# Could Shifting the Margin Between Community College and University Enrollment Expand and Diversify University Degree Production in STEM Fields?

Cheng Qian
Cory Koedel

March 2023

WORKING PAPER No. 244-0323-2

CALDER
National Center for Analysis of Longitudinal Data in Education Research

AIR®

# Could Shifting the Margin between Community College and University Enrollment Expand and Diversify University Degree Production in STEM Fields?

**Cheng Qian**
*University of Missouri*

**Cory Koedel**
*University of Missouri / CALDER*

# Contents

## Acknowledgments

---

Qian is in the Department of Economics, and Koedel is in the Department of Economics and Truman School of Government & Public Affairs, at the University of Missouri. We thank the Missouri Department of Higher Education and Workforce Development for providing access to data and the Center for the Analysis of Longitudinal Data in Education Research (CALDER), which is funded by a consortium of foundations, for support.

CALDER working papers have not undergone final formal review and should be cited as working papers. They are intended to encourage discussion and suggestions for revision before final publication. Any opinions, findings, and conclusions expressed in these papers are those of the authors and do not necessarily reflect the views of our funders or the institutions to which the authors are affiliated. All errors and opinions are our own.

CALDER • American Institutes for Research
1400 Crystal Drive 10th Floor, Arlington, VA 22202
202-403-5796 • www.caldercenter.org

***Could Shifting the Margin between Community College and University Enrollment Expand and Diversify University Degree Production in STEM Fields?***

Cheng Qian and Cory Koedel

CALDER Working Paper No. 244-0323-2
March 2023

## Abstract

We examine the potential to expand and diversify the production of university STEM degrees by shifting the margin of initial enrollment between community colleges and 4-year universities. Our analysis is based on statewide administrative microdata from the Missouri Department of Higher Education and Workforce Development covering enrollees in all public postsecondary institutions statewide. We find that the potential for shifting the enrollment margin to expand degree production in STEM fields is modest, even at an upper bound, because most community college students are not academically prepared for bachelor's degree programs in STEM fields. We also find that shifting the enrollment margin is unlikely to improve racial/ethnic diversity among university STEM degree recipients. This is because community college students at the enrollment margin are less diverse than their peers who enter universities directly.

## 1. Introduction

Improving and expanding education in science, technology, engineering, and mathematics (STEM) fields is a consistent policy priority at the highest levels of U.S. government and an active area of academic scholarship (Coleman, Smith and Miller, 2019; National Science & Technology Council, 2018; White House, 2016). Policy interest in STEM education is motivated by the concern that the United States is falling behind globally in the production of STEM human capital (Atkinson and Ezell, 2012; National Academy of Sciences, National Academy of Engineering, & Institute of Medicine of the National Academies, 2007). Diversifying the STEM workforce is also an explicit policy objective, as participation in STEM fields is low among women and underrepresented minorities.[1] An implication is that the size of the STEM workforce can be expanded by increasing participation (Anderson and Kim, 2006; Committee on Science, Engineering, and Public Policy, 2011); diversification may also make the workforce more productive given that diversity is associated with more innovation in firms (Nathan and Lee, 2013).[2,3]

We examine the potential to expand and diversify the production of university STEM degrees by simulating a shift of the enrollment margin between community colleges and universities (we use the terms "university" and "4-year institution" interchangeably). The community college population is an appealing group to consider for such a shift because it is

[1] Numerous government programs operate with the goal of improving STEM diversity. 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] Although this is not a consistent finding along the dimension of race/ethnicity—for a counterexample, see Østergaard, Timmermans, and Kristinsson (2011).
[3] We take the substantial policy (and scholarly) interest in expanding and diversifying the STEM pipeline as a point of departure for our study, noting there is some debate as to how much these objectives are driven by labor-market demand. A simple data point suggesting demand for STEM workers is strong is the large and persistent aggregate wage gap favoring graduates in STEM fields relative to their peers in non-STEM fields, although disaggregating the gap suggests some STEM fields are in greater demand than others (Webber, 2016). Other factors motivating the policy interest in STEM education include social equity and national security (per the citations in this paragraph).

large and demographically and socioeconomically diverse (Deming, Goldin and Katz, 2012; Provasnik and Planty, 2008; Wang, 2013). Community college students also have a revealed preference for the pursuit of higher education and most indicate university aspirations.[4] The potential benefits of infusing the STEM pipeline with community college students have been discussed extensively in prior research (e.g., see Bahr et al., 2017; Evans, Chen, and Hudes 2020; Hagedorn and Purnamasari, 2012; and Terenzini et al., 2014).

In our policy simulation, we shift the enrollment margin of academically qualified community college students so they initially enroll at universities instead. While it is useful to think of our hypothetical intervention conceptually as shifting the enrollment margin, in the process we do much more. This is because our intervention also implicitly aligns other differences between observationally similar community college and university students. These include differences in financial resources, personal circumstances that might tie some students closer to home, etc. We elaborate on our methodology below, but in short, our hypothetical intervention is ideal in the sense that it removes whatever barriers exist—ranging from financial to personal—that lead academically well-prepared students to enroll at community colleges rather than universities. This equating of circumstances between observationally similar community college and university students can be described as assuming selection on observables into college sector. We further assume that all the students we intervene with change their behavior by shifting their enrollment, 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, 2015; DellaVigna and Linos, 2020; Oreopoulos and Petronijevic, 2019). These assumptions are strong and lead to what we interpret as upper-bound estimates of the potential

[4] 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 goal as a bachelor's degree or higher.

for an enrollment shift to expand and diversify 4-year degree production in STEM fields.

Even with these upper-bound estimation conditions, we find what we view as only modest potential to expand the production of university STEM degrees by shifting the enrollment margin. Specifically, we estimate that under the ideal conditions of our policy simulation, university STEM degree production increases by just 6.4 percent. The expansion effect is limited because most community college students do not have the academic qualifications necessary to succeed in a university STEM program. While this is not entirely surprising, we show that the magnitude of the drop-off is substantial from the full population of community college students to those who have these qualifications.

In addition, among those community college students with such lofty qualifications, racial-ethnic diversity is limited. In fact, it is less than among existing university STEM entrants. Thus, while at a cursory glance community college students seem like an appealing source to diversify the university-trained STEM workforce, our analysis suggests policies that aim to shift the enrollment margin will likely have a limited impact in this regard.

Our findings are informative in the conditions we do and do not alter in our policy simulation. A broad policy type subsumed by our simulation is financial aid that puts community college students on equal financial footing with observationally similar university students. Our results suggest that such aid, even if well-targeted, is unlikely to change the macro-level features of the STEM pipeline. We do not alter (1) students' pre-college academic qualifications, or (2) the efficacy of the postsecondary system into which students enter. One way to interpret the weak upper-bound results for the enrollment-shift policy we test is as follows: they imply that without addressing students' pre-college academic preparation, or the ability of postsecondary systems to better cultivate student potential in STEM, the STEM pipeline is unlikely to meaningfully change.

