NBER WORKING PAPER SERIES

UNDERSTANDING GAPS IN COLLEGE OUTCOMES BY FIRST-GENERATION STATUS

Esteban M. Aucejo
Jacob French
Paola Ugalde Araya
Basit Zafar

Working Paper 34129
http://www.nber.org/papers/w34129

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
August 2025

We thank various seminars and conference participants for valuable feedback and comments. We also thank the University Office of Institutional Analysis at ASU for providing access to the data. All errors that remain are ours. The research in this paper has been approved by the IRB at Arizona State University The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Understanding Gaps in College Outcomes by First-Generation Status
Esteban M. Aucejo, Jacob French, Paola Ugalde Araya, and Basit Zafar
NBER Working Paper No. 34129
August 2025
JEL No. D83, I23, I24

ABSTRACT

Information frictions significantly shape students' academic trajectories, but their differential impact across student backgrounds remains understudied. Using a novel panel survey capturing incoming students' subjective expectations and anonymized transcript data from Arizona State University, we first show that parental education strongly predicts educational success, even after controlling for demographics and measurable college preparation. First-generation students enter college less informed and with more uncertain beliefs, facing substantial challenges stemming from limited understanding and uncertainty about the higher education setting. A Bayesian expected utility maximization model demonstrates that higher uncertainty alone can sustain persistent achievement gaps. Empirically, students update their beliefs and make academic decisions consistent with the model's predictions. Finally, leveraging a natural experiment involving a targeted first-year experience program for academically marginal students, we demonstrate that cost-effective interventions can successfully reduce knowledge frictions, improve retention, and encourage beneficial early major switching.

Esteban M. Aucejo
Arizona State University
Department of Economics
and NBER
Esteban.Aucejo@asu.edu

Jacob French
New York University
Leonard N. Stern School of Business
Economics
jff305@nyu.edu

Paola Ugalde Araya
Louisiana State University
Department of Economics
mugaldea@asu.edu

Basit Zafar
University of Michigan
Department of Economics
and NBER
basitak@gmail.com

1 Introduction

The persistent gaps in college completion among different sociodemographic groups are significant and well-documented. Notably, first-generation students — students without a college-educated parent — are substantially less likely to earn a bachelor's degree within six years of entering a four-year institution: 44% compared to 74.3% of their continuing-generation peers (National Center for Education Statistics, 2019). This disparity is particularly concerning because differences in college graduation rates contribute to growing wage inequality (Autor et al., 2020). While an extensive body of research investigates socioeconomic status (SES) differences in college *enrollment*, there is comparatively less focus on SES disparities in outcomes *post-enrollment*.[1] Yet these divergent post-enrollment experiences are crucial components of the composite college completion gap, as nearly half (47.9%) of undergraduates enrolled in four-year colleges during the 2019–2020 academic year were first-generation students (National Center for Education Statistics, 2022). Understanding the drivers of these completion disparities is critically important for designing effective policy interventions that can broaden educational opportunity and promote equity across socioeconomic groups

This paper attempts to improve our understanding of why outcomes in college diverge by first-generation status using a unique combination of rich (administrative and panel survey) data and a quasi-experimental intervention targeted at students early in college. The setting of our study is Arizona State University (ASU), one of the largest public universities in the US. With its large share of first-generation students, ASU offers an ideal context to understand the challenges faced by those from diverse academic backgrounds. ASU's charter emphasizes its commitment to "accessibility for all students qualified to study at a research university," a value reflected in the university's high acceptance rate. This accessibility means that ASU's student body is more likely to represent the marginal college enrollee—a critical group for addressing educational inequalities.

Starting with an analysis of anonymized student-level transcripts, we uncover key patterns in the outcomes of first-generation students. Even after flexibly accounting

[1] For example, only 37% of first-generation students attend 4-year colleges within eight years of high school graduation, versus 88% for students whose parents have a bachelor's degree or higher (Papay et al., 2015; Bui and Rush, 2016). The various explanations put forward for these enrollment gaps include differences in: credit constraints (Lochner and Monge-Naranjo, 2011), biases in information about college returns (Bleemer and Zafar, 2018), and perceived pecuniary and non-pecuniary returns to college enrollment (Boneva and Rauh, 2020).

for the joint distribution of characteristics – demographics, household income, initial major choices, early college performance, and various measures of college preparation – parental education remains a significant predictor of academic success. First-generation students are significantly less likely to graduate and more likely to drop out. This disparity in dropout rates is primarily a feature of the first academic year; conditional on continued enrollment after the first year, dropout rates are fairly similar by first-generation status. In addition, conditional on dropping out, first-generation students are less likely to switch majors. While dropping out may be the optimal strategy for some students, this suggests that first-generation students are less likely to explore alternative majors before dropping out of college. Our survey data reveal that, at the point of entry, first-generation students tend to be more over-optimistic about their educational prospects (relative to their continuing-generation counterparts), conditional on their level of preparation. This over-optimism is accompanied by greater uncertainty regarding their expected performance and less use of academic policies and resources, such as course withdrawals, that could help protect their GPA. In addition, our survey data reveal that first-generation students are ex ante less willing to change their field of study and less integrated with their social networks (i.e., they know fewer peers and are less likely to turn to family members for support when facing academic challenges). We argue that these information frictions collectively represent a knowledge barrier, which is a first-order impediment that first-generation students face in successfully navigating higher education.

To provide a more formal characterization of how knowledge frictions may impact student decision-making, we outline a simple model of re-enrollment that combines Bayesian belief updating with a basic expected utility model. Students do not have perfect information about their subjective returns to graduating in a given major, but can learn about these returns after enrollment through signals such as course grades. After updating their priors, but before finishing university, students must make a re-enrollment decision that allows them to switch majors or drop out. Through this stylized framework, we show that more uninformed students, i.e., those who enter college with higher initial uncertainty about the subjective returns to their enrolled major, undergo a more substantial update of their academic beliefs throughout their studies. We allow students to learn about the returns to alternative majors imperfectly. So, a negative signal may push a student to switch majors or even drop out if the signal is sufficiently negative. The key insight of the model is that two students who only differ in their ex-ante uncertainty will have different graduation probabilities, with the more uncertain student being less likely to graduate. The model posits that this difference in graduation rates should be concentrated among students with moderately poor first-year grades. Within this group, the model predicts that more uncertain students are more likely to drop out and less likely to switch majors.

The longitudinal nature of our surveys and transcript data allows us to test several key implications of our model for observationally similar first- and continuing-generation students. First, as already discussed, first-generation students have higher dropout and lower graduation rates (even after conditioning on the joint distribution of other observables). Second, conditional on receiving a positive (negative) grade signal, first-generation students update their subjective academic expectations (i.e., expected GPA) upward (downward) more than their continuing-generation peers. Third, first-generation students assign more weight to (early) grades when forming *dropout* expectations. Fourth, first-generation students are more likely to drop out and less likely to switch majors only when they receive negative grade signals. Fifth, the higher dropout and lower major-switching rates are specific to intermediate (not extremely poor) negative grade signals. The fact that heterogeneity in uncertainty alone is able to match such a rich set of empirical patterns in both reported beliefs and real-world academic decisions highlights knowledge frictions as a key candidate to explain first-generation students' lagging academic achievement.

Expanding on the model implications, the breadth of our transcript data allows us to employ a Coarsened Exact Matching (CEM) approach that accounts for the possibility that observable differences in first-generation college decisions are driven by factors such as household income, financial aid, levels of student preparation, performance in the first semester at ASU, or sorting into majors. Using this non-parametric approach, we find a notable disparity in dropout rates. First-generation students who encounter negative grade events have about a 40% likelihood of dropping out, which is around five percentage points higher than observationally identical continuing-generation students who face the same academic setback. Interestingly, there is no significant difference in dropout rates between similarly matched first-generation and continuing-generation students who do not experience such early academic challenges. Rather than dropping out, we find that continuing-generation students who face academic difficulties in their first year are more likely to switch majors.[2] Moreover,

[2] ASU students have to declare their field of study at the time of initial enrollment.

conditional on continuing into the second year at university, the major-switching rate among students who receive negative academic signals is statistically identical across first-generation status. We interpret this as evidence that the marginal first-generation student is facing a decision between dropping out and switching majors, as our model would predict; the empirical patterns suggest that some first-generation students are dropping out prematurely, even though they might have successfully pursued a different major.

While identifying specific knowledge frictions in the data is quite challenging due to their unobservable nature, the collective weight of evidence presented in this study provides a convincing picture of the role of these frictions in shaping the outcomes of first-generation students. More specifically, our results strongly suggest that differences in baseline uncertainty and environmental knowledge, combined with initial (negative) information signals, likely contribute to socioeconomic achievement gaps. In particular, our model and empirical evidence agree that the margin between switching majors and dropping out after experiencing academic setbacks is important to understanding how these knowledge frictions translate into academic outcomes. Given the sequential nature of academic decisions and the narrow window to intervene, our evidence suggests that targeted interventions to provide additional support and guidance early in the university experience may narrow achievement gaps.[3] To test this assertion, we examine a program at ASU named "Learn Explore Advance Design" (LEAD), which targets first-time students with weaker incoming academic credentials. Although not explicitly designed for first-generation students, the LEAD program contains a higher proportion of first-generation students than the overall ASU population. The LEAD curriculum replaces standard first-year courses, allowing participants to earn required credits through smaller, more interactive, and homogeneous classes, and provides additional coursework to build soft skills (such as time management, presentation skills, communication, and critical reading and thinking). In addition, participants benefit from peer and faculty/staff mentorship, aimed at creating closer interactions with professors and fostering a supportive learning community. Importantly, there is experimental variation in program take-up that we exploit to estimate causal program effects. We find that program participation leads to a sizable increase in the likelihood of graduating. In line with the model

[3] Most dropout, not surprisingly, happens early in students' college experience. Of all students who end up dropping out, 42% have dropped out before the start of the second year.

mechanisms, we find evidence that LEAD participants are more likely to switch majors and are less likely to switch into an undecided major, both indicators that the program provides useful information to students about how to navigate important academic decisions. Qualitative evidence from our surveys suggests that LEAD students are more likely to report that ASU helps them enhance their campus network as compared to non-LEAD students, indicating that the program can help students overcome initial information barriers. Together, this evidence suggests that programs such as LEAD may be able to alleviate the types of knowledge barriers we identify as contributing to first-generation achievement gaps. LEAD is a bundled treatment, so we cannot precisely determine which aspects of the program make it effective. However, our findings are encouraging, especially given that many college interventions have been shown to be either ineffective or only marginally effective (Angrist et al., 2009, 2014; Alamuddin et al., 2018; Oreopoulos, 2021).[4] In addition, successful programs generally have been quite intensive and expensive, which makes them infeasible for scalability (e.g., Scrivener et al., 2015; Carrell and Sacerdote, 2017). LEAD, on the other hand, is fairly cost-effective as it largely replaces the standard first-semester curriculum and leverages low-cost peer mentorship.

To our knowledge, this is the first paper to provide a comprehensive picture of why educational trajectories in college differ by first-generation status. Some of the facts that we document are not new. For example, prior work has shown that family background and race are systematically related to college major choice and major-switching patterns (Baird et al., 2016; Leighton and Speer, 2024), and that incoming preparation is related to college completion (Bound et al., 2010). Likewise, it is known that a significantly lower percentage of first-generation students utilize academic advising or support services compared to their peers with college-educated family members.[5] Our findings show that the challenges first-generation students face extend beyond the commonly cited factors such as credit constraints and academic preparation, indicating that first-generation achievement gaps cannot be closed by addressing these traditional barriers alone.[6]

[4]However, in line with our findings, it is important to note that several studies have shown that proactive counseling can improve college student outcomes; see Dynarski et al. (2023) for a comprehensive review of the literature on college access and success.

[5]While 72% of continuing-generation students sought academic advising services, only 55% of first-generation students did so, according to the Center for First-Generation Student Success (see `https://firstgen.naspa.org/files/dmfile/NASPA_FactSheet-03_FIN.pdf`).

[6]This is consistent with recent work documenting a class gap (as defined by parental education)

Our paper is also related to the literature on information frictions and how students learn (see Altonji, 1993; Arcidiacono, 2004; Zafar, 2011; Stange, 2012; Stinebrickner and Stinebrickner, 2012, 2014; Thomas, 2019; Arcidiacono et al., 2020; Larroucau and Rios, 2022). However, to date, there remains limited direct evidence on how the learning process varies by family background. This is largely because existing survey evidence on students comes from settings that are either selective (e.g., Arcidiacono et al., 2012; Zafar, 2013), or composed predominantly of disadvantaged students (e.g., Stinebrickner and Stinebrickner, 2012).[7] Our setting, with a large and fairly representative college-going population, allows us to fill this gap in the literature.

One implication that follows directly from our findings is that providing informative signals to students early in college can help mitigate gaps (since first-generation students arrive in college with less information and more uncertainty). In fact, Aucejo and Wong (2025) provide evidence that first-generation students can improve their academic performance when they receive personalized feedback on interpreting negative academic shocks—such as discrepancies between expected and actual course performance—along with encouragement from professors. In light of the findings in our paper, we believe that understanding the dynamics of information acquisition and processing is essential for developing targeted interventions that can better support disadvantaged students and help close the educational gaps.

The remainder of this paper is structured as follows. Section 2 outlines the various data sources utilized in our analysis. In Section 3, we provide a detailed characterization of first-generation students at ASU, highlighting key demographic and academic differences. Section 4, using survey data, examines the presence of knowledge frictions and their impact on students' academic trajectories. In Section 5, we introduce a model to formally analyze these knowledge frictions and their implications for student outcomes. Section 6 presents empirical evidence that validates the predictions derived from our model. Section 7 evaluates the effectiveness of the LEAD program. Finally, Section 8 concludes.

in US tenure-track academia (Stansbury and Rodriguez, 2024). They attribute part of the gap to a lack of social and cultural capital.

