
Share Asian	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04
	[0.02,0.05]	[0.02,0.05]	[0.02,0.05]	[0.02,0.05]	[0.02,0.06]	[0.02,0.06]	[0.02,0.06]	[0.02,0.06]
Share Other Race	0.02	0.02	0.02	0.02	0.02	0.03	0.02	0.03
	[0.01,0.04]	[0.01,0.04]	[0.01,0.04]	[0.01,0.04]	[0.01,0.04]	[0.01,0.04]	[0.01,0.04]	[0.01,0.05]
Share Race missing unknown	0.09	0.09	0.09	0.1	0.09	0.1	0.1	0.1
	[0.07,0.12]	[0.06,0.13]	[0.06,0.13]	[0.06,0.13]	[0.06,0.13]	[0.06,0.14]	[0.06,0.14]	[0.06,0.15]
Number of students or degrees (gross)	5024	1142	4037	1003	2515	742	1944	624
	[4645,5427]	[1039,1248]	[3696,4407]	[905,1105]	[2246,2769]	[656,827]	[1721,2163]	[546,702]

Notes: table reports averages and 95 percent confidence intervals over 500 bootstrap repetitions under different nudge threshold: nudge threshold= 40th percentile, nudge threshold= 45th percentile, nudge threshold= 55th percentile, nudge threshold= 60th percentile.

Table 11: Summary statistics for STEM-qualified community college students after dropping Biology majors.

	(1) STEM qualified	(2) Graduate with STEM	(3) STEM completers at universities
Avg ACT math	25.17	25.73	26.87
	[24.84,25.51]	[25.37,26.07]	[26.75,26.97]
Avg ACT English	21.86	22.07	25.79
	[21.51,22.18]	[21.7,22.42]	[25.67,25.91]
Avg HS percentile rank (/100)	0.75	0.77	0.81
	[0.74,0.77]	[0.76,0.79]	[0.81,0.82]
Share Female	0.08	0.07	0.28
	[0.06,0.1]	[0.05,0.09]	[0.27,0.29]
Share White	0.84	0.84	0.87
	[0.79,0.88]	[0.79,0.88]	[0.86,0.88]
Share Black	0.01	0.01	0.04
	[0.01,0.02]	[0.01,0.02]	[0.03,0.04]
Share Hispanic	0.01	0.01	0.02
	[0,0.03]	[0,0.02]	[0.01,0.02]
Share Asian	0.03	0.03	0.03
	[0.02,0.05]	[0.02,0.05]	[0.02,0.03]
Share Other Race	0.02	0.03	0.01
	[0.01,0.04]	[0.01,0.05]	[0.01,0.02]
Share Race missing unknown	0.08	0.08	0.04
	[0.05,0.12]	[0.05,0.12]	[0.03,0.04]
Number of students or degrees (gross)	2995	705	6441
	[2721,3339]	[636,787]	[6307,6584]

Notes: Table reports averages and 95 percent confidence intervals over 500 bootstrap repetitions when we exclude biology from STEM majors.

Table 12: Summary statistics for STEM qualified community college students: different levels of selection on unobservables.

(1) # of Nudged Community College Students	(2) Average STEM Completion Likelihood among four-year STEM-Qualified Entrants	(3) two-year Student Selection on Observables	(4) Selection on Unobservables	(5) Average Likelihood (community college)	(6) # of STEM Degrees Produced via Nudge (gross)
3209	0.31	-0.04	0	0.27	869
3209	0.31	-0.04	-0.02 (50%)	0.25	802
3209	0.31	-0.04	-0.04 (100%)	0.23	738
3209	0.31	-0.04	-0.08 (200%)	0.19	610
3209	0.31	-0.04	-0.12 (300%)	0.15	481

Notes: Table describes the number of STEM degrees produced with different levels of selection on unobservables: 0%, 50%, 100%, 200% and 300% times selection on observables. Selection on observables value are calculated from the average likelihoods of graduating in STEM corresponding to equation (1). Column (5)= column (2)- column (3)- column (4). Column (6)= column (1) * column (5).

Table 13: Summary statistics for STEM-qualified community college students with 5/10-percent nudge compliance rate.