## 2. Motivation and Background

The demographic and socioeconomic diversity of community college students, along with their large numbers, make them an appealing group to consider in efforts to expand and diversify bachelor's degree production in STEM fields. In terms of sheer numbers, enrollment in public community colleges is almost two-thirds as large as enrollment in public universities nationwide (Snyder, de Brey and Dillow, 2019). Moreover, most students who enroll in community colleges have university aspirations (Deming, Goldin, and Katz, 2012). In terms of diversity, Deming, Goldin, and Katz (2012) and Wang (2013) show that community college students are more likely than university students to be Black and Hispanic and in the lowest quartile of socioeconomic status (SES). Wang (2013) further shows that demographic and SES disparities between university and community college students remain conditional on interest in STEM fields. This implies the subpopulation of community college students with STEM interests reflects the broader diversity of the community college population as a whole.

However, despite the surface appeal, infusing the university STEM pipeline with community college students presents many challenges. To understand these challenges, it is important to first understand why students attend community colleges in the first place.[5] A review of the extant literature identifies three factors that primarily drive community college

[5] The most natural alternative to the pool of 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 appealing in some respects—most notably, many have strong academic qualifications—a drawback is that they have actively decided against pursuing STEM degrees. 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, shifting academically-qualified community college students to universities, then letting them self-select into STEM fields, ensures that students' inherent field-selection processes are preserved.

enrollment. First is inadequate academic preparation for university coursework. While some students who attend community colleges have strong academic qualifications, many do not, and on average, students from community colleges are much less prepared academically than their university peers (Long and Kurlaender, 2009; also see below). Whether there are enough community college students with the academic preparation necessary to succeed in a university STEM program is a point of debate in the literature. Hoxby and Avery (2013) argue there are many academically-qualified, low-income students undermatched to postsecondary institutions, including community colleges. However, Chetty et al. (2020) find the Hoxby and Avery numbers are inflated. Our study is an indirect, empirical test for these competing estimates—undergirding our analysis is the question of how many community college students have the academic preparation necessary to succeed in a university STEM program if we could shift their enrollment.

The second and third factors that primarily drive community college enrollment are location and price (Somers et al., 2006). Turley (2009) argues students attend colleges locally because it is more convenient logistically, financially, and emotionally. Turley's discussion highlights the confluence of location and price in promoting community college enrollment. In addition to having lower tuition, attending a local community college allows students to live at home, saving on the significant room-and-board costs of non-local attendance at a university.

Many mediating variables undergird these factors that drive community college enrollment. For instance, students who are underprepared for university coursework at the conclusion of high school may have lacked exposure to high-quality K-12 schools or had other experiences that led to underinvestment in their education. And the draw of a local and low-cost college may reflect a variety of student needs and preferences with respect to academic ambition, family priorities and responsibilities, etc. More broadly, Evans et al. (2020) provide evidence on the importance of a variety of barriers—what they term "life barriers"—that impact community college students. These include problems with housing, childcare, transportation, and work schedules, among others, all of which may contribute to the decision to enroll in a local community college.

Once students enroll in community colleges, institutional factors can impede their educational progress. Most community college students enter with university aspirations, but transfer rates are low, and the majority never enroll in a university (Horn, L., & Skomsvold, 2012). Many students have trouble navigating the complexities of college systems and institutional norms, especially first-generation students (Scott-Clayton, 2015; Stephens et al., 2013). Of particular relevance are the complexities of vertical transfer policies between community colleges and universities. Long and Kurlaender (2009) show that students who enter community colleges with plans to transfer are less likely to complete university degrees even when they are similar to university entrants in terms of measured academic preparation. Wang (2015) provides related evidence for degrees in STEM fields specifically. Xu et al. (2018) identify a mechanism that causes transfer frictions—they show that transfer students experience significant credit loss during the transfer process.

These studies contribute to the long-standing debate about the democratization versus diversion view of community colleges (Dougherty, 1987; Rouse, 1995). The democratization view posits that community colleges increase access to postsecondary education through open admissions, ease of access, and low costs. The diversion view posits that the flexibility afforded by community colleges allows students to maintain weaker postsecondary connections, ultimately worsening outcomes. Research suggests that community colleges have both democratization and diversion effects, but for different student populations (Dougherty, 1987; Rouse, 1995). Extending Adelman's (2006) general concept of academic momentum to the STEM context, Wang (2015) ultimately concludes that "community colleges have not yet evolved into a pathway to a baccalaureate in STEM that is comparable with public 4-year institutions" (p. 387). For high-achievers with interest in STEM fields, the evidence is more aligned with the diversion hypothesis.

This provides a jumping-off point for our policy simulation, in which we hypothetically shift the enrollment of community college students with strong pre-college academic preparation to universities. One way to conceptualize our policy simulation is that we identify community college enrollees for whom academic preparation at entry is *not* a barrier—i.e., these students have similar observable academic qualifications to successful university entrants—then consider the potential impacts of an intervention that could remove all other barriers that lead them to initially enroll in a community college rather than a university. The hypothetical intervention we consider would be quite powerful if it could remove all of these barriers, which is why we present our findings as upper bounds on the potential for interventions with community college students to impact the university STEM pipeline.

The idea that student outcomes can be improved by shifting their enrollment decisions at college entry is not new. It has been popularized in numerous studies of undermatching between students and colleges (e.g., see Belasco and Trivette, 2015; Hoxby and Avery, 2013; Smith, Pender, and Howell, 2013) and motivated direct intervention. For instance, a recent information intervention conducted by The College Board aimed to combat undermatching by promoting attendance at more selective colleges among low- and middle-income students (Gurantz et al., 2021). Hyman (2020) implements and evaluates a similar intervention in Michigan.

Like our policy simulation, these interventions do not aim to improve the quality of postsecondary institutions, nor do they aim to improve K-12 schools. Holding these factors fixed, the goal is to improve student outcomes by changing their enrollment decisions. There are several mechanisms by which shifting the enrollment margin can improve outcomes. First, to the extent that the leaky community college pipeline is the product of complex and burdensome transfer policies, it circumvents these policies. Second, universities have more resources than community colleges and this may lead to greater student success (Cohodes and Goodman, 2014). Third, shifting students from community colleges to universities changes their peer groups to include students with stronger academic attachment, on average. All these factors may help students develop academic momentum in STEM fields and lead to more STEM success.

Noting all of this, education institutions could also do more to improve the quality of STEM education. For instance, universities could improve the efficiency with which they convert academic potential into positive outcomes among underprepared students (Hrabowski III, Rous, and Henderson, 2019). And K-12 education systems could be improved to increase students' levels of academic preparation and interest in STEM. But we conduct our policy simulation under the condition that no changes are made to the current operational efficiency of postsecondary programs or K-12 schools. In the conclusion we discuss how in light of our disappointing findings, these missing aspects of our policy simulation provide insights about how to move forward in efforts to expand and diversify the university STEM pipeline.