[7]One exception is Qiu (2025), who, using data from the centralized college application system in China, documents how information gaps contribute to socioeconomic differences in major choice and long-term outcomes.

2 Data

2.1 ASU Transcript Sample

We use anonymized transcript-level data for 145,000 first-time in-person Arizona State University (ASU) freshmen. The administrative data covers the period from 2000 to 2022. It traces the trajectory of students as they progress through their college careers, including information about courses taken, grades obtained, and major switches. Notably, these data are linked with the National Clearinghouse (NSC), so we can separately identify transfers to another US university from college dropouts (as long as they move to an institution that is included in the database). Data on parental education and household income are sourced from FAFSA applications, covering approximately 62% of our sample. Appendix Table A1 shows how this selection impacts our sample. Students with known parental education are more likely to be a racial minority, have lower standardized test scores, and are less likely to be enrolled in the honors college. FAFSA applications are required to receive financial aid. While we cannot directly test this assertion, conditioning on FAFSA application likely yields a more socioeconomically comparable sample of first- and continuing-generation students.

Table A2 shows how ASU students compare to the average undergraduate student at other large flagship universities (specifically, the largest public universities in each state). Although the gender distribution at ASU is similar to that of other flagship institutions, the same is not true for its racial distribution. ASU has more Hispanic students (28% vs 12%), slightly fewer black students, and significantly fewer white students than the average flagship university.

ASU is well-suited for studying differences across demographic groups as it is one of the largest public universities in the country. The share of underrepresented minorities (URM) (30%) and first-generation students – defined as those without a college-educated parent – (32%) is relatively high. At the most selective schools, these shares are typically much lower: first-generation students make up only about 20% at top 20 institutions (Ivy Coach, 2024), and 27.5% of U.S. freshmen in 2012 (National Center for Education Statistics, 2019) and 27.6% of U.S. undergraduates in 2019 (National Center for Education Statistics, 2022) in the most selective category of colleges, as defined by National Center for Education Statistics (NCES). For reference, 47.9%

of undergraduate students enrolled in four-year institutions were first-generation in 2019-2020 (National Center for Education Statistics, 2022). It is worth noting that ASU admits most of its applicants regardless of their academic credentials, with an acceptance rate of 90% in Fall 2023 (National Center for Education Statistics, 2025). Thus, as shown in Table A2, it is unsurprising that its ACT 25th and 75th percentile scores are lower than those at the average flagship university. Therefore, ASU students are more likely to represent the marginal college enrollees, which is the student population to target if we want to close socioeconomic gaps in educational attainment. Table A3 compares first-year first-generation students at ASU to their counterparts at other four-year institutions using the Beginning Postsecondary Students Longitudinal Study 12/17 (BPS 12/17). The proportion of first-generation students at ASU is lower (37.5% vs. 48%), though the gender distribution among first-generation students is similar. ASU's first-generation students include fewer White and Black students and a higher share of Hispanic students. They also have higher ACT scores and first-year GPAs compared to the national average for first-generation students.

2.2 Survey Sample

We combine transcript data with panel student survey information. Our primary data come from original surveys fielded to in-person undergraduate students at ASU. The data collection took place during the school year 2021-2022 in two rounds.[8]

For the first round, in-person first-year students (around 13,000) were invited to participate via email early in the fall semester (in September 2021). Additionally, the study was advertised by the professors from the LEAD program in their classes. As compensation, participants entered a lottery for one of 600 $15 Amazon gift cards and one of 2 $250 Amazon gift cards. A total of 1,395 students completed the survey, with a median completion time of 30 minutes.

Data collection for the second round of surveys started in February 2022. A total of 7,748 freshmen students were invited to participate. Invitations were distributed via email and the ASU mobile app. Additionally, LEAD professors advertised the study in their classes. Respondents entered a lottery for one of 500 $10 Amazon gift cards. A

[8]The survey instruments can be accessed online here: Round 1 and Round 2.
It is worth noting that we also leverage a small amount of data from a separate 2023 survey of incoming first-year students. In particular, the data in Appendix Table A11 are taken from this survey.

total of 785 students completed the second round survey. The median completion time for the follow-up survey was 17 minutes. 57% of second-round survey participants had also completed the first-round survey, which resulted in a panel of 448 students across both surveys. Given that the sample for the second survey was targeted at students with a relatively higher ex-ante likelihood of dropout, the proportion of new respondents in the second survey (that is, those who had not taken the first survey) who had a high school GPA of below 3.5 was 59%, much higher than the corresponding share of 18% in the first survey sample. To improve response quality, we further restrict the sample to respondents who reported that they "somewhat" or "strongly" agreed with the statement "I answered all questions the best I could" at the end of each survey. This restriction reduces our panel sample from 448 to 423 respondents.

Table 1 compares the survey samples to the ASU 2021 freshmen cohort. The samples from the first and second rounds are denoted by R1 and R2, respectively. Columns (3) and (5) compare R1 and R2 samples to the 2021 freshman cohort. Both samples have a higher proportion of white and female students. Additionally, students come from families with lower incomes than the average ASU freshmen. The R1 sample specifically seems to be positively selected academically, with students being more likely to be in the Honors program, less likely to have missing ACT/SAT scores, scoring higher in admission tests (ACT/SAT), and earning better grades in high school than the average 2021 freshman student at ASU. While the selection is less pronounced in the R2 sample, qualitatively similar patterns hold for the subsample of students who participated in both rounds (column 7) when compared to the ASU average freshmen in 2021 (column 8).

The last column of the table analyzes selection into participating in both rounds by comparing the observable characteristics of the participants who completed both surveys to those who participated in R1. The average student who participated in both surveys is more likely to be an Honors student, and has higher admission scores and a higher high school GPA. However, it is not the case that students who participated in both surveys had systematically different beliefs about their future academic performance than the students who completed R1. This can be seen in the last three rows of the table. We see that there is no statistically significant difference between these two groups' beliefs about the likelihood of returning for the second year, the expected GPA after the freshman year, and the likelihood of graduating in

4 years.

3 Characterization of First-Generation Students

In this section, we characterize the population of first-generation students and document the substantial academic achievement gaps they face. Drawing on the breadth of our administrative data, we show that baseline differences in academic preparation, socioeconomic status, and other observable demographic characteristics cannot fully explain these gaps.

3.1 Differences Between First- and Continuing-Generation Students

We begin by comparing the entering characteristics of first-generation students to those of their continuing-generation peers. Panel A of Table 2 presents average observable characteristics of all first-time undergraduates who entered ASU between 2000 and 2017, disaggregated by parental education in columns (1) and (2), with the raw differences reported in column (3).[9]

Among the most pronounced disparities in Panel A, first-generation students have significantly lower household incomes, averaging about $63,000 compared to $118,000 for their peers (in 2019 dollars). Correspondingly, they are 16 percentage points (pp) more likely to be in-state students. In terms of racial and ethnic composition, first-generation students are less likely to identify as White or Asian and more than twice as likely to identify as Latino/Hispanic. They are also 5pp more likely to identify as female. With respect to academic preparation, first-generation students enter college with weaker academic credentials. While their high school GPAs differ by only 0.03 grade points—a negligible gap—first-generation students score about 2.2 points lower on the ACT-equivalent scale, roughly half a standard deviation.[10] First-generation students are more likely to be on financial aid (grants or loans) during their first year,

[9]Restricting the last entry cohort to 2017 allows us to observe outcomes for that cohort for five years.

[10]The first column in Table A4 shows that high school GPA is a much weaker predictor of ACT scores for first-generation students compared to their continuing-generation peers. This pattern suggests that first-generation students are more likely to have attended high schools where GPAs may be inflated or less informative of standardized test performance.

though the likelihood they have loans is similar to that of their counterparts. Finally, the last row of Panel A shows that about a third of first-generation students come in with transfer credits, similar to the share of continuing-generation students.

Given these observable differences, it is perhaps unsurprising that first-generation students underperform academically on a wide range of outcomes, as shown in Panel B of Table 2. They are 6pp (9%) less likely to graduate at all, 7pp (13%) less likely to graduate within four years, and 8pp (47%) more likely to drop out; note that students who leave ASU and enroll elsewhere within 12 months are recorded as transfers rather than dropouts. Those who eventually drop out tend to do so 0.12 years (approximately 0.24 semesters) earlier than their continuing-generation peers. Moreover, first-generation students are 27% less likely to earn STEM degrees. Strikingly, these disparities are at least as large as gaps associated with family income or minority status (see Tables A5 and A6).

Our rich administrative data also allows us to explore how differences in transcript-level decisions and behavior may contribute to these patterns. Panel C of Table 2 focuses on indicators of academic choices and performance. First-generation students are 1pp more likely to switch majors at least once, and this gap widens to 3pp among those who eventually graduate.[11] Notably, among students who drop out, first-generation students appear less likely to experiment with different fields of study before leaving. Those who persist toward graduation take 0.18 semesters longer, on average, to settle on a final major. Moreover, first-generation students are 1pp (20%) more likely to revert to an undecided major from a previously declared one.[12] Finally, there is little difference in first-generation students' initial preference for STEM majors, with a 1pp STEM gap at entry (3.8% of the continuing-generation STEM enrollment rate).[13]

Turning to markers of academic performance, first-generation students face substantially greater academic challenges. They are 8pp (38%) more likely to be placed on academic probation at some point, reflecting a higher risk of falling below the

[11] Major switching is defined as moving from one declared major (at the 6-digit CIP code level) to a different declared major or to an undecided status. Transitions from undecided to a declared major are not counted as switches.

[12] ASU actively discourages students from enrolling without a declared major, resulting in low baseline undecided rates. For example, for the 2021 cohort, only 3% of students were undecided at the end of the first semester.

[13] The top ten majors at the beginning of college and at graduation are somewhat similar for first-generation and continuing-generation students. See Appendix Table A7 for details.

university's minimum performance standards. They are also 4pp (17%) more likely to be designated *"Twice off Track"*, an official status signaling repeated underperformance relative to their major requirements. As shown in the last row of the panel, first-generation students earn first-semester GPAs that are, on average, 0.22 grade points lower.

We also see evidence that first-generation students are less adept at using withdrawal policies strategically to avoid low grades. They withdraw from 51% of the courses in which they eventually either fail or withdraw, compared to 58% for continuing-generation peers.[14] Moreover, when first-generation students do withdraw, they tend to do so later in the term (Appendix Figure A1). This suggests that first-generation students may be slower to recognize poor fit or performance problems, or slower to learn how and when to exercise the option to withdraw, both of which suggest a role for information to influence students' academic trajectories.

3.2 Academic Gaps Accounting for Observables: Coarsened Exact Matching

A natural question to ask is whether the academic achievement gaps by first-generation status documented above stem directly from parental education or are driven by observable differences in student background and preparation.

To disentangle forces specific to parental education from other factors, we employ the coarsened exact matching (CEM) algorithm, following the methodology of Iacus et al. (2012). This approach approximates a sample of first- and continuing-generation students that are identical in all observable characteristics except parental education. The matching process balances students along the following dimensions: race indicators (non-minority, Black, Hispanic, and other), gender, initial major category (based on 2-digit CIP family group codes), Arizona residency, transfer credit status, financial aid, household income, high school GPA, ACT/SAT scores, and first-semester GPA. Most matching variables are determined before university enrollment or within the first semester at university. Household income (and parental education)

[14]Students may be reluctant to withdraw if, for example, dropping the course leads them to miss the course credits threshold required for certain loans or for maintaining scholarship status. Notably, the gap in withdrawal rates is even more pronounced among students who have no academic loans, suggesting that financial constraints influence the withdrawal decision but cannot explain the first-generation withdrawal gap.

is derived from the earliest FAFSA record available but may be reported later than the first semester if initially missing. First-semester GPA, though determined post-enrollment, is included as a proxy for unobserved ability. Results are consistent when this variable is excluded from the matching process.

We first "coarsen" continuous variables into discrete categories.[15] Using these coarsened variables, we partition the sample into bins that represent the joint realization of matching variables. We then construct sample weights, ensuring that first-generation and continuing-generation students within each bin have equal weight while preserving the relative weight of bins with respect to the full ASU sample. These weights enable a non-parametric comparison of academic outcomes across parental education groups while accounting for observable characteristics.

Column (4) of Table 2 reports the differences between first- and continuing-generation students after applying CEM weights.[16] This method isolates the impact of parental education by constructing samples that are balanced on observable characteristics.[17] It is important to note that our purpose in this analysis is *not* to estimate the causal impact of parental education on students' academic outcomes. Rather, by jointly controlling for demographic, socioeconomic, and ability/preparation channels, we seek to demonstrate that the lack of a college-educated parent presents a unique set of barriers to a student's academic success. This analysis likely underestimates the "true" effect of parental education, as it controls for intermediate factors like household income and preparation that may mediate the overall impact of parental education.

Each of the variables in Panel A of Table 2 (along with first-semester GPA shown in the last row in Panel C) is used to create the CEM weights. Therefore, the last column in Table 2 validates that the CEM weights effectively balance the sample

[15]For continuous variables, we use five levels, except for household income, which is divided into eight levels to account for the thick tails of the household income distribution. Missing values are treated as a separate category. We exclude the bottom and top 1% of household income to mitigate outlier effects. Our findings are robust to alternative discretizations. See Appendix B for details on the CEM implementation.

[16]See appendix Table A8 for the difference between ASU population and the CEM weighted sample. The CEM sample differs from the full analysis sample because CEM weights can only be constructed for CEM strata, defined by the joint realizations of the matching variables, with at least one first-generation and at least one continuing-generation observation. To account for this sample selection, CEM weights rebalance the sample to reflect the overall ASU population.