	Randomly Selected 5 percent		Top 5 Percent		Randomly Selected 10 percent		Top 10 Percent	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	STEM qualified	Graduate with STEM	STEM qualified	Graduate with STEM	STEM qualified	Graduate with STEM	STEM qualified	Graduate with STEM
Avg ACT math	25.26	25.76	29.29	29.39	25.27	25.76	28.45	28.61
	[24.63,25.83]	[25.07,26.46]	[28.12,30.24]	[28.16,30.34]	[24.83,25.71]	[25.26,26.27]	[27.69,29.11]	[27.72,29.34]
Avg ACT English	22.46	22.68	24.37	24.43	22.47	22.68	23.8	23.91
	[21.76,23.22]	[21.92,23.53]	[23.35,25.36]	[23.33,25.46]	[21.9,23.1]	[22.06,23.36]	[23.05,24.54]	[23.15,24.63]
Avg HS percentile rank (/100)	0.79	0.81	0.92	0.93	0.79	0.81	0.9	0.9
	[0.76,0.82]	[0.77,0.84]	[0.88,0.96]	[0.89,0.96]	[0.77,0.81]	[0.78,0.83]	[0.87,0.92]	[0.87,0.93]
Share Female	0.16	0.14	0.06	0.06	0.15	0.14	0.07	0.06
	[0.09,0.22]	[0.08,0.2]	[0.02,0.12]	[0.02,0.11]	[0.11,0.2]	[0.1,0.18]	[0.03,0.11]	[0.03,0.11]
Share White	0.81	0.81	0.78	0.78	0.81	0.81	0.79	0.78
	[0.73,0.88]	[0.73,0.88]	[0.66,0.88]	[0.65,0.88]	[0.75,0.86]	[0.75,0.87]	[0.69,0.86]	[0.69,0.87]
Share Black	0.02	0.01	0.01	0.01	0.02	0.01	0.01	0.01
	[0,0.04]	[0,0.04]	[0,0.04]	[0,0.04]	[0,0.03]	[0,0.03]	[0,0.03]	[0,0.03]
Share Hispanic	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01
	[0,0.05]	[0,0.05]	[0,0.05]	[0,0.05]	[0,0.04]	[0,0.04]	[0,0.04]	[0,0.04]
Share Asian	0.04	0.04	0.04	0.03	0.04	0.04	0.04	0.04
	[0.01,0.07]	[0.01,0.08]	[0.01,0.08]	[0.01,0.08]	[0.02,0.07]	[0.01,0.07]	[0.01,0.07]	[0.01,0.07]
Share Other Race	0.02	0.02	0.04	0.05	0.02	0.02	0.03	0.04
	[0,0.05]	[0,0.06]	[0.01,0.09]	[0.01,0.1]	[0.01,0.05]	[0.01,0.05]	[0.01,0.07]	[0.01,0.08]
Share Race missing unknown	0.09	0.1	0.11	0.11	0.09	0.1	0.11	0.11
	[0.05,0.15]	[0.05,0.16]	[0.03,0.21]	[0.03,0.22]	[0.05,0.14]	[0.05,0.14]	[0.05,0.21]	[0.04,0.21]
Number of students or degrees (gross)	159	43	159	90	318	86	318	159
	[145,175]	[38,49]	[145,175]	[80,103]	[288,349]	[77,96]	[288,349]	[142,180]
Number of degrees (net)	29		54		58		95	
	[21,37]		[40,69]		[45,70]		[71,120]	

Notes: Table reports averages and 95 percent confidence intervals over 500 bootstrap repetitions when 5/10 percent STEM-qualified community college students actually choose to enroll at universities. Randomly selected 5/10 percent in column (1) and (2)/ (5) and (6); top 5/10 percent in terms of prediction likelihoods in column (3) and (4)/(7) and (8).