## 3. Data

### *3.1 Data Overview & Missouri Context*

We use administrative microdata from the Missouri Department of Higher Education and Workforce Development (DHEWD). The DHEWD data cover the universe of enrollees in public postsecondary institutions statewide in Missouri. For each student, we have information on demographic and background characteristics (race, gender, age, high school attended), pre-entry academic qualifications (high school class percentile rank, ACT test scores), and in-college outcomes (declared majors at entry, final majors, and graduation outcomes). We restrict our analytic sample to first-time, full-time, state-resident students who entered a public community college or university between 2006 and 2010 as college freshman (there are 14 community colleges and 13 universities in Missouri). We track students for up to six years after entry to determine whether they graduate with a bachelor's degree from any public university in Missouri.[6]

Table 1 shows summary statistics for all university and community college students in Missouri and for various restricted subsamples that lead to our primary analytic sample. The first restriction, imposed going from columns (1)-(2) and (4)-(5) in Table 1, focuses the analysis on in-state students. This is due to data limitations for out-of-state students, especially at community colleges.[7] 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 credit hours, (c) missing math or English ACT scores, and/or (d) are from a very small high school (defined as a school that sent five or fewer students to a public university during the

[6] Our data are comprehensive for public colleges and universities statewide but do not cover private or out-of-state institutions. Therefore, we cannot speak to the potential for expanding the STEM pipeline by redirecting community college students to these universities. We view this as a modest limitation in the context of our thought experiment because if we were to redirect 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).

[7] 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 collegegoers, which creates analytic problems in our empirical models.

period covered by our data panel) or the high school attended is missing. The age restriction is to focus on the population most likely to respond to an intervention that shifts enrollment. The ACT and full-time-enrollment restrictions are because we treat the steps of taking the ACT and enrolling full time as indicators of stronger interest and ability to pursue a higher degree among community college students (virtually all university 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.[8]

Students' class percentile ranks are important predictors of STEM success and Table 1 shows they are missing for many community college students. This is due to inconsistent institution-level reporting, which derives partly from community colleges' open enrollment policies. For students with missing class percentile ranks, we impute them via linear regression using demographic information, math and English ACT scores, and high schools of attendance, which are available for all students in the sample.[9]

Additional information about the construction of our dataset is provided in Appendix Tables A1-A3. In Appendix Table A1 we document how each data restriction we impose affects the final sample size, and in Appendix Tables A2 and A3 we provide more detailed breakdowns of how each data restriction affects the composition of the university and community-college samples. In subsequent appendix tables that we discuss below, we also show that relaxing the data restrictions described in this section does not impact our findings substantively.

[8] This restriction has no substantive implications for our analysis because it affects very few students.
[9] A feature of imputing using linear regression is that the imputed values are shrunken toward the mean. This is problematic because students in the upper tail of the distribution of academic qualifications 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 2-year and 4-year) to match the variance of observed percentile ranks at the same level. We report on the sensitivity of our findings to the use of our variance inflation procedure in Appendix Table A (Table A6).

There are strengths and weaknesses of using the Missouri data. A key strength is the administrative data cover a large sample of university and community college students—from Table 1, our combined analytic sample exceeds 110,000 students. This is a much larger sample than would be available in a typical national survey dataset and useful for our prediction-based framework described below (because the predictions are more precise with the large sample). Moreover, the fact that our sample is concentrated in just one state means we have large high-school-level sample sizes. We leverage this feature of the dataset to incorporate high school effects into our prediction models, which prior research suggests greatly improve the accuracy with which postsecondary outcomes can be predicted (Arcidiacono and Koedel, 2014).

The biggest limitation of our dataset is that the focus on Missouri raises questions about generalizability. While we cannot speak directly to the generalizability of our findings to other states (ultimately this is an empirical question), we provide context for interpreting our findings in two ways. First, we note that Missouri is a middle-of-the-distribution state in many respects. For instance, it is the middle third of states as measured by total population, the percent of the population that is Black, median household income, and the percent of the population with at least a bachelor's degree. Missouri is also in the middle third of states in terms of education-system indicators such as per-pupil spending on K-12 education, college-going rates among high school graduates, and tuition and fees at public universities. Missouri is in the bottom third of states for tuition and fees at public community colleges, although still outside the bottom ten.[10] In short, Missouri is not an outlying state along any obvious observable dimension, suggesting at least some degree of generalizability.

[10] The information provided in this paragraph was compiled by the authors using data from the U.S. Census, the St. Louis Federal Reserve Economic Data (FRED) system, and the 2018 Digest of Education Statistics (Snyder, de Bray, and Dillow, 2019). We would have liked to report on postsecondary state support, but no per-pupil measure is available, making cross-state comparisons difficult.

The second notable aspect of our context is the structure of the postsecondary education system in Missouri. For community colleges, one question is whether their geographic distribution is sufficiently dispersed so that they are widely accessible. This is important because our analysis is based on the thought experiment of shifting initial community college enrollees, and if a structural aspect of the system affects who enrolls, it could affect our sample and thus the generalizability of our findings. Although we do not have a simple quantitative metric to describe the geographic distribution of campuses, community colleges are dispersed throughout the state, and more (and larger) colleges operate in the primary population centers in Kansas City and St. Louis. Nothing seems atypical about Missouri in this regard.

The production of university STEM degrees in Missouri also merits discussion. Two Missouri public universities—University of Missouri-Columbia and Missouri University of Science and Technology—account for about 55 percent of all university STEM degrees awarded. The remaining 45 percent of STEM degrees are spread across the other 11 public universities. Because our analysis is based on empirical models that predict student success in university-based STEM programs, the efficacy of the two major STEM-producing universities play an important role in driving our findings. While the level of concentration of STEM education in Missouri seems in line with most state public university systems, it is unclear if there are substantive differences in the efficacy of Missouri's STEM-producing universities compared to their analog institutions in other states. While we have no reason to expect substantial differences, it is outside of the scope of our study to test for differences directly. This issue merits attention in future research aimed at exploring the generalizability of our results. [11]

[11] Our claim that Missouri seems typical in terms of STEM concentration is based on a separate investigation of the concentration of engineering degree production in states' public university systems using the Integrated Postsecondary Education Data System (IPEDS). This work is part of a parallel project and the exact findings cannot

Finally, historically, Missouri has been among the states with stronger statewide vertical transfer policies. The closest rating of Missouri's policies to the time period of our study is from Ignash and Townsend (2000), who rated Missouri's statewide articulation agreement as "fairly strong," which is their second-highest rating category and puts Missouri in the top half of states. That said, while Missouri rates well compared to other states historically, transfer policies in many states have strengthened in recent years, including Missouri. This means the transfer policies facing our cohorts were more complex and difficult to navigate than those faced by their contemporary peers.[12] The transfer landscape in Missouri is mostly peripheral to our analysis because our policy experiment circumvents community colleges entirely, with one exception noted below.

### *3.2 Insights from the Raw Data*

The racial/ethnic information in Table 1 previews our diversity findings. While the full community college population has a higher share of Black students than the university population, the gap is not large, at 0.14 versus 0.12 (given Missouri demographics, the proportion Black is the most relevant consideration for diversity). 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), then to just 0.08 after we add the restrictions for age and taking the ACT. In the analytic sample, the Black share among community college entrants is below the Black share among university entrants.

In Table 2 we provide descriptive statistics for university students by STEM status. We use Classification of Instructional Programs (CIP) codes to identify majors in STEM fields, then

---

be disclosed here, but with the caveat that engineering is just one STEM subfield, we find that Missouri is near the middle of the national distribution in terms of the concentration of engineering degree production.