[17]Regression analysis with extensive controls is an alternative approach but may fail to capture complex non-linear interactions between variables. Given our data's scope, CEM offers a more robust, non-parametric alternative for isolating the effects of parental education.

across these targeted matching variables. While continuous variables (e.g., income, standardized test scores, GPA) show small residual differences, these are not (statistically or economically) significant and, by construction, are not present once the variables are coarsened.

Turning to academic outcomes in Panel B, column (4) shows that many of the naive gaps in column (3) decline in magnitude (which is expected) but remain robust after applying CEM weights. For instance, first-generation students are still 2pp less likely to graduate and 3pp more likely to drop out. They also drop out nearly a fifth of a semester earlier. This pattern is also evident in Appendix Figure A2, which illustrates the timing of dropout among students who leave college. The largest difference between first-generation and continuing-generation students occurs early in their college careers, with a significant gap in the first semester.

Intermediate outcomes in Panel C reveal no significant differences in the likelihood of switching into an undecided major or being placed on academic probation. These findings suggest that differences in these measures are attributable to observable characteristics. Interestingly, first-generation students switch majors less often in the CEM-matched sample, primarily among those who drop out. This suggests that first-generation students are less likely to explore alternative majors (relative to their continuing-generation peers) before leaving college. While relatively small in overall magnitude, the gap in major-switching rates within the CEM sample appears economically meaningful, given that it could account for two-thirds of the 3pp (residual) dropout gap. While dropping out may certainly be optimal for some students, these patterns suggest that some first-generation students may be leaving college prematurely, when they could have successfully switched to a different major and graduated. We also see that residual differences in withdrawal rates persist. The residual first-generation differences in column 4 are not specific to capacity-constrained majors where a minimum GPA is required for continued enrollment - results are qualitatively identical when business and engineering students are dropped from the sample.

The achievement gaps in column 4 of Table 2 represent barriers specific to first-generation students. A similar CEM matching exercise by income rather than parental education generates a strikingly different pattern of outcomes. In particular, lower-income students with identical observables are *more* likely to graduate (1pp, $p < 0.05$) and less likely to transfer (0.03pp, $p < 0.01$) compared to their higher-income peers.[18]

[18]See appendix Table A5 for details.

They are also more likely to switch majors (3pp, $p < 0.01$), remain enrolled longer conditional on dropping out (0.38 semesters, $p < 0.01$), and are equally likely to use withdrawals when at risk of failing. The unique pattern of outcomes by parental education, conditional on preparation and demographic/socioeconomic factors, motivates our search for a specific mechanism capable of rationalizing first-generation achievement gaps.

Transcript data alone are unable to illuminate the mechanisms underlying these patterns. In the next section, we introduce new (survey) data to further explore differences between first- and continuing-generation students.

4 Evidence of Knowledge Frictions

In this section, we investigate mechanisms that may contribute to the unexplained achievement gap by parental education. In particular, we provide evidence that first-generation students face a distinct set of information constraints or "knowledge frictions" that influence their academic expectations and subsequent decision-making.

4.1 Differential Belief Updating by Parental Education

Informational gaps can indirectly affect academic outcomes by shaping students' beliefs and expectations regarding their future performance, such as their end-of-year GPA. To examine this channel, we analyze how students update their GPA expectations over time.

Figure 1 plots the distribution of students' "belief accuracy"—defined as the difference between their expected end-of-year GPA and the historical average GPA of a comparable reference group (same parental education and similar first-semester GPAs). A positive value indicates that the student's expectation exceeds the typical realized GPA of similar students. We present these distributions separately by parental education for both survey rounds.

Several patterns emerge. First, both first- and continuing-generation students exhibit optimism, with the mass of the distribution lying mostly above zero in both rounds. Second, the initial (Round 1) beliefs of first-generation students are more dispersed and positively biased compared to those of their peers. On average, first-generation students overestimate their GPA by 0.35 points relative to historic patterns, whereas continuing-generation students overestimate by 0.17 points. Third, over time (Round 2), expectations tighten around zero, indicating that students become more accurate as they gain experience. This narrowing of beliefs is much more pronounced for first-generation students, whose average overestimation declines from 0.35 to 0.23, compared to a more modest drop from 0.17 to 0.16 for continuing-generation students. These findings suggest that all students update their expectations, but first-generation students begin with more biased priors and make larger adjustments, consistent with more severe initial information constraints.

A natural explanation for why first-generation students adjust their beliefs more dramatically is that they receive systematically worse first-semester grades. Indeed, in the first semester, 49% of first-generation students earned grades below their initial expectations, compared to only 32% of continuing-generation students ($p = 0.001$). More pessimistic updates could simply reflect the fact that first-generation students start with overly optimistic priors and then confront more negative news. However, GPA expectation gaps alone cannot fully explain the pattern. Conditioning on the sign of the GPA expectation gap—measured as actual minus expected first-semester GPA—reveals that first- and continuing-generation students experience gaps of similar magnitude on average. Panel A of Figure 2 shows that, conditional on experiencing a positive GPA expectation gap, both groups exceed their expectations by about 0.3 grade points, on average; those with a negative gap underperform by roughly 0.5 grade points.

However, despite experiencing similar average GPA expectation gaps (conditional on the sign), first-generation students adjust their future GPA expectations more than their peers. This can be seen in Panels B and C of Figure 2. Even when conditioning on the sign of the GPA expectation gap, first-generation students revise their expectations nearly twice as much as continuing-generation students, a difference that is statistically significant for both positive and negative expectation gaps. A natural concern is that some omitted characteristic – beyond the GPA expectation shock itself – might be driving the observed gap in revisions between first-generation and continuing-generation students. However, any latent factor that explains why first-generation students revise nearly twice as much would have to satisfy two requirements. First, it would need to be entirely unique to first-generation students (i.e., absent among continuing-generation students). Second, it would have to affect belief updating in the same direction regardless of whether the shock is positive or negative—that is, it would have to amplify upward revisions when the shock is favorable and amplify downward revisions when the shock is unfavorable. These patterns make it difficult to attribute the differential updating solely to unobservables unrelated to grades. Supporting this intuition, we find that the patterns in GPA expectations updating are robust to demographic, socio-economic, preparation, and grade-signal controls. In particular, regressions of the GPA expectations gap (Panel A of Figure 2), and GPA expectation revisions (Panels B and C of Figure 2) onto a first generation dummy and controls find an even more robust response to grade signals among first-generation students (see Appendix Table A9 for details).[19]

These results suggest that while all students learn about their own performance upon entering college, first-generation students place more weight on observable signals such as grades when forming posterior expectations.

4.2 Differential Uncertainty by Parental Education

One possible reason why first-generation students respond more aggressively to similar average GPA expectation gaps is that they may have greater ex-ante uncertainty about their own academic abilities. To assess this, our survey elicited subjective first-year GPA distributions from each student in 0.5 grade-point intervals.[20] Using these reported probabilities, we calculate the variance of students' beliefs under two different distributional assumptions: uniform mass within a bin and a fitted beta distribution.

First-generation students report significantly higher variance in GPA expectations as compared to continuing-generation students. Table 3 reports differences in uncertainty by parental education. The average GPA variance was 0.118 (0.075) for the uniform mass (fitted beta) distributional assumption. On average, first-generation students had 0.033 (0.023) higher variance under the uniform mass (fitted beta) distributions. The pattern of higher uncertainty among first-generation students remains robust after including controls for household income, high school GPA, and other

[19]Controls include log household income, high school GPA, gender, a non-white indicator, a STEM major indicator; for future grade expectations (Panels B and C of Figure 2), we also control for the the GPA expectations gap (Panel A of Figure 2)

[20]Specifically, students were asked: *"What do you believe is the likelihood (chance out of 100) that your cumulative GPA will fall within the specified ranges below at the end of the Spring 2022 Semester (at the end of your freshman year)? Note: your responses should sum to 100."* Continuing in the survey required a response that summed up to 100.

background characteristics, as shown in Columns (2) and (4).

Interestingly, we do not observe a similar pattern by household income – once conditioning on observables, there is no difference in subjective grade uncertainty between above/below median household income students (see Appendix Table A10 for details). This provides further evidence that achievement gaps related to parental education represent a unique mechanism compared to barriers associated with socio-economic status. Further, these results indicate that subjective uncertainty may be a candidate for this mechanism. If students act as Bayesians when forming posterior beliefs about future academic performance, higher uncertainty among first-generation students could explain their stronger updating behavior in response to realized grade information. We build on this intuition through the model presented in Section 5.

4.3 Differential Perceived Major-Switching Costs by Parental Education

Our matched-sample comparisons in column 4 of Table 2 reveal that first-generation students are less likely to change their majors than continuing-generation peers (especially among the subset of students who drop out). The survey data suggests that this pattern can be observed in beliefs data elicited within the first few weeks of the fall semester, indicating that factors discouraging first-generation major switching are likely already active before students have stepped on campus. Specifically, the survey instrument asked students about their likelihood of switching majors given different realized first-year GPA outcomes. For both groups, intent to switch majors decreases as GPA increases, featuring a clear discontinuity at the 3.0 threshold.[21] However, starting from the 2.5–3.0 GPA range, first-generation students consistently report a 6–10 percentage points lower probability of switching majors than their peers. While differences in major preferences could potentially explain this gap, controlling for initial major subject (CIP code family) only strengthens the gaps. Further, conditional on graduating, we find no evidence that first-generation students who switch majors make more drastic moves across majors in terms of the majors' course content.[22] Both groups' initial major choices and the major switches made by successful

[21] Figure A3 plots expected conditional major switching probabilities by parental education.

[22] We measure a major's course content as the distribution of course load across academic departments. See Appendix Figure A4 for details.

graduates suggest that preference heterogeneity is unlikely to explain first-generation students' reluctance to switch majors, leaving systematic information frictions as a more plausible explanation.[23,24]

Additional survey results support this interpretation. Appendix Table A11 demonstrates that first-generation students perceive greater similarity in expected earnings across different fields of study.[25] This pattern remains robust even after residualizing expectations on observable characteristics such as household income, race, gender, high school GPA, and standardized test scores, suggesting that the stronger correlations within first-generation students are specific to their parental education status and not due to demographic factors (see the lower panel of Table A11). One way to rationalize this finding, in light of the evidence presented above, is to posit that first-generation students have less precise information about major-specific returns, leading them to anchor on a shared prior for post-graduation earnings across all majors. Under this assumption, poor grades in one major reduce expected returns not only in that field but also in alternative fields, thereby deterring major switching.

4.4 Additional Evidence on Frictions

Before concluding this section, we present additional findings that support the interpretation that first-generation students face more information frictions.

Panel A of Table 4 shows that first-generation students report fewer close friends at ASU (2.0 vs. 2.3, $p = 0.033$), know fewer students already enrolled at ASU (4.7 vs. 6.9, $p < 0.001$), and know fewer students in their incoming class (4.9 vs. 7.4,

[23]In addition, Appendix Figure A5 shows that *conditional on incoming math proficiency* (defined based on a normed math test ASU administers to all students), first-generation students sort into majors that are traditionally harder to graduate in. If students were fully informed about their subjective likelihood of success in their initial major, the ex-ante likelihood that first-generation students switch majors should be higher than that of their peers, not lower.

[24]There could also be some norm-based explanations for why first-generation students ex-ante have less flexible major-switching plans. For example, they may be more likely to view switching majors as a failure. However, it is important to highlight that, conditional on graduation, first-generation students are ex-post not less likely to change majors (when compared to continuing-generation students), suggesting that preferences or norm-based explanations cannot fully drive their initial reluctance to change major while in college.

[25]Students were asked about their expected earnings at age 35 if they were to graduate in each of the majors. The exact wording of the question was: *"Below is a list of major categories at ASU. You are asked to report how much you expect YOUR annual income at age 35 to be, if you graduate with a major in that category. When answering, assume a dollar when you are 35 is worth a dollar today, i.e., ignore the effects of inflation."*

$p < 0.001$) compared to continuing-generation peers. Since social connections may facilitate the exchange of academic and institutional knowledge, smaller networks may restrict the ability of first-generation students to acquire valuable information. Additionally, first-generation students are less likely to report that they will seek academic support or guidance from parents and family members, and are more likely to report that they will rely on friends outside the college environment. Both patterns are consistent with more limited access to reliable, experienced academic advisors and mentors.

Consistent with the idea that first-generation students are less well-informed upon arriving in college, column (2) of Table A4 shows that high school GPA is perceived to be similarly predictive of expected first-semester GPA by first- and continuing-generation students. That is, both groups place similar weight on high school GPA when forming expectations, even though, as shown in column (1), high school GPA is a less informative signal of academic preparedness for first-generation students (presumably because they tend to attend worse schools).

The lower panel of Table 4 shows that first-generation students' ex-ante beliefs about the economic returns to graduation are broadly similar to those of their continuing-generation counterparts. On average, first-generation students expect that earning a degree will more than double their age 35 earnings (2.17 times the earnings without a degree), which is statistically indistinguishable from the 2.22 of continuing-generation students. Thus, their higher dropout probability cannot be attributed to lower perceived returns to graduating or better outside options.

Overall, our findings reinforce the notion that first-generation students begin their college careers operating under greater information frictions. Compared to continuing-generation students, they: (1) arrive with more biased and overly optimistic expectations, (2) revise their beliefs downwards more over time, (3) adjust more aggressively to similar GPA expectation shocks, (4) hold greater uncertainty about their future grades, (5) are less likely to anticipate switching majors and perceive greater similarity in expected earnings across majors, and (6) possess smaller social networks with fewer sources of informed guidance. We use these patterns to motivate the conceptual framework that we outline next.