Appendix A
Supplementary Tables

Appendix Table A1: Construction of the analytic sample.

first time entrants at four year universities from 2006 to 2010		97749	first time entrants at community colleges from 2006 to 2010		110695
	record lost	Remaining sample		record lost	remaining sample
Out of state/ Foreign student	14486	83263	Out of state/ Foreign student	2497	108198
Not full time	3320	79943	Not full time	32647	75551
Older than 20	2656	77287	Older than 20	12228	63323
Missing high school code	6053	71234	Missing high school code, or out of state high school	5529	57794
Missing ACT Math score or ACT English Score	398	70836	Missing ACT Math score or ACT English Score	14537	43257
Missing high school rank*	8199	70836	Missing high school rank*	28299	43257
Drop extremely small high schools **	99	70737	Drop extremely small high schools **	43	43214

* We do not drop those students whose high school percentile ranks are missing, instead we impute their ranks based on other covariates
** We drop high schools that sent five or fewer students to a public college during the period covered by our data panel

Appendix Table A2: Summary statistics for the university sample.

	(1) Raw	(2) In state	(3) In state full time	(4) In state full time <=20	(5) In state full time <=20, in state hs	(6) In state full time <=20, in state hs, nomissing ACT	(7) Analytic sample
ACT math	22.84	22.7	22.83	22.89	22.88	22.89	22.89
	(4.88)	(4.87)	(4.82)	(4.81)	(4.78)	(4.78)	(4.78)
ACT English	23.62	23.49	23.64	23.7	23.68	23.68	23.68
	(5.45)	(5.46)	(5.39)	(5.37)	(5.34)	(5.34)	(5.34)
HS percentile rank	0.69	0.69	0.7	0.7	0.7	0.71	0.71
	(0.23)	(0.23)	(0.23)	(0.22)	(0.22)	(0.22)	(0.22)
HS percentile rank missing indicator	0.16	0.16	0.15	0.14	0.12	0.12	0.12
	(0.37)	(0.36)	(0.35)	(0.34)	(0.32)	(0.32)	(0.32)
Female	0.55	0.55	0.55	0.55	0.55	0.55	0.55
	(0.5)	(0.5)	(0.5)	(0.5)	(0.5)	(0.5)	(0.5)
White	0.77	0.79	0.79	0.79	0.8	0.81	0.81
	(0.42)	(0.41)	(0.41)	(0.4)	(0.4)	(0.39)	(0.39)
Black	0.12	0.12	0.11	0.11	0.11	0.1	0.1
	(0.32)	(0.32)	(0.31)	(0.31)	(0.31)	(0.3)	(0.3)
Hispanic	0.02	0.02	0.02	0.02	0.02	0.02	0.02
	(0.15)	(0.15)	(0.15)	(0.15)	(0.14)	(0.14)	(0.14)
Asian	0.02	0.02	0.02	0.02	0.02	0.02	0.02
	(0.14)	(0.14)	(0.14)	(0.14)	(0.14)	(0.14)	(0.14)
Other Race	0.03	0.02	0.02	0.02	0.01	0.01	0.01
	(0.17)	(0.14)	(0.13)	(0.13)	(0.11)	(0.11)	(0.11)
Race missing unknown	0.04	0.04	0.04	0.04	0.04	0.04	0.04
	(0.2)	(0.19)	(0.19)	(0.19)	(0.19)	(0.19)	(0.19)
Number of observations	97749	83263	79943	77287	71234	70836	70737

Appendix Table A3 Summary statistics for the community college sample.