[12] Beginning in the 2018-19 school year, Missouri implemented the CORE 42 curriculum, which is a robust transfer curriculum shared by all public two- and four-year institutions statewide.

divide students into groups of STEM and non-STEM entrants. 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 in the social and behavioral sciences. Table 2 shows that 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 among STEM entrants (Chen, 2009) and a slightly lower percentage of underrepresented minority students (Hill, 2017). But while 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 going from STEM entrants to completers (also see Arcidiacono, Aucejo, and Spenner, 2012).[13] Unsurprisingly, both STEM and non-STEM entrants at universities possess substantially stronger academic qualifications than their community college peers, on average.

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 STEM degrees at 4-year institutions.

[13] It is beyond the scope of our analysis to consider interventions that would reduce attrition among Black students between STEM entry and completion. For a deeper discussion of this and related issues from the perspective of higher education institutions, see Hrabowski III, Rous, and Henderson (2019).

## 4. Methodology

### *4.1 Prediction Model*

The first step in constructing our hypothetical intervention is to identify the margin for the enrollment shift. If the objective function is simply to maximize the number of STEM degrees produced, then the optimal policy would shift all community college students. However, it would be costly and undesirable to shift students to universities who are underprepared or uninterested in STEM fields and have low likelihoods of success. Therefore, we focus only on students who we identify as "STEM qualified." We use this term broadly to describe students who are likely to have necessary academic preparation, and interest, in STEM fields as reflected by their characteristics and qualifications prior to college entry. We operationalize the notion of a STEM-qualified student empirically—i.e., a STEM-qualified student is one whose observable characteristics predict the attainment of a STEM degree with (relatively) high likelihood.

To find these students, we begin with the following model that predicts STEM degree completion within six years of initial enrollment, which we estimate using data from all university entrants and specify as a logistic regression:

$$Y_{ijt}^{*} = \boldsymbol{X_i \beta_1} + \gamma_j + \delta_t + \varepsilon_{ijt} \qquad (1)$$

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 4-year public institutions in year *t*. Students who complete a STEM degree within six years—i.e., $Y_{ijt} = 1$—have latent utility above zero. $\boldsymbol{X_i}$ 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. $\gamma_j$ is a fixed effect for high school *j*, $\delta_t$ a fixed effect for year *t*, and $\epsilon_{ijt}$ is an idiosyncratic error.

The vectors of control variables and fixed effects in equation (1) are included based on previous research. Arcidiacono and Koedel (2014) show that students' class ranks and high school fixed effects combine as powerful predictors of postsecondary outcomes. It is a significant advantage of our dataset that we can include the high school fixed effects in the prediction model. Our inclusion of students' math and English ACT scores is motivated by evidence from Bettinger, Evans, and Pope (2013), who show these are the strongest predictors of college performance among available ACT subjects. The inclusion of the math score is further supported by evidence that that math skills in particular are critical for postsecondary success in STEM (e.g., see Hagedorn and DuBray, 2010).

The fitted values from equation (1), $\hat{P}_{ijt} = Pr(Y_{ijt} = 1|\boldsymbol{X_i}, \gamma_j, \delta_t)$, give predicted likelihoods of completing a STEM degree among university entrants based on pre-entry characteristics and qualifications. We apply the parameter estimates from equation (1) to the profiles of community college students to generate 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 4-year degree in a STEM field if the student initially enrolled in a university instead of a community college, and if the student was similar in every way to a student with the same observable attributes $(\boldsymbol{X_i}, \gamma_j, \delta_t)$ who initially enrolled in a university, on average. Said another way, the predicted value $\hat{P}_{ijt}^{cc}$ gives the likelihood of STEM degree attainment for community-college student *i* assuming we could design an intervention that put the student on equal footing in all respects to an observationally similar university entrant. This aspect of the predictions contributes (likely strongly) to the upper-bound interpretation of our estimates.

The predicted values also embed an assumption about student interest in STEM among community college students: that their interest level in STEM fields is the same as among observationally similar university entrants, on average. Despite our inability to observe the interest levels of individual students, we embed this assumption in our empirical procedure by letting students' observable attributes translate to STEM success at the same rate regardless of which college sector they enter. It is unclear whether this is a reasonable assumption. Prior research shows that community college students are less interested in pursuing STEM degrees compared to their university peers unconditionally (Wang, 2013), but we are not aware of any evidence on conditional differences in STEM interest. If community college students are *conditionally* less interested in STEM fields, our assumption that their interest level is conditionally equal to that of university entrants, on average, would further contribute to the upper-bound interpretation of our estimates.

### *4.2 Policy Simulation*

We identify the subpopulation of community college students who are STEM qualified using the distribution of $\hat{P}_{ijt}$ among university STEM entrants. In our preferred set up, we identify student *i* as STEM qualified if $\hat{P}_{ijt}^{cc} > \tilde{\hat{P}}$, where $\tilde{\hat{P}}$ is the median predicted likelihood of STEM degree attainment among initial STEM entrants in the university sample. Whether the median value is an appropriate threshold is a normative question. On the one hand, we want a threshold high enough that students would have a reasonable likelihood of success in a STEM field. On the other hand, total STEM degree production is at least weakly increasing as the threshold falls. We use the median value among university STEM entrants as our primary threshold because it is an intuitive anchor for this value. We consider the sensitivity of our findings to modifications of this threshold below.

We assume that (a) all students respond to our hypothetical intervention by shifting to enroll at a university and (b) the shifted students succeed in STEM at the same rates as their observationally similar peers who initially enroll at universities. Under these assumptions, the number of university STEM degrees produced by our policy simulation is given by the sum of $\hat{P}_{ijt}^{cc}$ over the number of students with whom we intervene (denoted by $N_s^{cc}$):

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

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 summation to be over targeted groups of students only (e.g., female students). To obtain error bands for our estimates that account for error throughout the process described in this section, we bootstrap our entire procedure 500 times and report 95-percent empirical confidence intervals based on the bootstrapped values.[14]

## 5. Results

### *5.1 Primary Results*

Table 3 shows logit coefficients and bootstrapped 95 percent confidence intervals for equation (1) estimated on the university sample. As expected based on previous research, among the pre-entry academic qualifications, class rank is by far the strongest predictor of STEM success. The ACT math score is also a significant predictor of STEM success.[15] In terms of demographics, the familiar gender difference in STEM outcomes is clearly present in our data (Kahn and Ginther, 2017). Similarly, the well-documented lack of differences in outcomes by

[14] Our use of 500 bootstrap repetitions is based on a sensitivity analysis to alternatives (omitted for brevity). Because our outputs of interest are sample means of the predicted values, they are highly stable. Correspondingly, the sensitivity analysis shows that there is virtually no change in the mean values, or the ranges of our confidence intervals, with bootstrap repetitions above 500. In fact, even going from 500 down to 250 repetitions has no impact on our confidence intervals, but we use 500 bootstrap repetitions to be conservative.

[15] 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).