5 A Model of Reenrollment under Uncertainty

We propose a stylized model that combines Bayesian belief updating with a simple utility framework. The model demonstrates that ex-ante differences in uncertainty alone are enough to capture many key empirical differences in first- and continuing-generation students' subjective beliefs and real-world academic decision-making, and therefore provides a useful framework to rationalize persistent academic achievement gaps among first-generation students.

5.1 Model

Consider an agent deciding whether to continue their studies or drop out of college. We model this decision as a 2-period problem and assume risk neutrality. In period 1, the agent enters college enrolled in major m, with limited information about the costs and benefits of continued enrollment. In particular, the agent is uncertain about the subjective net value of staying in their current major. This net value may reflect the agent's beliefs about a range of factors such as the time and resources required, psychological costs, job prospects upon graduation, and time to completion. We summarize these factors as a single random variable representing the net benefit of continuing in major m, denoted as ν_m. To capture the ability of students to switch majors, the agent has similarly limited information about the returns of graduating in their next-most-preferred major, m', with $m' \neq m$, and denote this value as $\nu_{m'}$. We assume ν_m and $\nu_{m'}$ follow a bivariate normal distribution:

$$\begin{pmatrix} \nu_m \\ \nu_{m'} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_m \\ \mu_{m'} \end{pmatrix}, \begin{pmatrix} \sigma_m^2 & \rho\sigma_m\sigma_{m'} \\ \rho\sigma_m\sigma_{m'} & \sigma_{m'}^2 \end{pmatrix} \right),$$

where μ_m and $\mu_{m'}$ represent the mean of ν_m and $\nu_{m'}$, respectively; σ_m^2 and $\sigma_{m'}^2$ represent the variance of ν_m and $\nu_{m'}$; and ρ represents the correlation between ν_m and $\nu_{m'}$. We assume the true means of both random variables ν_m and $\nu_{m'}$ are unknown to the agent, but the variances and the correlation are known. We denote the agent's beliefs about μ_m and $\mu_{m'}$ at the start of period 1 as $\mu_{m,1}$ and $\mu_{m',1}$.

We assume the agent's prior beliefs about the returns to major m are unbiased, *i.e.* $\mu_{m,1} = \mu_m$. To rationalize the agent's initial enrollment in major m over m', we assume $\mu_{m',1} < \mu_{m,1}$. Additionally, we assume a positive correlation of strictly less than 1 between the agent's subjective returns to majors m and m' $(0 < \rho < 1)$, such that an agent who believes they have a high return to major m also expects high returns in major m'. Finally, for simplicity, we impose symmetry on the agents' uncertainty: $\sigma_m = \sigma_{m'}$.[26]

During period 1, the agent receives grades based on their academic performance in major m. We model this information as an imprecise observation of ν_m with a noise term $\epsilon \sim N(0, \sigma_\theta^2)$. In other words, the agent observes a signal $\theta = \nu_m + \epsilon$.

At the end of period 1, the agent decides whether to continue enrollment in their current major m into period 2, receiving a payoff of ν_m; switch to major m' and receive $\nu_{m'}$, or; drop out and receive some outside option value v_o. We assume v_o is constant, known to the agent, and lower than their prior about their next-most-preferred major. Given that $\mu_{m,1} > \mu_{m',1}$, it follows that enrolling in major m was optimal at the start of period 1, and that major m' was also preferred to the outside option, *i.e.* major m' represented a viable major at the start of period 1.[27] This setup leads to the following value function:

$$u_1(\theta) = \max\left(\mathbb{E}\left[\nu_m|\theta\right], \, \mathbb{E}\left[\nu_{m'}|\theta\right], \, v_o \right). \tag{1}$$

To concisely describe how agents update their beliefs, we further assume that the agent acts as a Bayesian when updating her priors about the value of ν_m and $\nu_{m'}$. Then, the agent's updated beliefs about ν_m follow a normal distribution, with the posterior expectation of ν_m given by:

$$\mu_{m,2} = \frac{\sigma_\theta^2}{\sigma_m^2 + \sigma_\theta^2}\mu_{m,1} + \frac{\sigma_m^2}{\sigma_m^2 + \sigma_\theta^2}\theta. \tag{2}$$

Note that the relative contribution of the signal (θ) to the posterior $(\mu_{m,2})$ depends

[26]The value of $\sigma_{m'}$ is not crucial for the key model results, and so we impose symmetric uncertainty only for algebraic and notional convenience. By assuming that the correlation between returns to majors is positive, we impose that a negative signal about performance in major m induces a downward revision in the agent's beliefs about the return to both major m and m'. If the correlation were negative, a negative signal about ν_m would induce an upward revision in the agents' expectation for their return to m'. It seems unlikely that performing poorly in one major would increase a student's expected return to their next-best major in an absolute sense. Of course, poor performance in m may make m' appear to have a higher *relative* return, and may even lead to major switching. A positive but less than one correlation coefficient would allow for such a case.

[27]The assumption that v_o is less than $\mu_{m',1}$ is not strictly necessary. However, if the agent enters university with their second-choice major offering less value than their outside option, major switching will be a degenerate choice.

on the precision of the signal (σ_θ^2) relative to the precision of the agent's prior (σ_m^2). The more accurate the signal relative to the agent's prior, the greater its influence on the agent's updated beliefs. Similarly, the posterior expectation of $\nu_{m'}$ is given by:[28]

$$\mu_{m',2} = \mu_{m',1} + \frac{\rho\sigma_m^2}{\sigma_m^2 + \sigma_\theta^2}(\theta - \mu_{m,1}). \tag{3}$$

An agent will decide to drop out rather than continue enrollment into period 2 if $\mu_{m,2} < v_o$ and $\mu_{m',2} < v_o$.[29] The decision to drop out can, therefore, be summarized as a threshold on the signal θ. To see why, note that the only unknown factor at the start of period 1 in equations (2) and (3) is θ. Further, $\mu_{m,2}$ and $\mu_{m',2}$ are monotone functions of θ. Because v_o is constant, $\mu_{m,2}$ and $\mu_{m',2}$ will cross v_o at most once. Thus, it is optimal for the agent to drop out if their observed θ falls below both thresholds. Similar to the dropout decision, the agent will optimally choose to switch majors if the expected value of $\mu_{m',2}$ is greater than both $\mu_{m,2}$ and v_o. We can then derive similar threshold conditions on θ under which major switching is optimal. Appendix C.1 provides a more detailed derivation of these thresholds.

In order for the model to capture the empirically documented pattern of major switching, we will assume there is a non-zero probability that the agent will ever switch majors. Under this assumption, two θ thresholds fully characterize the agent's policy function. A lower threshold between dropping out and switching majors that we will refer to as θ^1, and a higher threshold between switching majors and remaining in the initial major that we will refer to as θ^2.

5.2 Comparative Statics: High- and Low-Knowledge agents

Consider two types of agents: f for first-generation and c for continuing-generation. These types are identical in every way, except that first-generation agents begin with less prior knowledge about the university environment. Specifically, we make the

[28] See Bishop (2006), section 2.3.1, for a detailed derivation of these posterior means.

[29] Without loss, we assume that the agent breaks ties by favoring the status quo, *i.e.*, prefers to stay enrolled over dropping out and prefers staying in major m over switching to m'.

following assumptions about agents' ex-ante beliefs:

$$\sigma^2_{m,f} > \sigma^2_{m,c} \tag{A1}$$

$$\rho_f > \rho_c \tag{A2}$$

The first assumption (A1) states that first-generation agents are more uncertain about the returns to continuing in school than continuing-generation agents. Section 4.2 provided empirical evidence that respondents to our survey report subjective belief distributions that align with this assumption. This assumption directly implies the following corollary:

Corollary 1. *Upon receiving identical grade signals, a first-generation agent will update their beliefs weakly more than a continuing-generation agent,* i.e., $|\mu_{m,2,f} - \mu_{m,1}| \geq |\mu_{m,2,c} - \mu_{m,1}|$. *If the signal is informative ($\theta \neq \mu_{m,0}$), this will strongly hold.*

Assumption A2 states that first-generation agents perceive a stronger correlation between the subjective returns to their first and second-choice majors (section 4.3 provides a discussion of empirical evidence that supports this assumption).

The following comparative statics exercise examines how uncertainty affects agents' academic decision making. In mapping these predictions to the achievement gaps of first-generation students, we must assume that otherwise identical first-generation students differ from their continuing-generation peers as described in Assumptions A1 and A2. While we are unable to fully match first- and continuing-generation survey respondents on observable characteristics, Tables 3 and A11 demonstrate that first-generation students report higher uncertainty in both empirical analogs to assumptions A1 and A2, even after controlling for observable factors.

The main implications of our model are summarized in the following two propositions (see Appendix C.3 for details):

Proposition 1. *Continuing-generation agents have a lower drop out signal threshold compared to first-generation agents: $\theta^1_c < \theta^1_f$.*

Proposition 2. *Continuing-generation agents have a higher major switching signal threshold compared to first-generation agents: $\theta^2_c > \theta^2_f$.*

Figure 3 provides a visual representation of the relationships between signal thresholds implied by Propositions 1 and 2. Given their greater initial uncertainty, first-generation agents assign more weight to the signal θ when forming their posterior beliefs about the returns to major m. Further, because they perceive a stronger correlation between returns to major m and major m', they also assign more weight to the signal when updating their beliefs about m'. This additional weight on grade signals implies that if a first-generation and a continuing-generation agent receive the same negative grade signal ($\theta < \mu_{m,1}$), the first-generation agent will have larger downward revisions of both μ_m and $\mu_{m'}$.

5.3 Testable Implications of the Model

Several testable implications about the relative behavior and beliefs of first- and continuing-generation students can be derived from the model's comparative statics. In particular, the following five implications capture key aspects of the agents' decision-making and can be directly evaluated using our survey and administrative data:

1. Otherwise similar first-generation students have higher dropout rates and lower major-switching rates compared to their continuing-generation peers.

2. When otherwise similar first- and continuing-generation students receive the same grade signal, the first-generation student will more drastically update their expected subjective returns to college, both positively and negatively.

3. When forming posterior beliefs about subjective college dropout, otherwise similar first-generation students will put more weight on observed grade signals.

4. Conditional on receiving negative signals, first-generation students have higher dropout rates and lower major-switching rates compared to otherwise similar continuing-generation students; they have similar rates conditional on receiving positive signals.

5. There is a range of intermediate grade signals for which first-generation students will be more likely to drop out and less likely to switch majors as compared to their otherwise similar continuing-generation peers.

Implication 1 follows directly from Propositions 1 and 2. Implication 2 follows directly from Corollary 1.

Implication 3 directly links empirically observable beliefs (likelihood of dropout) and observed information (grades). This implication is the tightest connection between our model and empirically observable objects, as it provides a sharp prediction of how students' beliefs respond to informative signals. To bring the data to the model directly requires an extension of the model to produce non-deterministic posteriors, along with a few additional assumptions discussed in more detail in section 6.3. The details of this extension are discussed in Appendix Section C.2.[30]

Implication 4 follows from the observation that only agents with sufficiently negative signals will choose to drop out or switch majors. Then, the model predicts that the difference in unconditional dropout rates is completely driven by agents with negative grade signals. In contrast, agents who receive positive signals are expected to have similar outcomes, as a positive signal reinforces agents' initial belief that enrollment in major m is optimal.

Implication 5 follows directly from Proposition 1. As illustrated in Figure 3, for grade signals between the first-generation and continuing-generation dropout thresholds ($\theta \in \left(\theta_c^1, \theta_f^1\right)$) first-generation agents will dropout and continuing-generation agents will switch majors. For signals above the first-generation threshold, both types switch major; for signals below the continuing-generation threshold, both types drop out. Therefore, the model predicts that a dropout/major-switching gap should exist for an interior (intermediate) range of grade signals away from the extremely negative or merely weakly negative signals.

5.4 Model Discussion

The model's comparative statics demonstrate that when students make reenrollment decisions under uncertainty, differences in ex-ante uncertainty alone will result in achievement gaps. The model's key testable implications described above hold in our survey and administrative data, evidence of which is discussed in detail in the following section.

The purpose of the model is to articulate a mechanistic connection between un-

[30]More specifically, adding an unobserved utility shock after receiving the grade signal leads to a model (sub)period in which agents have probabilistic posterior beliefs; after observing their grade signal, they await for additional uncertainty to be resolved before making their reenrollment decision. This extension generates a sharp prediction about the weight first-generation and continuing-generation students place on grade information when forming their posterior dropout beliefs.

certainty and achievement in the context of higher education. To this end, many aspects of students' real-world experiences are abstracted away in the model. Several of these abstractions merit further discussion. First, the model does not attempt to explain major switching for reasons other than negative information shocks. Many students who do well in their initial courses nevertheless choose to switch to a major that better suits their interests or goals. Our framework has nothing to say about the motivations/mechanisms behind these major switches.[31]

Second, the model does not indicate whether first- or continuing-generation agents are more likely to switch majors conditional on continued enrollment. The difference in conditional switching rates depends on the probability that θ falls between θ_c^1 and θ_f^1 relative to the probability that θ falls between θ_c^2 and θ_f^2. The relationship between these probabilities is not determined by the model's assumptions.

Third, the model summarizes all information a student receives, relevant to their reenrollment decision, into a one-dimensional signal. In reality, students rely on many information sources/signals (e.g., family, friends, media). Empirically, first-generation students have smaller social networks and discuss academic issues less frequently, thus relying more heavily on grades to update their expectations. Therefore, incorporating multiple signals would likely reinforce the existing mechanism.