	(1) Raw	(2) In state	(3) In state full time	(4) In state full time <=20	(5) In state full time <=20, in state hs	(6) In state full time <=20, in state hs, nomissing ACT	(7) Analytic sample
ACT math	18.84	18.84	19.01	19.05	19.04	19.04	19.04
	(3.85)	(3.85)	(3.8)	(3.79)	(3.78)	(3.78)	(3.78)
ACT English	18.83	18.83	19.02	19.04	18.99	18.99	18.99
	(4.95)	(4.95)	(4.82)	(4.81)	(4.78)	(4.78)	(4.78)
HS percentile rank	0.48	0.49	0.51	0.52	0.52	0.56	0.56
	(0.25)	(0.25)	(0.24)	(0.24)	(0.24)	(0.23)	(0.23)
HS percentile rank missing indicator	0.55	0.55	0.46	0.41	0.37	0.35	0.35
	(0.5)	(0.5)	(0.5)	(0.49)	(0.48)	(0.48)	(0.48)
Female	0.54	0.54	0.53	0.52	0.52	0.54	0.54
	(0.5)	(0.5)	(0.5)	(0.5)	(0.5)	(0.5)	(0.5)
White	0.71	0.71	0.76	0.77	0.78	0.79	0.79
	(0.45)	(0.45)	(0.43)	(0.42)	(0.41)	(0.41)	(0.41)
Black	0.14	0.13	0.1	0.09	0.09	0.08	0.08
	(0.34)	(0.34)	(0.3)	(0.29)	(0.28)	(0.27)	(0.27)
Hispanic	0.03	0.03	0.03	0.03	0.02	0.02	0.02
	(0.17)	(0.16)	(0.16)	(0.16)	(0.15)	(0.15)	(0.15)
Asian	0.01	0.01	0.01	0.01	0.01	0.01	0.01
	(0.12)	(0.12)	(0.11)	(0.11)	(0.11)	(0.11)	(0.11)
Other Race	0.03	0.03	0.02	0.02	0.02	0.02	0.02
	(0.16)	(0.16)	(0.15)	(0.15)	(0.14)	(0.14)	(0.14)
Race missing unknown	0.09	0.09	0.08	0.08	0.08	0.07	0.07
	(0.28)	(0.28)	(0.27)	(0.27)	(0.26)	(0.26)	(0.26)
Number of observations	110695	108198	75551	63323	57794	43257	43214

Appendix Table A4: Summary statistics for university students.

	STEM-entrants		non-STEM entrants	
	STEM qualified	Not STEM qualified	STEM qualified	Not STEM qualified
ACT math	28.7	22.45	27.53	20.84
	[28.59,28.8]	[22.34,22.56]	[27.42,27.65]	[20.79,20.9]
ACT English	27.01	23.36	26.9	22.38
	[26.87,27.14]	[23.23,23.5]	[26.76,27.06]	[22.3,22.45]
HS percentile rank	0.86	0.68	0.86	0.65
	[0.86,0.87]	[0.67,0.69]	[0.86,0.87]	[0.65,0.65]
HS percentile rank missing indicator	0.1	0.11	0.12	0.12
	[0.09,0.11]	[0.1,0.11]	[0.11,0.13]	[0.11,0.12]
Female	0.21	0.5	0.33	0.67
	[0.2,0.23]	[0.49,0.51]	[0.31,0.35]	[0.66,0.67]
White	0.86	0.77	0.87	0.79
	[0.85,0.87]	[0.76,0.78]	[0.86,0.88]	[0.78,0.79]
Black	0.03	0.13	0.02	0.13
	[0.02,0.04]	[0.12,0.14]	[0.02,0.03]	[0.13,0.13]
Hispanic	0.02	0.03	0.02	0.02
	[0.01,0.02]	[0.02,0.03]	[0.01,0.02]	[0.02,0.02]
Asian	0.04	0.01	0.04	0.01
	[0.04,0.05]	[0.01,0.02]	[0.04,0.05]	[0.01,0.01]
Other Race	0.01	0.02	0.01	0.01
	[0.01,0.02]	[0.01,0.02]	[0.01,0.01]	[0.01,0.01]
Race missing unknown	0.03	0.04	0.04	0.04
	[0.03,0.04]	[0.03,0.04]	[0.03,0.04]	[0.03,0.04]
Number of students	7569	7570	10940	44658
	[7466,7665]	[7466,7668]	[10549,11391]	[44163,45118]

Appendix Table A5: In-sample and out-of-sample predictive validity among the university sample, equation (1).