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 2-year and 4-year samples, we use $\tilde{\tilde{P}}$ as the threshold to identify STEM qualified students. Table 4 shows that about a fourth of all university entrants have academic qualifications that align with our definition of STEM qualified. Specifically, the bottom row of the table shows that out of 70,737 university entrants in our dataset, 18,509, or 26.2 percent, are STEM qualified. These students have stronger academic qualifications along all dimensions than their non-STEM-qualified peers. The magnitudes of the differences in core qualifications—ACT scores and percentile ranks—are large, ranging from 0.8–1.4 standard deviations of these variables.

STEM-qualified students are also more likely to be male and more likely to be Asian or White than non-qualified students, but for different reasons. At universities, 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—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 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 academic qualifications are sufficient to meet our definition of STEM-qualified is much smaller than the fraction of university students, at around 7.4 percent (i.e., 3209/43214, per the bottom row of Table 4). Moreover, their academic qualifications are clearly below those of their STEM-qualified peers at universities. This is because the distribution of observed academic readiness at community colleges is to the left of the distribution at universities. The implication is that among students above the threshold, those at community colleges are closer to the threshold value, on average, than their university peers. A notable result in Table 4 is that the STEM-qualified population at community colleges does not include a larger proportion of Black students than the STEM-qualified population at universities.

Our estimates for STEM degree production from the community college sample, based on equation (2), are reported in the first column of Table 5. We report the total number of university STEM degrees produced and the characteristics of completers. The second column replicates descriptive statistics for STEM completers among university entrants from Table 2 for comparison.

We focus first on the potential to expand the STEM pipeline, then turn to diversity. In total, recall from Table 4 that we move 3,209 STEM-qualified community college students to universities. Of these, we predict that 869 would complete a STEM degree within 6 years (in the second-to-last row of the table). Noting that there were 9,060 STEM degrees awarded to university entrants over the sample period (column 2), our estimate of 869 degrees produced corresponds to an increase of 9.6 percent. However, this is a gross estimate. In the raw data, 289 of the STEM-qualified community college students we identified completed a STEM degree without intervention via vertical transfer. Thus, on net, our hypothetical intervention increased the number of STEM degrees by 560 (869-289), or about 6.4 percent of observed university STEM degrees awarded. Whether this net increase is large or small is in the eye of the beholder, but our interpretation is that it is quite small given the scope of the intervention and the upper-bound assumptions built into it.[16]

Turning to the characteristics of the new STEM completers, column (1) shows that their academic qualifications on average are broadly similar to, but below, their peers who start at universities. There is also no diversity improvement by race-ethnicity among STEM degree recipients in the community college sample relative to the university sample. In fact, the fraction of community college students who complete a STEM degree and are Black (0.01) is substantially lower than the fraction of university entrants who complete a STEM degree and are Black (0.04). This is driven by the low share of Black students at the enrollment margin who attend community colleges (per Table 4).

Finally, the new STEM graduates from community colleges are even more male-dominated than university entrants. 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. Said differently, female community college students are more negatively selected in terms of academic qualifications, within their gender-specific distribution, than their male counterparts. We are not aware of previous documentation of this phenomenon in the literature and its generalizability merits attention in future research.

[16] Our net-increase calculations are the one place where the vertical transfer landscape in Missouri matters for our calculations. For instance, if modern improvements to transfer policies have made vertical transfer easier in Missouri, we would expect an updated version of our analysis to find even smaller net effects of shifting the enrollment margin because a larger number of initial community college enrollees would earn university STEM degrees via vertical transfer. Since these students are subtracted from the gross increase in STEM degrees to arrive at the net increase, the net increase will be lower. Said another way, ongoing improvements to the postsecondary system that reduce the frictions community college students face will lower the returns to an enrollment-shift policy. Our review of the literature on vertical transfer rates and post-transfer success suggests that the effect of recent improvements to Missouri's transfer policies is likely be modest, but ultimately this is an empirical question.

### *5.2 Robustness & Minor Extensions*

We conduct a large number of robustness tests and minor extensions of our analysis. We discuss the findings in this section and show the results in Appendix A.

First, we assess the predictive validity of our core prediction model in and out of sample among university entrants (for whom true outcomes are observed, which is required to test predictive validity). To facilitate the in-sample and out-of-sample tests, we use 80% of the university-entrant data to comprise the training dataset and the remaining 20% to test predictive validity. We follow the same general procedure described in the methodology section: 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 university entrants. In-sample and out-of-sample prediction accuracies are shown in Appendix Table A5 and confirm 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. Recall that we inflate the variance of the imputed ranks in order to offset the shrinkage inherent to the imputation process. This is important because without the inflation procedure, fewer students have rankings in the tails of the distribution and since we draw from the upper tail when we identify STEM qualified students, fewer students are identified. Appendix Table A6 shows that as predicted, failing to inflate the variance of the high school percentile ranks reduces the number of community college students identified as STEM qualified. In turn, this leads to fewer predicted STEM degrees.

In Appendix Table A7 we modify how we account for student demographics in the prediction model. 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. Appendix Table A7 shows results if we replicate our 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 show 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 Table 3, which shows that outside of Asian students, racial/ethnic designations are not important predictors of STEM completion conditional on academic preparation. In terms of gender diversity, the female share jumps markedly if we remove all information about gender from the prediction model. That is, if we ignore the preference among female students for non-STEM fields, 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 community colleges have fundamentally different preferences for STEM education than their university peers, which seems unlikely.

Next, we relax the data restrictions imposed on our preferred estimation sample. First, in Appendix Table A8 we relax the credit-hour and age restrictions, which are set to at least 12 first-semester credits and age $\leq 20$ 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 maximum entry age to 22 and then 24. These changes result in small increases in the numbers of students identified as STEM qualified, and correspondingly, small increases in STEM degrees produced. The changes are small because relatively few part-time or older community college students are STEM qualified based on their academic profiles.

In Appendix Table A9, we revisit to our decision to drop students who did not take the ACT prior to college enrollment. Incorporating students with missing ACT scores into the community college sample (by imputing the missing values) leads to a substantial increase in its size—the sample increases from 43,214 (per Table 1) to 57,382 students, a 33 percent increase. However, Appendix Table A9 shows this translates into only a very small increase in the number of students identified as STEM qualified and who complete STEM degrees. Indeed, the number of gross STEM degrees produced increases by just 115 relative to baseline. The reason is that community college students without ACT scores are negatively selected, by which we mean that their other attributes suggest lower academic achievement (on average). As such, when we impute their ACT scores, the imputed values are low.[17] Consistent with the spirit of our initial exclusion of students who do not take the ACT, adding these students back into our sample does not meaningfully impact our findings.

Our next extension separates out the field of Biology. Biology is one of the largest STEM majors in Missouri and nationally (Snyder, de Brey and Dillow, 2019), but differs from other STEM fields in that it is less mathematically oriented and biology degrees have lower earnings returns.[18] The lower earnings returns imply that compared to other STEM majors, there is less market demand for biology degrees. Therefore, it may be appropriate for policies designed to increase STEM degrees to focus on fields outside of biology. Appendix Table A10 shows results from a complete replication of our analysis that excludes biology majors from the list of STEM fields. The total increase in degree production among community college students is similar as a percent of current university production in non-biology STEM fields. Like in our main analysis,

[17] Note that like with the imputed class ranks, we inflate the variance of students' imputed ACT scores to match the variance of observed scores.
[18] 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.

racial/ethnic and gender diversity are lower 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 produced by our hypothetical intervention are awarded to women (Table 5), versus just 7 percent when we exclude biology (Appendix Table A10).