Finally, the model assumes students have unbiased expectations. Empirically, all students overestimate their chances of graduating. This overestimation is even larger among first-generation students. If we incorporated overconfident priors into the model, first-generation students would show even higher dropout rates, due both to greater uncertainty and more pronounced optimism bias.[32]

[31]It is worth noting that differential major switching costs — specifically, higher switching costs for first-generation students (for example, due to higher opportunity costs of being enrolled in college) — could also help explain lower major switching rates for first-generation students. However, a model with only differential major switching costs would not be able to generate the second or third model predictions (i.e., larger updating and more weight on grade signals for first-generation students), or the fourth prediction (first-generation gaps in major-switching patterns in response to negative signals but not positive signals).

[32]To understand the role of ex-ante bias, consider two students with similar ex-ante expected returns to college enrollment. If one student is more optimistic than the other, then her actual returns to reenrollment (ν_m) must be lower. This means that her grade signal will be downward-biased relative to her peers' distribution, and thus, she will be more likely to observe a grade below her dropout threshold.

6 Validating Model Predictions

This section presents empirical evidence that students' beliefs and real-world academic decision-making follow the comparative statics generated by the simple model outlined in the previous section.

6.1 Model Implication 1: Achievement Gaps

Model implication 1 predicts that, among otherwise comparable students, first-generation students should exhibit higher dropout and lower major-switching rates. Evidence from the Coarsened Exact Matching (CEM) analysis in Section 3 aligns with this prediction. Specifically, column 4 of Table 2 (based on a CEM sample balanced on a wide range of observables) shows that first-generation students are less likely to switch majors and more likely to drop out.

6.2 Model Implication 2: Differential Belief-updating

Model implication 2 states that, controlling for initial qualifications, first-generation students should adjust their expected returns to continued enrollment more strongly in response to both positive and negative signals than their continuing-generation peers. Although we cannot directly observe changes in continuation value, we proxy these returns using students' expected future GPA.

As discussed in Section 4.1, we find evidence that first-generation students more drastically update their GPA expectations after both positive and negative grade signals (see Figure 2 and the associated discussion in Section 4.1). The model predicts this type of expectation updating for *otherwise identical* first- and continuing-generation students. While we lack the sample size to fully match first- and continuing-generation respondents in our survey, we find these patterns are robust to including controls for race, gender, household income, high school GPA, and average GPA within declared ASU major. Further, in three of four conditional outcomes evaluated, including controls widens the first-generation revision gap (Appendix Table A9).

Finally, Appendix Table A12 shows that first-generation students with smaller networks (i.e, those with network sizes in the bottom tercile) exhibit weakly larger updating in response to both positive and negative GPA signals. This is consistent with the discussion in section 5.4, where we hypothesized that, under an information friction mechanism with multiple signals, responsiveness to GPA shocks should be greater for groups that are less likely to receive signals from other sources.

6.3 Model Implication 3: Posterior Signal Weights

Model implication 3 predicts that, among otherwise comparable students, those without a college-educated parent should put more weight on observed grade signals when forming posterior beliefs about subjective college dropout. Both prior and posterior beliefs about subjective dropout and first-semester grades are directly observable in our survey data. Therefore, the empirical analog to the agents' Bayesian updating equation can be estimated. Marginal effects of students' grade signals on their posterior reenrollment beliefs are presented in Figure 4 separately by first-generation status. Appendix Table A13 presents the full estimation results. These marginal effects are derived from fractional response probit models, controlling for initial beliefs, outside option values (log expected income if not graduating), baseline GPA expectations, and the interaction of these terms with first-generation status.[33,34]

The model extension discussed in Appendix Section C.2 predicts first-generation students will be more sensitive to grades when updating their dropout expectations under the assumptions that (i) first- and continuing-generation students have identi-

[33]The full empirical specification is:

$$\begin{aligned}\mathbb{P}\left[dropout_i|\phi_i\right]_{i,2} =&\alpha + \beta_1\phi_i + \beta_2 FG_i + \beta_3\phi_i FG_i + \beta_4\mathbb{P}\left[dropout_i\right]_{i,1} + \beta_5\mathbb{P}\left[dropout_i\right]_{i,1} FG_i \\ &+ \beta_6 ln\left(\mathbb{E}\left[Earn|no\,grad\right]_{i,1}\right) + \beta_7 ln\left(\mathbb{E}\left[Earn|no\,grad\right]_{i,1}\right) FG_i + \beta_8\mathbb{E}\left[\phi_i\right]_{i,1} + \varepsilon_i,\end{aligned}$$

where ϕ_i is student i's semester 1 GPA, FG_i is an indicator equal to 1 if student i is first-generation, $\mathbb{P}\left[dropout_i|\phi_i\right]_{i,2}$ is student i's posterior subjective dropout likelihood elicited in survey wave 2, $\mathbb{P}\left[dropout_i\right]_{i,1}$ is student i's prior subjective dropout likelihood elicited in survey wave 1, $\mathbb{E}\left[Earn|no\,grad\right]_{i,1}$ is student i's expected earnings at age 35 conditional on not graduating from college elicited in survey wave 1, and $\mathbb{E}\left[\phi_i\right]_{i,1}$ is student i's prior expectation about their 1st semester GPA elicited in survey wave 1.

[34]Fractional response models, introduced by Papke and Wooldridge (1996), are designed for dependent variables that represent proportions or probabilities and thus lie within the [0,1] interval. Unlike linear models, which can yield predictions outside this feasible range, fractional response models ensure that fitted values remain between 0 and 1. In our setting, a nontrivial proportion of students report corner probabilities (exactly 0 or 1) for outcomes such as returning for the second semester or graduating within four years. We, therefore, employ a fractional response probit specification, which uses a probit link function to model the conditional mean of the probability within [0,1], providing a natural fit for self-reported probabilities while properly handling corner responses. For further details, see Papke and Wooldridge (1996).

cal grade distributions, (ii) first- and continuing-generation students relate observed grades to expected subjective outcomes in the same way, and (iii) first- and continuing-generation students have the same marginal effects of the net utility of continued enrollment on posteriors. We ensure assumption (i) is satisfied by estimating the marginal effects using sample weights to balance the first-generation and continuing-generation samples across 10 bins of first-semester GPA.[35] Assumption (ii) is a technical assumption required to directly relate observed GPA to the theoretical "grade signals" in the model. One way to evaluate how reasonable this assumption may be is to test the relationship between expected earnings conditional on graduating and expected GPA in our first-round survey. In the CEM-weighted sample used for the fractional response estimation, we cannot reject the null hypothesis that first- and continuing-generation students have the same relationship between expected earnings conditional on graduating and expected first-semester GPA.[36] Finally, assumption (iii) can be directly evaluated by testing for differential marginal effects of students' outside option value; see appendix Section C.2 for more details about why this marginal effect is relevant. Appendix Table A13 shows precisely estimated zero effects of expected log earnings conditional on not graduating on students' posterior reenrollment expectations. Given this, we can rule out the possibility that the differences in marginal grade effects can be explained by differences in net-utility.

Under assumptions (i), (ii), and (iii), the model provides a sharp prediction that first-generation students put more weight on grades when forming posterior dropout expectations. The estimates in Figure 4 show that the marginal effect of first-semester grades on posterior dropout expectations is indeed larger for first-generation students. Specifically, first-generation students are roughly 3-4 times more sensitive to grade information as compared to their continuing-generation peers. The model can only rationalize these heterogeneous marginal effects as differences in Bayesian weights due

[35]These weights are constructed using the CEM procedure as described in Appendix B with only one matching variable (GPA) coarsened into 10 bins.

[36]Specifically, in a regression of log expected earnings conditional on graduating from ASU on expected first semester GPA using the same sample weights as in Figure 4, we cannot reject the null hypothesis that first and continuing-generation students have identical coefficients on GPA expectations: the coefficient (standard error) on the interaction of GPA expectations and the first-generation dummy is -0.01 (0.22) (p=0.961). Further, to the extent that assumption (ii) is violated and first-generation students do not relate grades to expected outcomes in the same way as continuing-generation students, it seems likely that information gaps would be a strong candidate for the underlying mechanisms generating differential mappings. For this reason, we view this assumption as primarily technical, given the structure of the model.

to higher ex-ante subjective uncertainty among first-generation students.[37]

6.4 Model Implication 4: Outcomes Conditional on Grade Signals

Model implication 4 considers how key academic outcomes for otherwise comparable first- and continuing-generation students vary depending on their early grade signals. To empirically assess these predictions, we adopt a CEM strategy similar to the one outlined in Section 3.2. However, here we first stratify students into two groups based on their first-semester grades — those who performed well ("good grades") and those who did not ("poor grades"). Within each group, we use CEM to balance first- and continuing-generation students on a rich set of pre-determined characteristics (the same set used in column 4 of Table 2).[38] These matched samples reflect two comparable samples for first- and continuing-generation students conditional on students' early grades. To ensure our results are not sensitive to the exact definition of poor performance, we construct several sets of weights using different discretizations of the grade distribution.[39]

A key identifying assumption is that, conditional on the matched observables, first- and continuing-generation students who earned poor grades are comparable in ways relevant to their academic decisions. In particular, even if the reasons for earning poor grades are endogenous, the combination of controls for major, household income, pre-college academic credentials (such as high school GPA and standardized test scores), and first semester GPA should mitigate concerns that first-generation

[37]Note, that the model does not provide any prediction about the relative weights first- and continuing-generation agents put on their prior dropout expectations when forming posterior beliefs about subjective dropout (*i.e.*, the coefficients on prior expectations in Table A13). This is because the agent's subjective expectation about dropout does not directly obey Bayes' Theorem but is derived from Bayesian updating over subjective returns to reenrollment (ν_m, $\nu_{m'}$). Because of this, the prior beliefs in our estimates of the marginal grade effects on dropout expectations should be viewed as controls to ensure marginal grade effects are estimated for two students with identical prior beliefs, rather than as an (indirect) measure of a Bayesian weight.

[38]This approach is robust to performing the CEM weighting on the four subgroups (first-generation/poor grades, first-generation/good grades, continuing-generation/poor grades, continuing-generation/good grades) simultaneously.

[39]Specifically, we define poor grades using thresholds such as receiving a C or lower in one major-specific course, receiving a C or lower in two major-specific courses, or holding a cumulative GPA below 2.33 (which corresponds to C+). Major-specific courses are defined as the eight most common first-year courses in a given major.

and continuing-generation students fail courses for systematically different reasons.[40] Although we cannot rule out all forms of selection, we will provide further evidence supporting our identification strategy by applying the method developed in Oster (2019) to test for selection on unobservables.

Figure 5 displays estimates for various academic outcomes in both the good-grade and poor-grade sub-samples, with first-generation averages represented by blue circles and continuing-generation averages by red triangles. In this figure, we define "poor" grades as earning at least one C or lower in a major-specific first-year course. Appendix Figures A6 and A7 present analogous results for alternative poor-grade definitions, yielding qualitatively similar findings. Results are also robust to using a GPA cutoff and excluding business and engineering students, who often face minimum GPA requirements for continued enrollment due to major-specific capacity constraints.

The first part of model implication 4 predicts that relative dropout rates between otherwise similar first- and continuing-generation students should be larger among students who received negative early signals. Panel A of Figure 5 shows that, among students receiving poor grades, the dropout rate gap is approximately 6 percentage points (with first-generation students being more likely to drop out). A similar pattern emerges for four-year graduation rates, as shown in Panel B.

Panels A and B also speak to the second part of implication 4, which predicts that students who receive positive early signals should have similar outcomes regardless of parental education background. Consistent with this prediction, we find no statistically meaningful differences among students who performed well. For example, for students who received good grades, dropout rates are nearly identical for matched first- and continuing-generation students, differing by only 0.004pp ($p = 0.597$). This similarity in outcomes across all five panels of Figure 5 for the good-grade group strengthens the interpretation that the main unobserved difference driving first-generation outcomes under poor performance scenarios is indeed related to the knowledge frictions our model posits.

Finally, we consider major-switching rates in Panel C of Figure 5. Among students who receive poor grades, first-generation students are 2 percentage points less likely to switch majors ($p = 0.088$). This difference is larger (4pp $p = 0.090$, and 3pp

[40]Many studies rely on endogenous signals when studying the role of information frictions on academic outcomes (e.g., Arcidiacono (2004); Thomas (2019); Arcidiacono et al. (2020); Larroucau and Rios (2022), among others). Moreover, we do not require any assumption over why students earn poor grades, which is endogenous.

$p = 0.030$) under alternative definitions of poor performance shown in Figures A6 and A7. Again, consistent with our model, there is no difference in major switching rates among students who receive positive early signals. These findings suggest that first-generation students who do poorly early on are less likely than their continuing-generation counterparts to adjust their academic path by switching majors.

Why do poorly performing first-generation students switch majors at lower rates? The model predicts that some marginal first-generation students would have switched majors rather than dropped out had they possessed better initial information. Panel D provides additional support for this interpretation. We find no difference in major-switching conditional on reenrollment into the second year, implying that first-generation students who remain enrolled long enough to acquire more information behave like their ex-ante better-informed peers. Taken together, the panels of Figure 5 indicate that the gap in major-switching rates after poor performance arises because marginal first-generation students who might otherwise have switched majors instead drop out prematurely.

The analysis in Figure 5 can be interpreted as causal if our observables (on which we match) fully account for sorting into good- and bad-grade events by first-generation status. However, there could be unobservable (to the econometrician) differences between these groups, which may explain the differential reactions to poor grades. In Appendix Table A14, we show that to overturn our main results regarding differences in dropout, graduation, and major switching within the poor-grade group, the degree of selection on unobserved factors would have to exceed that on the observed factors by a factor of 5 for dropout and graduation, and a factor of 3 in the opposite direction for switching majors, suggesting our conclusions are robust to plausible forms of omitted variable bias.[41]

As an additional robustness check, because high school GPA may not be fully comparable across different secondary schools, we restrict the sample to in-state students (for whom we observe school names which we then use to geolocate to a census tract) and include neighborhood income as an additional matching variable to account for potential differences in school quality between first- and continuing-generation students. Appendix Figure A8 shows that our qualitative conclusions remain unchanged,

[41]The observable factors include major fixed effects, household income, highest standardized test score (ACT or SAT equivalent score), high school GPA, high school Math GPA, high school English GPA, number of transfer credits, cohort fixed effects, athlete dummy, high school state fixed effects, age, military veteran dummy, and a US citizen dummy.

further supporting the robustness of our findings.