	(A) In-sample		(B) Out-of-sample	
	(1) Actual Observed	(2) Predicted Value	(3) Actual Observed	(4) Predicted Value
Avg ACT math	26.64	26.64	26.62	26.60
	[26.57,26.72]	[26.57,26.72]	[26.47,26.77]	[26.5,26.71]
Avg ACT English	26.07	26.07	26.02	26.02
	[25.98,26.14]	[25.98,26.14]	[25.82,26.21]	[25.92,26.14]
Avg HS percentile rank	0.82	0.82	0.82	0.82
	[0.82,0.82]	[0.82,0.82]	[0.82,0.83]	[0.82,0.83]
Share Female	0.36	0.36	0.36	0.36
	[0.35,0.37]	[0.35,0.37]	[0.34,0.4]	[0.35,0.38]
Share White	0.86	0.86	0.86	0.85
	[0.85,0.86]	[0.85,0.86]	[0.84,0.87]	[0.85,0.86]
Share Black	0.04	0.04	0.04	0.04
	[0.04,0.05]	[0.04,0.05]	[0.03,0.05]	[0.04,0.05]
Share Hispanic	0.02	0.02	0.02	0.02
	[0.02,0.02]	[0.02,0.02]	[0.01,0.02]	[0.02,0.02]
Share Asian	0.03	0.03	0.03	0.03
	[0.03,0.04]	[0.03,0.04]	[0.03,0.04]	[0.03,0.04]
Share Other Race	0.01	0.01	0.01	0.01
	[0.01,0.01]	[0.01,0.01]	[0.01,0.02]	[0.01,0.02]
Share Race missing unknown	0.04	0.04	0.04	0.04
	[0.03,0.04]	[0.03,0.04]	[0.03,0.05]	[0.03,0.04]
Number of students	7158	7159	1820	1839
	[6965,7302]	[6965,7302]	[1749,1890]	[1777,1887]

Notes: Table shows the in-sample and out-of-sample comparison of predicted values versus true outcomes using equation (1) and the corresponding sample of initial STEM entrants. We use 80% of the data for the "training dataset" and the remaining 20% to test out-of-sample predictive validity. Averages and 95 percent confidence intervals over 500 bootstrap repetitions are provided.

Appendix Table A6: Summary statistics for STEM-qualified community college students in supplementary predictive models.

	(1) Graduate with Non-STEM	(2) Drop-out
Avg ACT math	25.08	25.09
	[24.84,25.34]	[24.76,25.43]
Avg ACT English	22.72	22
	[22.45,23.01]	[21.65,22.35]
Avg HS percentile rank	0.83	0.74
	[0.82,0.84]	[0.72,0.75]
Share Female	0.23	0.1
	[0.19,0.26]	[0.07,0.12]
Share White	0.84	0.78
	[0.81,0.87]	[0.74,0.82]
Share Black	0.01	0.02
	[0.01,0.02]	[0.01,0.03]
Share Hispanic	0.02	0.02
	[0.01,0.03]	[0.01,0.04]
Share Asian	0.03	0.04
	[0.02,0.05]	[0.03,0.07]
Share Other Race	0.02	0.03
	[0.01,0.03]	[0.01,0.05]
Share Race missing unknown	0.08	0.11
	[0.05,0.11]	[0.07,0.14]
Number of non-STEM degrees or dropouts	1145	1180
	[1036,1269]	[1065,1312]

Notes: Table reports averages and 95 percent confidence intervals over 500 bootstrap repetitions for nudged community college students: graduate with a non-STEM degree in column (1) and fail to graduate with any bachelor's degrees in column (2).

Appendix B
Supplementary Predictive Models

In addition to the main model predicting STEM attainment, we also estimate two separate, supplementary models to predict the likelihood of graduating with a non-STEM degree and the likelihood of failing to earn any four-year degree (within six years). These models are of the same structure as equation (1):

$$P^*_{ijt} = X_i\beta_1 + \gamma_{1j} + \delta_{1t} + \varepsilon_{1ijt} \tag{A1}$$

In equation (A1), P^*_{ijt} is either the latent utility of completing a non-STEM degree within six years, or failing to complete a bachelor's degree within six years.

We estimate the models for these outcomes independently, but note that in conjunction with the main model estimated in the text for STEM degrees, the model can be further modified to account for outcome-dependence. That is, we can specify a single multinomial outcome and model the outcomes jointly. We did not do this here because we view this as an add-on to the main analysis and do not wish to overwrite the main model. That said, in unreported results we have confirmed that inference from the main analysis, and Appendix Table A6, is very similar if we use a multinomial model that accounts for outcome-dependence in the data to generate the predictions for the three categorical outcomes we consider in this brief extension: STEM degree attainment, non-STEM degree attainment, and dropout.