In Appendix Table A11 we consider the implications of setting the intervention threshold at a range of values between the 40th and 60th percentiles of the predicted likelihood of success among initial STEM entrants in the university sample, rather than at the median. As we lower the threshold we identify more students as STEM qualified, but the STEM success rate among those who shift falls because the marginally-induced students are less prepared academically. Alternatively, as we raise the threshold the likelihood of obtaining a STEM degree conditional on being identified as STEM qualified rises, but fewer students are identified so total degree production falls.

These results illustrate the tradeoff of varying the intervention threshold. 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 4-year degree. On this point, supplementary predictive models using our baseline settings indicate that among students who we intervene with but who fail to complete a STEM degree, about half (49 percent) would be expected to complete a non-STEM university degree based on their observable characteristics and qualifications, and the other half (51 percent) would fail to complete a degree within 6 years. These findings are reported on in Appendix Table A12, which shows outcomes, characteristics, and qualifications of students who we intervene with but do not ultimately complete STEM degrees. A notable result is that the non-STEM degree completers in Table A12—like their STEM-completer peers—are less racial-ethnically diverse than the current population of non-STEM university graduates.

## 6. A More Realistic Policy

Our simulation is useful for generating upper-bound impact estimates for policies that aim to shift the enrollment of community college students. But it is unrealistic. A more realistic policy would not be able to overcome all the barriers faced by community college students and put them on par with observationally similar university entrants. Perhaps a particularly well-crafted and well-funded policy could remove institutional barriers such as transfer frictions, and/or financial barriers, or at least greatly reduce them, but even then, personal barriers—including preferences and life barriers—would remain. If we were to add these barriers back into our framework, they would be operationalized through two channels. First, the rate at which we could change students' enrollment decisions would fall below 100 percent—i.e., some students who we identify as STEM qualified would not change their enrollment decisions in response to our hypothetical intervention. Second, conditional on an enrollment change, the students we shift would be less likely than observationally similar university entrants to complete STEM degrees.

In this section we modify our simulation to allow these barriers to impact student decisions and outcomes. First, we reduce the rate at which our intervention shifts enrollment to below 100 percent. Empirical evidence from real-world policies suggests the true rate would be well below 100 percent. If our intervention were a textbook nudge (Thaler and Sunstein, 2008), available research suggests that the behavioral response would be very small, in the range of 0-5 percentage points (Barr and Turner, 2018; Castleman and Page, 2015; Gurantz et al., 2021; Oreopoulos and Petronijevic, 2019). The response rate could be increased if the intervention was more substantial, such as if tuition subsidies or stipends were offered, but even then, available research suggests a modest behavioral response is likely.[19]

Based on this evidence, in Table 6 we rescale the intervention effect so that only 5 or 10 percent of students change their enrollment decisions, rather than all 3,209 students we identify as STEM qualified. For each scenario, we show results for two cases: one where the 5-10 percent of students who change their enrollment are a random sample and the other where they are the most likely to succeed in a university STEM program. To isolate the change in our results through this channel, we continue to assume that the students who respond would be equally likely to succeed as observationally similar university entrants after the enrollment shift. Table 6 shows that when the intervention take-up rate declines to plausible levels, there are large reductions in degrees produced and no diversity improvements.

Next, we revert to the baseline condition where we shift the enrollment of all students, but we allow community college students to underperform relative to their university-entrant peers after they shift. To do this, we begin by estimating the magnitude of *selection on observables* between 2-year and 4-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 because the distribution of academic qualifications among community college students is to the left of the university-entrant distribution. The difference in observed selection can be summarized by the average difference in the predicted likelihood of STEM success between the groups, which is 4 percentage points.

[19] For example, Deming and Walters (2017) find 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 four percentage points more likely to transfer to a university.

We assume that selection into college sector on unobservables is in the same direction as observed selection (a common assumption in the literature—e.g., see Oster, 2019) and consider magnitudes of unobserved selection from 50 to 300 percent as large as observed selection. Table 7 shows how our findings change as unobserved selection increases. The first row replicates our baseline condition with no unobserved selection. We intervene with 3,209 students and 27 percent of these students are predicted to complete a STEM degree. As unobserved selection becomes more severe, gross and net 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 7), the net number of STEM degrees produced declines by 141, to 449 degrees.[20]

The broad takeaway from this section is that if we impose real-world constraints, the impacts of our policy simulation decline from the upper-bound levels, sometimes markedly. Figure 1 illustrates this by showing estimates for total degree production under a range of evaluation scenarios.

## 7. Conclusion

Researchers and policymakers have shown great interest in identifying interventions that can expand and diversify the university STEM pipeline (and subsequently, the STEM workforce). We assess the upper-bound potential for influencing these objectives via policies that would infuse the university pipeline with academically qualified community college students. Our policy simulation can be broadly described as implementing an intervention that removes all barriers—ranging from financial to personal—that lead academically-qualified students to enroll in community colleges instead of universities, and correspondingly, removes

[20] The net number of degrees produced is always equal to the gross number minus the number of actual degrees observed among the intervention sample, which is 289.

all impacts of these barriers on educational attainment. We then ask how such an intervention would influence the university STEM pipeline.

We find that the potential to expand the number of university STEM degrees produced is modest. The exact magnitude of the change depends on assumptions and policy-design details, but under our preferred settings, we estimate our hypothetical intervention would generate a net increase in university STEM degrees of about 6.4 percent. This extensive-margin estimate can be made somewhat larger by modifying aspects of the intervention. More likely, though, it is well above what could be feasibly achieved through a real-world policy because of the upper-bound assumptions built into our analysis.

Our findings for diversifying STEM degree production are even less promising. Community college students are more racial/ethnically diverse than their peers at 4-year institutions overall. However, among the pool of community college students on the enrollment margin, racial-ethnic diversity is below diversity among existing bachelor's degree recipients in STEM fields. The implication is that the diversity of individuals who are predicted to earn STEM degrees from our community college sample is lower than among university entrants who already earn STEM degrees. We also find no scope for increasing the gender diversity of university STEM graduates. This result is partly tautological because we assume that female aversion to STEM is similar among 2-year and 4-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 in our data.

The broad takeaway is that policies and interventions at the point of college entry, and within the context of current STEM education programs, 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 student outcomes, and indeed there is evidence that for individual students with appropriate academic qualifications, shifts in enrollment from 2-year to 4-year institutions are beneficial (Goodman, Hurwitz, and Smith, 2017; Mountjoy, 2019). However, the potential for shifting the enrollment of community college students in a way that meaningfully impacts overall STEM degree production at universities seems limited.