Finally, a poor-grade event could be correlated with other factors that differ by first-generation status. For example, first-generation students are more likely to work in college and face financial distress. Thus, it is conceivable that a poor grade event for first-generation students is accompanied by financial shocks. To consider this situation, we restrict the analysis to low-income students (defined as students with below median household income ($74,280 in 2019 dollars) and show that our conclusions are largely unchanged (Appendix Figure A9).

6.5 Model Implication 5: Academic Decisions Conditional on Grade Signals

Model implication 4 considers students' outcomes conditional on receiving a positive or negative signal. Implication 5 expands this analysis to consider the grade signals for which first- and continuing-generation students' academic decisions most diverge. Specifically, it posits that there should be an interior range of grades for which first-generation students are both less likely to switch majors and more likely to drop out. Figure 6 plots CEM weighted dropout rates (Panel A) and major switching rates (Panel B) for student-semester observations within the first academic year, separately for 15 semester-GPA bins. We find that, for GPAs between 1.8 and 2.0, first-generation students are 3.4pp more likely to drop out and 8.4pp less likely to switch majors ($p = 0.016$ and $p < 0.000$, respectively). These GPAs correspond to the 14th-18th percentiles of the (CEM-weighted) semester GPA distribution and thus represent poor, though non-extreme, ranges of the grade distribution, as predicted by the model. A similar, though less statistically significant, pattern emerges for the preceding GPA bin (1.6-1.8).[42]

The results in Figure 6 suggest that the dropout and major-switching gaps among poor-performing students documented in Figure 5 are driven by the same range of intermediate semester GPAs signals, precisely as the model would predict.

[42] For grades significantly above the 1.6-2.0 GPA range, first- and continuing-generation students have economically/statistically similar responses in terms of dropout and major-switching decisions. Turning to major switching, though statistically insignificant ($p = 0.205$), the 2.4-2.6 GPA bin has an economically relevant point estimate of -1.8pp. This is an order of magnitude larger than the estimates in adjacent bins. Though not precisely estimated, this jump is consistent with model predictions for agents' behavior within the set of signals above the major-switching threshold for first-generation students and below this threshold for continuing-generation students ($\theta \in \left(\theta_c^2, \theta_f^2\right)$).

In conclusion, our analysis shows that the key testable implications of our model of reenrollment under uncertainty are supported by both our survey and administrative data. While knowledge frictions are inherently difficult to measure, the various pieces of evidence converge to present a consistent picture: parental education emerges as a critical determinant of student success, beyond the factors typically emphasized in the literature. We now turn to evidence on a policy intervention with potential to alleviate these frictions.

7 Policy Scope to Address Barriers - LEAD Program

This section analyzes the impact of the "Learn Explore Advance Design" (LEAD) program at Arizona State University. LEAD is a holistic program designed to support academically at-risk students through structured coursework and mentorship. Although the program was not explicitly created to address information frictions, several of its key components—such as structured guidance, personalized academic support, and peer mentoring—are well-positioned to reduce the uncertainty and knowledge gaps that first-generation (and other at-risk) students often encounter when navigating college.

It is important to acknowledge that the program encompasses a range of support mechanisms beyond addressing information frictions. In this context, we consider the LEAD program a valuable case study for understanding how comprehensive interventions—including those targeting information barriers—can influence student success. Finally, while our analysis is not solely focused on first-generation students due to limitations in statistical power, it takes a holistic approach by demonstrating how programs like LEAD may serve as effective models for addressing the unique challenges faced by first-generation students within broader student support initiatives.[43] Supporting this approach, we find suggestive evidence that LEAD improves

[43] It should be noted that, to date, very few college-level interventions have successfully narrowed the SES achievement gaps within higher education (Oreopoulos, 2021). The most successful program in this regard has been ASAP (Accelerated Study in Associate Program), which is quite expensive (cost of $4,676 per student per year) and likely quite infeasible for scalability. The program provided incoming first-year students tutoring, counseling, career advising, free public transportation passes, and textbook funding. In addition, students were required to set regular meetings with their tutors and attend a student success seminar (Scrivener et al., 2015).

outcomes by addressing the same knowledge channels identified as contributing to the first-generation achievement gap.

7.1 The LEAD Program

Since 2016, ASU has offered the LEAD program to incoming students, with a special emphasis on those who enter with weaker academic credentials—an at-risk population that disproportionately includes first-generation students.[44] Although any first-year student may join, those whose composite index (CI) score (a weighted sum of high school GPA components and SAT/ACT scores) falls below a specific threshold receive a strong nudge to enroll. This cutoff creates quasi-random variation to estimate the program's causal impact on academic performance and dropout risk.

The LEAD curriculum replaces standard first-year courses, allowing participants to earn required credits through smaller, more interactive, and homogeneous classes that emphasize soft skills (e.g., time management) and consistent engagement. More importantly, students benefit from peer and faculty/staff mentorship, creating closer interactions with professors and fostering a supportive learning community. These elements align with prior research (Carrell and Kurlaender, 2023; Bettinger and Long, 2018; Carrell and Sacerdote, 2017; Lu and Anderson, 2015) highlighting the importance of social integration and academic support for student success (see Appendix D for more details of the program). Because LEAD relies on existing faculty/infrastructure and peer mentors (and largely replaces courses students would have to take otherwise), it is also relatively low-cost—expanding the program by 350 students would cost under $650 per student per year.

7.2 Characterization of LEAD Students

Table 5 provides a comparison of LEAD and non-LEAD students in terms of the transcript data variables. We focus on students who entered college between Fall 2016 and Fall 2018 as these are the cohorts for which LEAD was active and sufficient time has passed for outcomes to be evaluated. LEAD students are more likely to come from more disadvantaged backgrounds: they are more likely to be first-generation (12pp), racial minorities (16pp), and from below median-income families (11pp) when

[44]Our main findings from previous sections using administrative data are qualitatively robust to excluding the 2016 and 2017 cohorts, who entered when LEAD was already in place.

compared to their non-LEAD counterparts. Moreover, LEAD students enter college less prepared than their peers, with lower ACT scores (5.7 points on average) and a lower high school GPA (0.5 grade points on average). This is to be expected as LEAD targets students based on their incoming academic preparation. While the LEAD program was not explicitly designed to reduce SES gaps in educational attainment, the fact that LEAD students are disproportionately from low-SES backgrounds suggests that LEAD may provide a reasonable test of the scope of interventions to overcome barriers faced by first-generation and low-SES students.

7.3 LEAD Program Evaluation Strategy

We exploit the CI cutoff score, which determines what students are nudged to enroll in LEAD, to identify the program's causal impact via a regression discontinuity (RD) design. Appendix Figure A10 confirms that students scoring below the threshold are substantially more likely to join LEAD.[45] We implement the robust bias-corrected RD estimator proposed by Calonico et al. (2014), using the coverage-error-optimal bandwidth from Calonico et al. (2020).[46] Our baseline specifications include controls for cohort, high school GPA, and transfer credits, with standard errors clustered at the CI score level to account for the coarseness of the running variable (see Appendix section D.2 for details concerning the identification strategy).

7.4 Impact of LEAD

Table 6 shows the estimated effect of LEAD on participants based on the reduced-form sharp RD (ITT) estimates on students' academic outcomes. Column 1 presents estimated impacts on 1st-semester GPA. Perhaps unsurprisingly, given the structure of the LEAD program, we find a positive effect of 0.10 grade points on participants' first semester GPA. We estimate a similar impact on the final cumulative GPA (see column 2). However, column 3 shows we lack statistical precision to establish whether LEAD participation impacts grades after leaving the program (*i.e.*, excluding the first

[45] In all analyses of the LEAD program, we exclude observations directly on the threshold score. There was some uncertainty about whether the threshold was inclusive or exclusive, and so observations at the threshold CI score appear partially treated relative to those below the cutoff.

[46] Because the running variable (CI scores) is relatively discrete, we increase the bias (pilot) bandwidth by 18% beyond the initial estimate (i.e., we set $\rho = 0.85$), the ratio that minimizes asymptotic variance with a triangular kernel (Calonico et al., 2020).

year), where the point estimate is slightly lower than in columns (1) and (2), but is too noisy to rule out a null effect. Nevertheless, columns 4, 5, and 6 indicate that LEAD substantially reduces early dropout: it raises two-year retention and overall graduation rates by 4 percentage points, though it has no impact on 4-year graduation rates. Taken together, these findings suggest that LEAD meaningfully boosts students' eventual graduation largely by improving their early academic performance and persistence.

We also conducted a placebo test using the 2003–2015 pre-LEAD cohorts, when the program did not exist (Appendix Table A15). Not a single outcome exhibits a statistically significant discontinuity at the cutoff in the same direction as in the LEAD sample. If anything, students below the cutoff in the pre-LEAD period tended to fare worse (which is what one would have expected). These placebo findings reinforce our confidence that the main RD estimates reflect LEAD's true causal impact rather than being driven by unobserved confounders.

7.5 Why does LEAD Work?

Although the LEAD program encompasses multiple components, several pieces of evidence suggest it operates—at least in part—through improving students' access to information and guidance, as highlighted in our model.

Evidence from the Student Survey. Our survey of first-year students in the incoming 2021 cohort sheds light on how LEAD may have enhanced their college experience. Compared with non-LEAD respondents with below-median high school GPAs, Table 7 shows that LEAD participants were more likely to agree that ASU staff helped them overcome initial struggles (75% vs. 53%), and that ASU's efforts to promote student-professor interaction helped them develop a strong support network (54% vs. 33%). They were also more likely to report that the small class sizes, a key feature of LEAD, helped them form friendships (43% vs. 24%), and that ASU's focus on student interaction was beneficial (57% vs. 34%). Finally, LEAD students were more than twice as likely to say their experience at ASU improved their mental health (46% vs. 17%). Although these differences are non-causal, they suggest that LEAD fosters stronger peer and mentor networks, facilitating the flow of information and improving students' overall well-being.

Evidence from Major Switching Rates. Another indicator that knowledge frictions play a central role is LEAD's effect on students' likelihood of switching majors. Our model predicts that students with higher uncertainty—often first-generation students—are more likely to drop out rather than switch majors when they encounter negative academic signals. Consistent with this, columns 7 and 8 of Table 6 show that LEAD participants were 5 percentage points more likely to switch majors and 5 percentage points less likely to switch from a declared major to "undecided".[47]

One concern might be that LEAD pushes students into majors that perhaps have low labor market returns. Such a switch would not be socially efficient. However, column 9 in Table 6 shows that it is not the case. It indicates that LEAD nudged students toward majors where the median wage for past ASU graduates is not lower than those of majors chosen by comparable non-LEAD students.

Our model of reenrollment under uncertainty would suggest that LEAD may have improved outcomes by improving grades early in students' academic career, when grade signals have the most impact on expectations, and may have directly reduced uncertainty through mentorship/network formation. These reinforcing channels may explain the relative effectiveness of the program. Taken together, our findings are consistent with LEAD reducing uncertainty and improving students' willingness to persist.

8 Conclusion

This study provides new insights into the role of knowledge frictions in shaping the academic trajectories of first-generation students at Arizona State University (ASU), one of the largest public universities in the U.S. Our findings suggest that first-generation students face significant informational barriers that hinder their ability to navigate college effectively. These frictions lead to larger revisions in their academic expectations following setbacks, lower rates of major switching, and higher dropout rates compared to their continuing-generation peers.

Through a model of student reenrollment under uncertainty, we demonstrate that first-generation students' limited initial knowledge makes them more sensitive to

[47]Moving to "undecided" typically reflects significant uncertainty or second-guessing. Hence, a reduced rate of declared-to-undecided transitions suggests students are making more informed major decisions.

early academic setbacks. Our empirical analysis confirms the framework's predictions, showing that first-generation students are more likely to drop out rather than switch majors when confronted with poor academic performance. These patterns persist even after accounting for a broad set of observable characteristics, underscoring the critical role of informational barriers in perpetuating academic disparities.

While some dropout is optimal, our findings – especially regarding the effectiveness of the LEAD program – suggest that *some* first-generation students are prematurely dropping out relative to the benchmark defined by their (otherwise similar) continuing-generation peers. Students who arrive in college uninformed and uncertain are "rationally" going to respond to any early signals of academic performance. Thus, interventions that lower barriers to information acquisition and/or provide additional informative signals early in a student's college career could be particularly effective at improving retention, narrowing socioeconomic achievement gaps, and enhancing intergenerational educational mobility.

Finally, although this study leverages both administrative and survey data to document the broad patterns of belief updating and dropout behavior, future research would benefit from a more micro-based approach – for example, by collecting detailed information on the channels through which students from different backgrounds receive academic feedback (e.g., advisor emails, peer networks, automated portal notifications, etc) – to understand how students acquire, process, and act upon information in real time.

References

Alamuddin, R., D. Rossman, and M. Kurzweil (2018). Monitoring advising analytics to promote success (maaps): Evaluation findings from the first year of implementation.

Altonji, J. G. (1993). The demand for and return to education when education outcomes are uncertain. *Journal of Labor Economics 11*(1), 48–83.

Angrist, J., D. Lang, and P. Oreopoulos (2009). Incentives and services for college achievement: Evidence from a randomized trial. *American Economic Journal: Applied Economics 1*(1), 136–163.

Angrist, J., P. Oreopoulos, and T. Williams (2014). When opportunity knocks, who answers?: new evidence on college achievement awards. *Journal of Human Resources 49*(3), 572–610.

Arcidiacono, P. (2004). Ability sorting and the returns to college major. *Journal of Econometrics 121*(1-2), 343–375.

Arcidiacono, P., V. J. Hotz, and S. Kang (2012). Modeling college major choices using elicited measures of expectations and counterfactuals. *Journal of Econometrics 166*(1), 3–16. Annals Issue on "Identification and Decisions", in Honor of Chuck Manski's 60th Birthday.

Arcidiacono, P., V. J. Hotz, A. Maurel, and T. Romano (2020). Ex Ante Returns and Occupational Choice. *Journal of Political Economy 128*(12), 4475–4522.

Aucejo, E. M. and K. Wong (2025). The effect of feedback on student performance. *Journal of Public Economics 241*, 105274.

Autor, D., C. Goldin, and L. F. Katz (2020, May). Extending the race between education and technology. *AEA Papers and Proceedings 110*, 347–51.

Baird, M., M. Buchinsky, and V. Sovero (2016). Decomposing the racial gap in stem major attrition: A course-level investigation. *RAND Working Paper Series WR-1171*.

Bettinger, E. P. and B. T. Long (2018, 01). Mass Instruction or Higher Learning? The Impact of College Class Size on Student Retention and Graduation. *Education Finance and Policy 13*(1), 97–118.

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Information Science and Statistics. Springer.

Bleemer, Z. and B. Zafar (2018). Intended college attendance: Evidence from an experiment on college returns and costs. *Journal of Public Economics 157*, 184–211.

Boneva, T. and C. Rauh (2020). Socio-economic gaps in university enrollment: The role of perceived pecuniary and non-pecuniary returns. *Working Paper*.

Bound, J., M. F. Lovenheim, and S. Turner (2010, July). Why have college completion rates declined? an analysis of changing student preparation and collegiate resources. *American Economic Journal: Applied Economics 2*(3), 129–57.

Bui, K. and R. A. Rush (2016). Parental Involvement in Middle School Predicting College Attendance for First-Generation Students. *Education 136*(4), 473–489.

Calonico, S., M. D. Cattaneo, and M. H. Farrell (2020). Optimal bandwidth choice for robust bias-corrected inference in regression discontinuity designs. *23*(2), 192–210.

Calonico, S., M. D. Cattaneo, and R. Titiunik (2014). Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs: Robust Nonparametric Confidence Intervals. *82*(6), 2295–2326.

Carrell, S. and B. Sacerdote (2017, July). Why do college-going interventions work? *American Economic Journal: Applied Economics 9*(3), 124–51.

Carrell, S. E. and M. Kurlaender (2023). My professor cares: Experimental evidence on the role of faculty engagement. *American Economic Journal: Economic Policy 15*(4), 113–141.

Davidson, R. and J.-Y. Duclos (2013). Testing for restricted stochastic dominance. *Econometric Reviews 32*(1), 84–125.

Dynarski, S., A. Nurshatayeva, L. C. Page, and J. Scott-Clayton (2023). Addressing nonfinancial barriers to college access and success: Evidence and policy implications. In *Handbook of the Economics of Education*, Volume 6, pp. 319–403. Elsevier.

Heckman, J. J. and T. Kautz (2012). Hard evidence on soft skills. *Labour Economics 19*(4), 451–464. European Association of Labour Economists 23rd annual conference, Paphos, Cyprus, 22-24th September 2011.

Iacus, S., G. King, and G. Porro (2012). Causal inference without balance checking: Coarsened exact matching. *Political Analysis 20*(1), 1–24.

Ivy Coach (2024). Who qualifies as a first-generation college student? https://www.ivycoach.com/the-ivy-coach-blog/college-admissions/who-qualifies-first-generation-college-student/. Accessed: 2025-07-29.

Larroucau, T. and I. Rios (2022). Dynamic college admissions. *Working Paper*.

Leighton, M. and J. Speer (2024). Family background and college major choice: Evidence on major earnings growth. *Education Finance and Policy*.

Lochner, L. J. and A. Monge-Naranjo (2011, October). The nature of credit constraints and human capital. *American Economic Review 101*(6), 2487–2529.

Lu, F. and M. Anderson (2015). Peer effects in microenvironments: The benefits of homogeneous classroom groups. *Journal of Labor Economics 33*(1), 91 – 122.

McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *142*(2), 698–714.

National Center for Education Statistics (2019). Beginning Postsecondary Students Longitudinal Study (BPS:12/17). https://nces.ed.gov/datalab. U.S. Department of Education, Institute of Education Sciences. Accessed via NCES DataLab.

National Center for Education Statistics (2022). National Postsecondary Student Aid Study (NPSAS:20). https://nces.ed.gov/surveys/npsas/. U.S. Department of Education, Institute of Education Sciences.

National Center for Education Statistics (2025). Ipeds institution profile. IPEDS Data Center. Accessed July 28, 2025.

Oreopoulos, P. (2021, June). What limits college success? a review and further analysis of holzer and baum's making college work. *Journal of Economic Literature 59*(2), 546–73.

Oster, E. (2019). Unobservable Selection and Coefficient Stability: Theory and Evidence. *Journal of Business & Economic Statistics 37*(2), 187–204.

Papay, J. P., R. J. Murnane, and J. B. Willett (2015, May). Income-Based Inequality in Educational Outcomes: Learning From State Longitudinal Data Systems. *Educational Evaluation and Policy Analysis 37*(1S), 29S–52S.

Papke, L. E. and J. M. Wooldridge (1996). Econometric methods for fractional response variables with an application to 401 (k) plan participation rates. *Journal of applied econometrics 11*(6), 619–632.

Qiu, X. (2025). Information inequality in college major choices. Working Paper.

Scrivener, S., M. J. Weiss, A. Ratledge, T. Rudd, C. Sommo, and H. Fresques (2015). Doubling graduation rates: Three-year effects of cuny's accelerated study in associate programs (asap) for developmental education students. *MDRC*.

Stange, K. M. (2012, January). An empirical investigation of the option value of college enrollment. *American Economic Journal: Applied Economics 4*(1), 49–84.

Stansbury, A. and K. Rodriguez (2024). The class gap in career progression: Evidence from us academia. *Working Paper*.

Stinebrickner, T. and R. Stinebrickner (2012). Learning about academic ability and the college dropout decision. *Journal of Labor Economics 30*(4), 707–748.

Stinebrickner, T. and R. Stinebrickner (2014). A major in science? initial beliefs and final outcomes for college major and dropout. *The Review of Economic Studies 81*(1 (286)), 426–472.

Thomas, J. (2019). The signal quality of grades across academic fields. *Journal of Applied Econometrics 34*(4), 566–587.

Zafar, B. (2011). How do college students form expectations? *Journal of Labor Economics 29*(2), 301–348.

Zafar, B. (2013). College major choice and the gender gap. *The Journal of Human Resources 48*(3), 545–595.

Table 1: Survey Sample Balance

	2021 Cohort	R1	Diff. (2)-(1)	R2	Diff. (4)-(1)	Diff. (4)-(2)	R1 + R2	Diff. (7)-(1)	Diff. (7)-(2)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Observations	13,189	1,316		722			423		
White	0.48	0.65	0.17***	0.63	0.15***	-0.01	0.64	0.16***	-0.01
Hispanic	0.28	0.27	-0.01	0.26	-0.02	-0.01	0.23	-0.04**	-0.03
Black	0.05	0.05	0.00	0.06	0.01	0.01	0.04	-0.01	-0.01
Female	0.52	0.62	0.10***	0.61	0.09***	-0.01	0.63	0.10***	0.01
First-Generation	0.26	0.34	0.08***	0.26	0.00	-0.08***	0.31	0.06**	-0.03
Med. HH Income ($1ks)	98.1	82.6	-15.5**	82.6	-15.5**	0.0	117.9	19.9	35.4
Honors	0.15	0.23	0.08***	0.21	0.06***	-0.02	0.29	0.15***	0.06**
ACT or SAT eq.	24.58	27.40	2.82***	26.42	1.84***	-0.98***	28.42	3.85***	1.03***
Missing ACT/SAT	0.57	0.38	-0.19***	0.53	-0.04**	0.15***	0.35	-0.21***	-0.02
HS GPA	3.54	3.70	0.16***	3.59	0.05***	-0.11***	3.77	0.22***	0.07***
R1 Belief: Return for Y2		92.38					92.97		0.59
R1 Belief: GPA after Y1		3.61					3.63		0.02
R1 Belief: Prob. Grad in 4		88.46					89.07		0.62

Notes: Table displays average student characteristics for several samples. Column 1 corresponds to the 2021 cohort in the administrative data. Column 2 includes all respondents to the first survey round, and Column 4 corresponds to all respondents to the second round. Column 7 includes those who responded to both the first and second rounds.. Columns 3, 5, 6, 8, and 9 report mean differences between the indicated samples. Stars denote statistical significance. ***: p<0.01, **: p<0.05, *: p<0.10

Table 2: Achievement Gaps by Parental Education

	First-Gen	Continue-Gen	Basic Difference	CEM Weighted Difference
	(1)	(2)	(3)	(4)
Observations	31,769	58,632	90,401	19,411
Panel A) Baseline Characteristics				
Household Income (1,000s $2019)† m	63.10	118.28	-55.18***	-0.96
In State (0/1)m	0.78	0.62	0.16***	-0.00
Race: Black (0/1)m	0.06	0.06	0.00	-0.00
Race: Latino/Hispanic (0/1)m	0.38	0.16	0.23***	0.00
Race: White/Asian (0/1)m	0.49	0.71	-0.22***	-0.00
Female (0/1)m	0.58	0.52	0.05***	0.00
ACT (or SAT eq.)m	23.07	25.28	-2.22***	-0.10
HS GPAm	3.38	3.42	-0.03***	-0.00
Financial Aid (0/1)m	0.95	0.90	0.04***	-0.00
Transfer Credits (0/1)m	0.28	0.28	-0.00	0.00
Panel B) Educational Outcomes				
Graduated (0/1)	0.61	0.67	-0.06***	-0.02***
Graduated in 4 (0/1)	0.44	0.51	-0.07***	-0.00
Transferred (0/1)	0.13	0.15	-0.02***	-0.01
Dropped-out (0/1)	0.25	0.17	0.08***	0.03***
Years before dropout | dropout	1.70	1.81	-0.12***	-0.17***
Graduated in STEM major (0/1)	0.12	0.15	-0.04***	-0.01
Panel C) Intermediate Outcomes				
Ever Switch Major	0.46	0.45	0.01***	-0.02**
Ever Switch Major | dropout	0.34	0.35	-0.01*	-0.03*
Ever Switch Major | grad	0.54	0.52	0.03***	-0.01
ASU semesters before in grad major | grad	2.05	1.87	0.18***	-0.03
Move into Undec.	0.05	0.04	0.01***	-0.00
Started in STEM major (0/1)	0.27	0.28	-0.01***	-0.00
Twice Off Track (0/1)	0.27	0.23	0.04***	0.01
Ever Academic Probation (0/1)	0.29	0.21	0.08***	0.01
Num W per W/F††	0.51	0.58	-0.06***	-0.01**
First semester GPAm	2.79	3.00	-0.22***	-0.02

Notes: Sample includes all in-person first-time freshmen from the 2000–2017 entering cohorts. Columns 1 and 2 report the average of each row variable for first- and continuing-generation students, respectively. Column 3 shows the unweighted mean difference between the two groups, and Column 4 presents the mean difference using the CEM-weighted sample. Stars denote statistical significance. ***: p<0.01, **: p<0.05, *: p<0.10

† Income winsorized below 1st percentile ($0) and above 99th percentile (∼ $680,000)

†† Number of withdrawn courses divided by number of courses withdrawn or failed.

m Matching variables.

Table 3: First Generation Students and Subjective GPA Uncertainty

	Dep Variable: Variance in Subj. GPA Expectations			
	Distriubtional Assumptions			
	Uniform Mass Within Bin		Fitted Beta Distribution	
	(1)	(2)	(3)	(4)
First gen. (0/1)	0.033***	0.024**	0.023***	0.017*
	(0.009)	(0.010)	(0.008)	(0.009)
Controls		✓		✓
Dep. mean	0.118	0.118	0.075	0.075
N	1,307	1,307	1,307	1,307
R^2	0.013	0.049	0.008	0.021
Adjusted R^2	0.013	0.042	0.007	0.013

Notes: Table reports estimates from separate regressions of variance in Round 1 GPA expectations onto a first-generation indicator. Variance calculated assuming a uniform mass within a bin (Columns 1 and 2) and using a fitted beta distribution (Columns 3 and 4). Columns 2 and 4 include controls for a non-white indicator, gender, the logarithm of household income, high school GPA, and major fixed effects. The sample includes Round 1 survey respondents with known parental education. The variance is winsorized above the 99th percentile. Robust standard errors are reported in parentheses. ***: p<0.01, **: p<0.05, *: p<0.10

Table 4: Survey: College Networks

	First-gen.	Cont.-gen.	P-value (1)-(2)
	(1)	(2)	(3)
Panel A) Network			
Numb. close friends at ASU	1.96	2.28	0.033
Numb. students already at ASU known	4.70	6.89	0.000
Numb. incoming class known	4.91	7.37	0.000
Likelihood (1-5) of seeking academic help from:			
Parents	3.30	4.23	0.000
Sibling/Other family	3.34	3.60	0.015
Friends not attending college	2.50	2.22	0.002
Panel B) Expected Returns			
E[earn at 35 | grad] / E[earn at 35 | no grad]	2.17	2.22	0.540
Observations	445	862	

Notes: Columns 1 and 2 report the average of each row variable for first- and continuing-generation students, respectively. Column 3 reports the p-value from a test of the difference in means between the two groups. The sample includes Round 1 survey respondents with known parental education.