Our analysis is based on administrative microdata from Missouri. Several aspects of the data make them well-suited for our evaluation, but the generalizability of our findings is an open question. A notable Missouri-specific factor that plays an important role in our analysis is the student success rate in Missouri's university STEM programs, particularly for students who are marginally qualified. We treat the efficacy of Missouri's university STEM programs as a fixed contextual feature of our evaluation. While we have no reason to believe Missouri differs from other states in this regard, if there are differences, it could lead to different conclusions in other states. The extent to which this and other factors influence the generalizability of our findings merits attention in future research.

While disappointing, what our findings suggest will not work to bolster the STEM pipeline can be used to refocus efforts on other strategies that may be more promising. Two alternative strategies are most readily apparent. The first is to improve postsecondary institutional quality, about which our study is silent because we do not allow for changes in institutional quality in our simulation. If universities could improve their ability to leverage the academic potential of their students, the STEM pipeline would expand and perhaps diversify as well, depending on the nature of the improvement. This could involve improvements to skill development during college or improvements in fostering interest in STEM fields. This is not a trivial task (Hrabowski, Rous, and Henderson, 2019), but our analysis suggests it may be necessary if meaningful changes to the STEM pipeline are desired.[21] Relatedly, although community colleges are not focal in our analysis (other than for providing the intervention pool), increasing the prevalence of effective STEM initiatives in community colleges could provide the necessary real-world supports to boost student success (Hagedorn and Purnamasari, 2012). Indeed, the 289 community college students in our sample who we observe ultimately earning university STEM degrees are proof-of-concept data points that the community college pathway is viable. The other strategy would involve improvements to educational efficacy prior to college, which would result in students with better academic qualifications and/or stronger interests in STEM at the point of college entry. This would also expand the STEM pipeline and may increase diversity—again depending on the nature of the improvements—but will be challenging.

[21] Although we do not pursue this extension here, one could incorporate the impact of increases in institutional quality into our framework by imposing *ad hoc* increases in STEM degree completion rates conditional on observed student qualifications at entry.

**References**


Adelman, C. (2006). The toolbox revisited: Paths to degree completion from high school through college. Washington, DC: U.S. Department of Education. Retrieved 10.11.21 at https://files.eric.ed.gov/fulltext/ED490195.pdf

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., & 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.

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.

Belasco, A.S., and Trivette, M.J. (2015). Aiming low: Estimating the scope and predictors of postsecondary undermatch. *Journal of Higher Education* 86(2), 233-63.

Bettinger, E., P., Evans, B.J., and Pope, D.G. (2013). Improving college performance and retention the easy way: Unpacking the ACT exam. *American Economic Journal: Economic Policy* 5(2), 26-52.

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.

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.

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.

Cohodes, S.R., and Goodman, J.S. (2014). Merit aid, college quality, and college completion: Massachusetts' Adams Scholarship as an in-kind subsidy. *American Economic Journal: Applied Economics* 6(4), 251-85.

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.

Dougherty, K. (1987). The effects of community colleges: Aid or hinderance to Socioeconomic attainment. *Sociology of Education* 60(2), 86-103.

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.

Evans, W.N., Kearney, M.S., Perry, B., and Sullivan, J.X. (2020). Increasing community college completion rates among low-income students: Evidence from a randomized controlled trial evaluation of a case-management intervention. *Journal of Policy Analysis and Management* 39(4), 930-65.

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

Gurantz, O., Howell, J., Hurwitz, M., Larson, C., Pender, M., & White, B. (2021). A national-level informational experiment to promote enrollment in selective colleges. *Journal of Policy Analysis and Management* 40(2), 453-79.

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.

Hrabowski III, F.A., Rous, P.J., and Henderson, P.H. (2019). *The Empowered University*. Baltimore, MA: Johns Hopkins University Press.

Hagedorn, L.S., and DuBray, D. (2010). Math and science success and nonsuccess: Journeys within the community college. *Journal of Women and Minorities in Science and Engineering* 16(1), 31-50.

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.

Ignash, J. M., & Townsend, B. K. (2000). Evaluating state-level articulation agreements according to good practice. *Community College Review*, 28(3), 1-21.

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

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

Long, B.T., and Kurlaender, M. (2009). Do community colleges provide a viable pathway to a baccalaureate degree? *Educational Evaluation and Policy Analysis* 31(1), 30-53.

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.

Nathan, M., and Lee, N. (2013). Cultural diversity, innovation, and entrepreneurship: Firm-level evidence from London. *Economic Geography* 89(4), 367-94.

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.

Oster, E. (2019). Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics* 37(2), 187-204.

Østergaard, C.R., Timmermans, B., and Kristinsson, K. (2011). Does a different view create something new? The effect of employee diversity on innovation. *Research Policy* 40, 500-509.

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*.

Rivera, L.A. (2015). *Pedigree: How Elite Students Get Elite Jobs*. Princeton, NJ: Princeton University Press.

Rouse, C.E. (1995). Democratization or Diversion? The effect of community colleges on educational attainment. *Journal of Business & Economic Statistics* 13(2), 217-24.

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

Scott-Clayton, J. (2015). The shapeless river: Does a lack of structure inhibit students' progress at community colleges? In S. Baum, B. L. Castleman, and S. Schwartz (Eds.), *Decision Making for Student Success: Behavioral Insights to Improve College Access and Persistence*. London, UK: Routledge

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*.

Smith, J., Pender, M., and Howell, J. (2013). The full extent of student-college academic undermatch. *Economics of Education Review* 32, 247-61.

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

Somers, P., Haines, K., Keene, B., Bauer, J., Pfeiffer, M., McCluskey, J., Settle, J., and Sparks, B. (2006). Towards a theory of choice for community college students." *Community College Journal of Research & Practice* 30, 53–67.

Stephens, N.M., Fryberg, S.A., Markus, H.R., Johnson, C.S., & Covarrubias, R. (2012). Unseen disadvantage: How American universities' focus on independence undermines the academic performance of first-generation college students. *Journal of Personality and Social Psychology* 102(6), 1178-1197.

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.

Turley, R.N.L. (2009). College proximity: Mapping access to opportunity. *Sociology of Education* 82, 126-146.

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.

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

Xu, D., Jaggars, S.S., Fletcher, J., and Fink, J.E. (2018). Are community college transfer students "a good bet" for 4-year admissions? Comparing academic and labor-market outcomes between transfer and native 4-year college students. *Journal of Higher Education* 89(4), 478-502.

**Tables and Figures**

***Table 1: Summary statistics for 2-year and 4-year 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 4-year 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 4-year 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 & -gender interactions | X |
| Percentile Rank-Race/Ethnicity & -gender interactions | 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 2-year and 4-year institutions.*

| | Four Year University | | Community College | |
|---|---|---|---|---|
| | (1) STEM qualified | (2) Not STEM qualified | (3) STEM qualified | (4) 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 universities compared to observed STEM completers at universities.***

| | (1)<br>Graduate with STEM | (2)<br>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 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: Summary statistics for STEM-qualified community college students with 5/10-percent intervention 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] | |

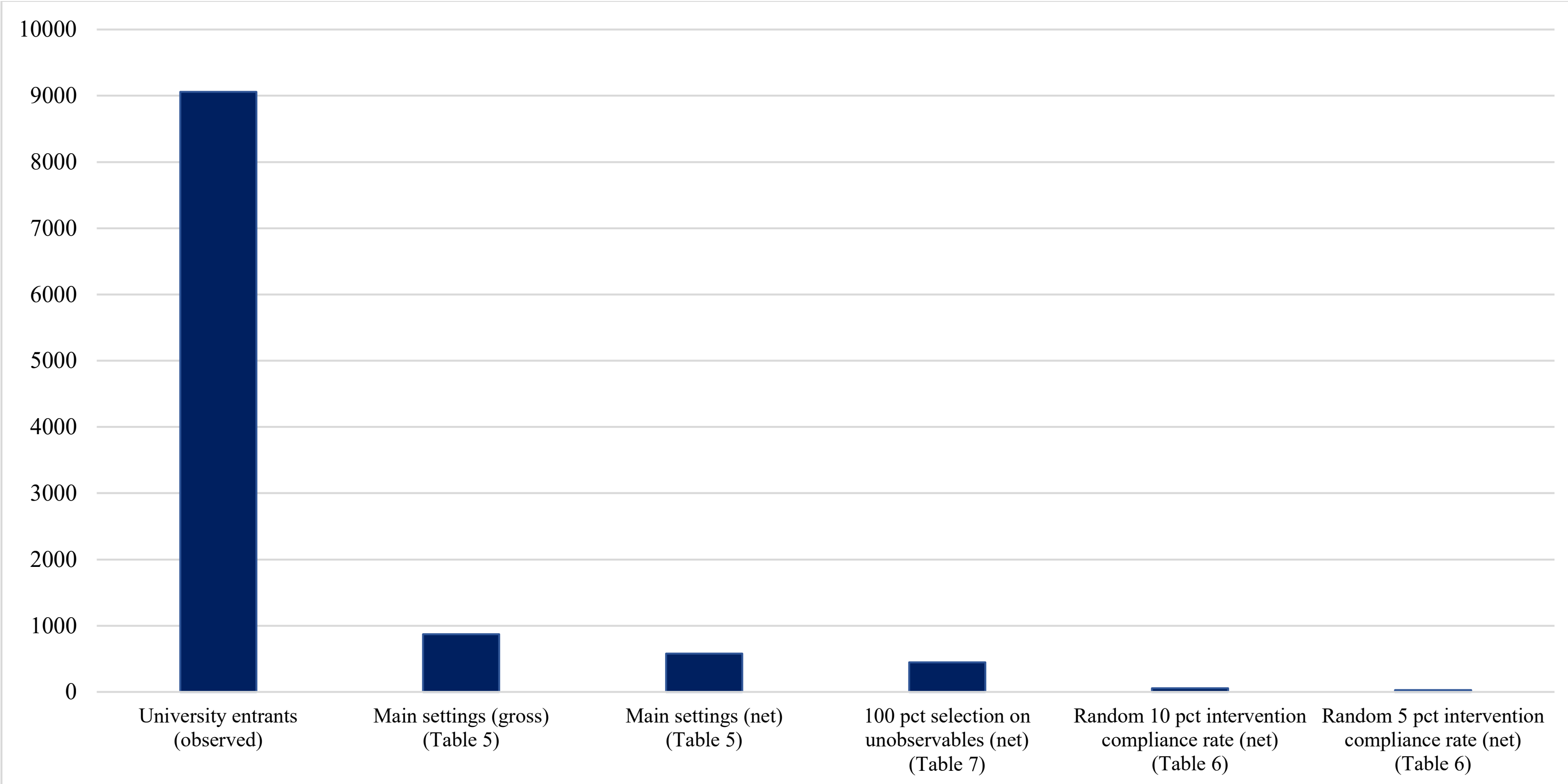
Notes: Table reports averages and 95 percent confidence intervals over 500 bootstrap repetitions when 5/10 percent STEM-qualified community college students actually choose to enroll at universities. Randomly selected 5/10 percent in column (1) and (2)/ (5) and (6); top 5/10 percent in terms of prediction likelihoods in column (3) and (4)/(7) and (8). The net-degrees-produced numbers are equal to the gross degrees produced minus the actual number of degrees observed among the STEM-qualified sample that we identify in each scenario.

***Table 7: Summary statistics for STEM qualified community college students: different levels of selection on unobservables***

| (1) Number of Community College Students | (2) Average STEM Completion Likelihood among University STEM-Qualified Entrants | (3) Community College Student Selection on Observables | (4) Selection on Unobservables | (5) Average Likelihood of STEM Degree for Community College Students | (6) # of STEM Degrees Produced (gross/net) |
|---|---|---|---|---|---|
| 3209 | 0.31 | -0.04 | 0 | 0.27 | 869/580 |
| 3209 | 0.31 | -0.04 | -0.02 (50%) | 0.25 | 802/513 |
| 3209 | 0.31 | -0.04 | -0.04 (100%) | 0.23 | 738/449 |
| 3209 | 0.31 | -0.04 | -0.08 (200%) | 0.19 | 610/321 |
| 3209 | 0.31 | -0.04 | -0.12 (300%) | 0.15 | 481/192 |

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 values are calculated from the average likelihoods of graduating in STEM corresponding to equation (1). Column (5)= column (2)-column (3)-column (4). Column (6) (gross)= column (1)*column (5). The net calculations subtract the 289 individuals who we observe transferring and earning a STEM degree from the gross production numbers.

***Figure 1. STEM Degrees Produced by the Missouri Public University System Under Various Parameterizations of our Hypothetical Intervention.***



**Appendix A: Supplementary Tables**

***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 |
| | Records lost | Remaining sample | | Records 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

***Table A2. Summary statistics for the university sample.***

| | (1)<br>All | (2)<br>In state | (3)<br>In state full time | (4)<br>In state full time, age<=20 | (5)<br>In state full time, age<=20, in state hs | (6)<br>In state full time, age<=20, in state hs, nomissing ACT | (7)<br>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 | 0.70 | 0.70 | 0.70 | 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 |

*Table A3. Summary statistics for the community college sample.*

| | (1) All | (2) In state | (3) In state full time | (4) In state full time, age<=20 | (5) In state full time, age<=20, in state hs | (6) In state full time, age <=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 |

***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] |

***Table A5. In-sample and out-of-sample predictive validity in 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.

***Table A6. 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 A7. Robustness of findings to dropping race-gender indicators and/or race-gender indicator interactions in the model that predicts university STEM 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 A8: 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) | (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 | 25.24 | 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 A9: Robustness of findings to removing the ACT-taking requirement among community college students.***

| | 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 A10: 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 A11: Findings using different intervention thresholds for identifying STEM-qualified community college students based on the percentile of the distribution among 4-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 at different intervention thresholds.

***Table A12: Summary statistics for STEM-qualified community college students in supplementary predictive models who are not predicted to earn STEM degrees.***

| | (1)<br>Graduate with Non-STEM | (2)<br>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 community college students who 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 4-year degree (within six years). These models are of the same structure as equation (1):

$$P_{ijt}^{*} = \boldsymbol{X_i}\boldsymbol{\beta_1} + \gamma_{1j} + \delta_{1t} + \varepsilon_{1ijt} \tag{B1}$$

In equation (B1), $P_{ijt}^{*}$ 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 A12, 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.