Table 5: LEAD vs Non-LEAD at Entry

	LEAD Gap (1)	LEAD (2)	Non-LEAD (3)
Observations		2,595	30,133
Demographics			
First-Generation (0/1)	0.12***	0.44	0.32
Minority (0/1)	0.16***	0.54	0.38
Lower-Income (0/1)	0.11***	0.60	0.49
Preparation			
ACT (or SAT eq.)	-5.66***	19.71	25.37
HS GPA	-0.49***	3.02	3.50

Notes: Columns 2 and 3 report the average of each row variable for LEAD and non-LEAD students, respectively. Column 1 reports the mean difference between LEAD and non-LEAD students. Stars denote statistical significance. The sample includes all students in the 2016–2018 cohorts. ***: $p<0.01$, **: $p<0.05$, *: $p<0.10$

Table 6: LEAD Regression Discontinuity Estimates

	1st sem GPA	Cum GPA in Last Sem.	Cum GPA in Last Sem. (excl. 1st yr)	Finish Year 2 (0/1)	Grad. (0/1)	Grad. in 4 Years (0/1)	Ever Switch Major (0/1)	Ever Switch into Undec. (0/1)	Median Grad. Maj. Salary (1,000s)†
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Nudged (ITT)	0.10***	0.08*	0.06	0.04**	0.04***	0.00	0.05*	-0.05***	-0.10
	(0.03)	(0.05)	(0.05)	(0.02)	(0.01)	(0.02)	(0.03)	(0.02)	(0.82)
Observations	6,712	6,747	6,816	6,165	6,816	5,020	6,816	4,390	2,307
Within BW Mean	2.88	2.75	2.31	0.76	0.53	0.40	0.46	0.06	48.39
Optimal BW	8	8	8	7	8	6	8	5	5

† Missing major salaries estimated by LASSO on the course content of the major.

Notes: Table reports robust bias-corrected sharp RD (ITT) estimates of the effect of LEAD participation on key academic outcomes. Outcomes are indicated at the top of each column. Estimates use coverage-error optimal bandwidths following Calonico et al. (2020). All specifications control for cohort fixed effects, high school GPA, and transfer credits. The sample includes students in the 2016–2018 cohorts with reported high school GPA and standardized test scores in the administrative data, and with 12 or fewer transfer credits. Standard errors, clustered at the CI score (running variable) level, are reported in parentheses. ***: p<0.01, **: p<0.05, *: p<0.10

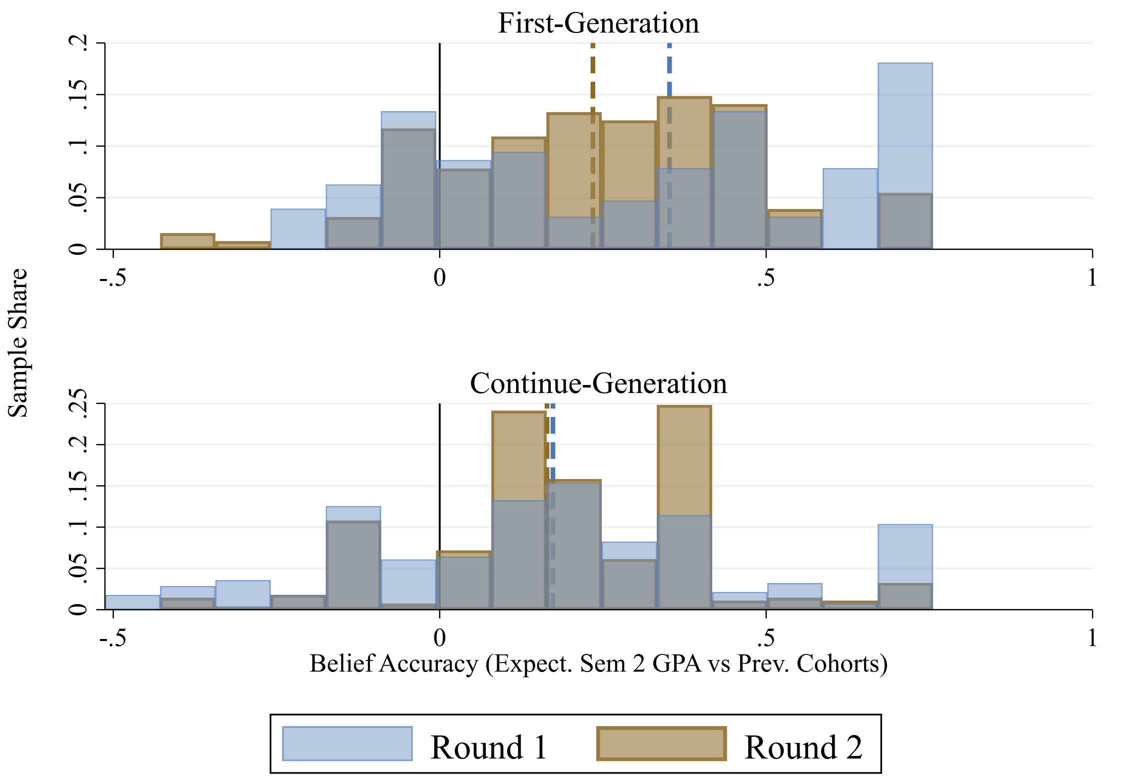
Table 7: First Year Experiences: LEAD vs Non-LEAD in Survey

	LEAD (1)	Non-LEAD Restricted† (2)	Diff. (1)-(2) (3)	Non-LEAD All (4)	Diff. (1)-(4) (5)
I [Somewhat/Strongly] Agree...					
ASU [staff] understood my needs so they could help me to overcome initial struggles	0.75	0.53	0.22***	0.56	0.19***
ASU's attempts to promote student-professor interaction helped me to create a strong support network at ASU	0.54	0.33	0.21***	0.33	0.21***
My typical class size helped me to create a strong group of friends at ASU	0.43	0.24	0.19***	0.25	0.18***
ASU's attempts to promote student interaction helped me to create a strong group of friends at ASU	0.57	0.34	0.24***	0.35	0.22***
My ASU experience improved my mental health	0.46	0.17	0.28***	0.21	0.25***
Observations	103	271		614	

† Non-LEAD restricted to participants with a high school GPA below the median (3.65).
Notes: Table shows differences in reported first-year experiences between LEAD participants and non-LEAD participants among respondents to the second-round survey. The sample is restricted to students who answered all questions corresponding to the row variables. Column 1 reports average responses for self-reported LEAD participants. Columns 2 and 4 report averages for two different samples of non-LEAD. Columns 3 and 5 present the corresponding mean differences. Stars denote statistical significance. ***: $p<0.01$, **: $p<0.05$, *: $p<0.10$

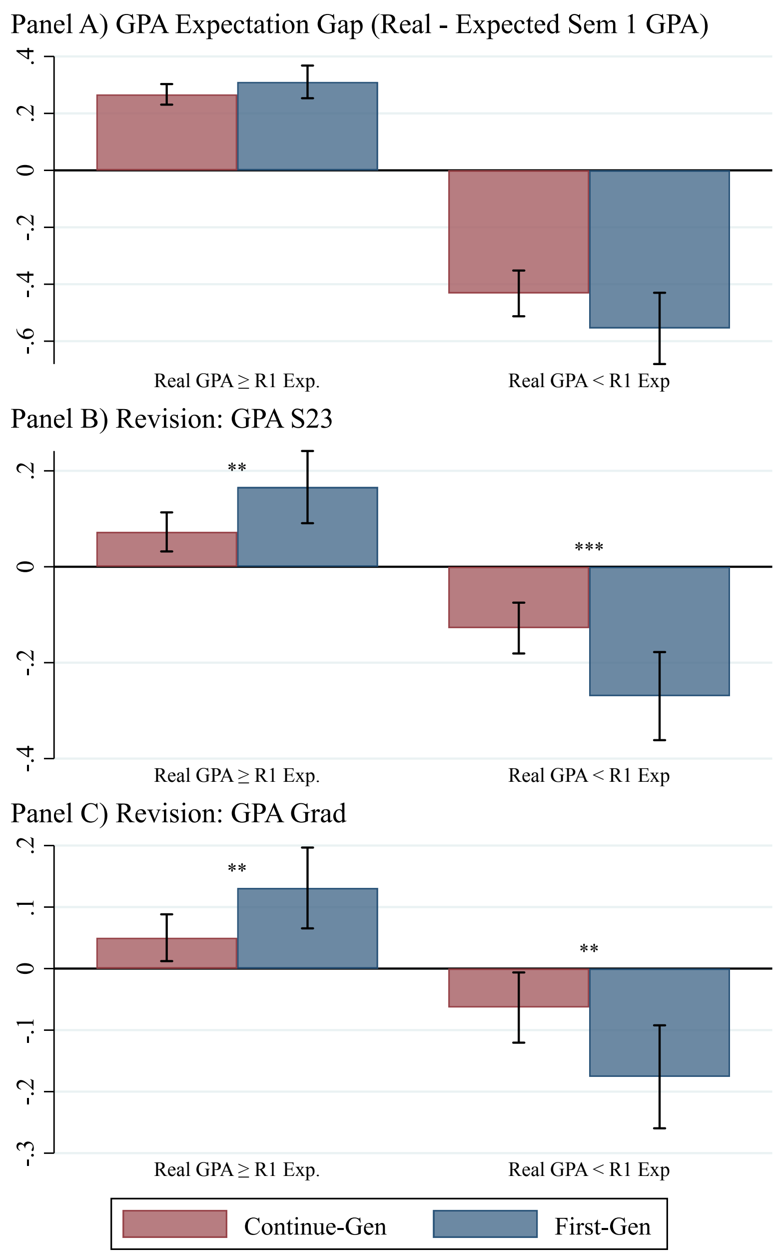
Figures

Figure 1: Revision in GPA Beliefs Between Survey Rounds



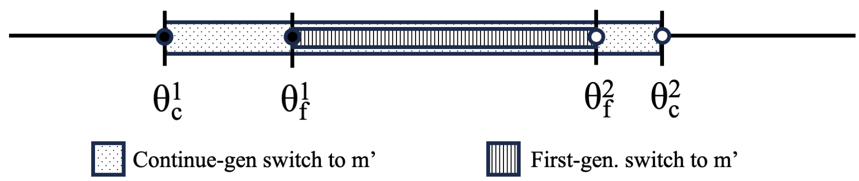
Notes: Figure plots the distribution of students' expected GPA at the end of their first year relative to the average 1st year GPA of students with the same parental education (first- or continuing-generation) and similar first semester GPA (15 equally sized bins) in the 2007-2018 cohorts. Beliefs reported in units of grade points, with a 1 corresponding to a student who believes she will earn a first-year GPA 1 grade point (i.e., an A average rather than a B average) higher than the average student with her same level of parental education and first semester GPA would have earned. Distributions are displayed separately by parental education and survey round. Dashed lines correspond to mean values for each group/period.

Figure 2: Grades and Subjective Expectation Updating



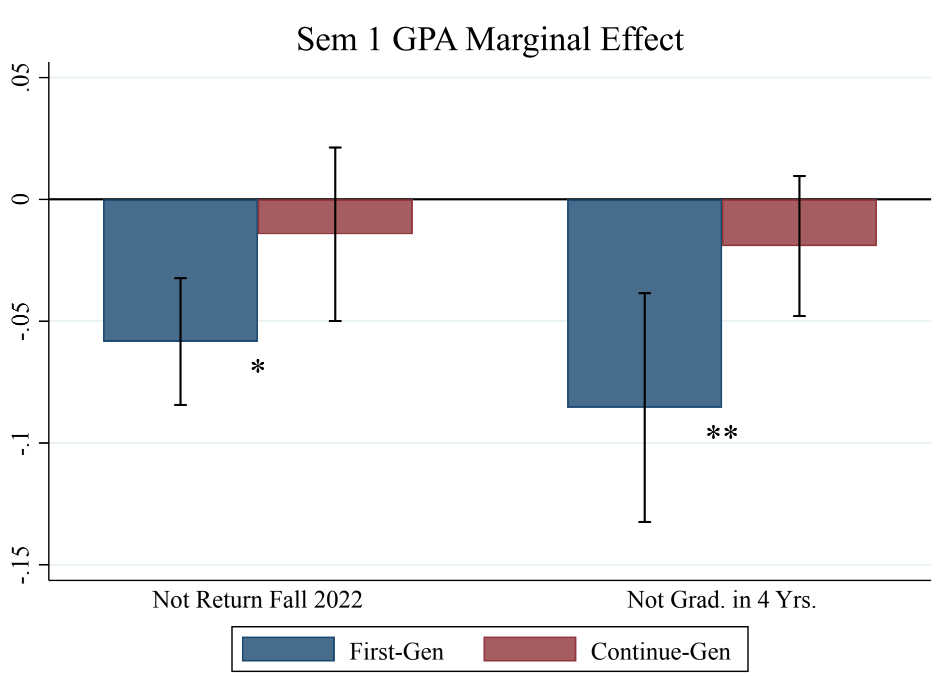
Notes: Figures show average GPA expectation gap (Panel A) and revisions in subjective expectations (Panels B and C) separately by parental education (indicated by color) and sign of grade signal (positive grade signals on the left, negative grade signals on the right). GPA expectation gap measured as the difference between a respondent's reported 1st semester GPA in survey round 2 (R2) and their expected first-semester GPA in survey round 1 (R1). Revision in expectations measured between R1 (start of fall semester) and R2 (mid spring semester) survey waves. I-bars correspond to 95% confidence intervals. Stars correspond to statistical difference in means between first- and continuing-generation samples conditional on grade signal. ***: $p<0.01$, **: $p<0.05$, *: $p<0.10$

Figure 3: Signal Thresholds by Type



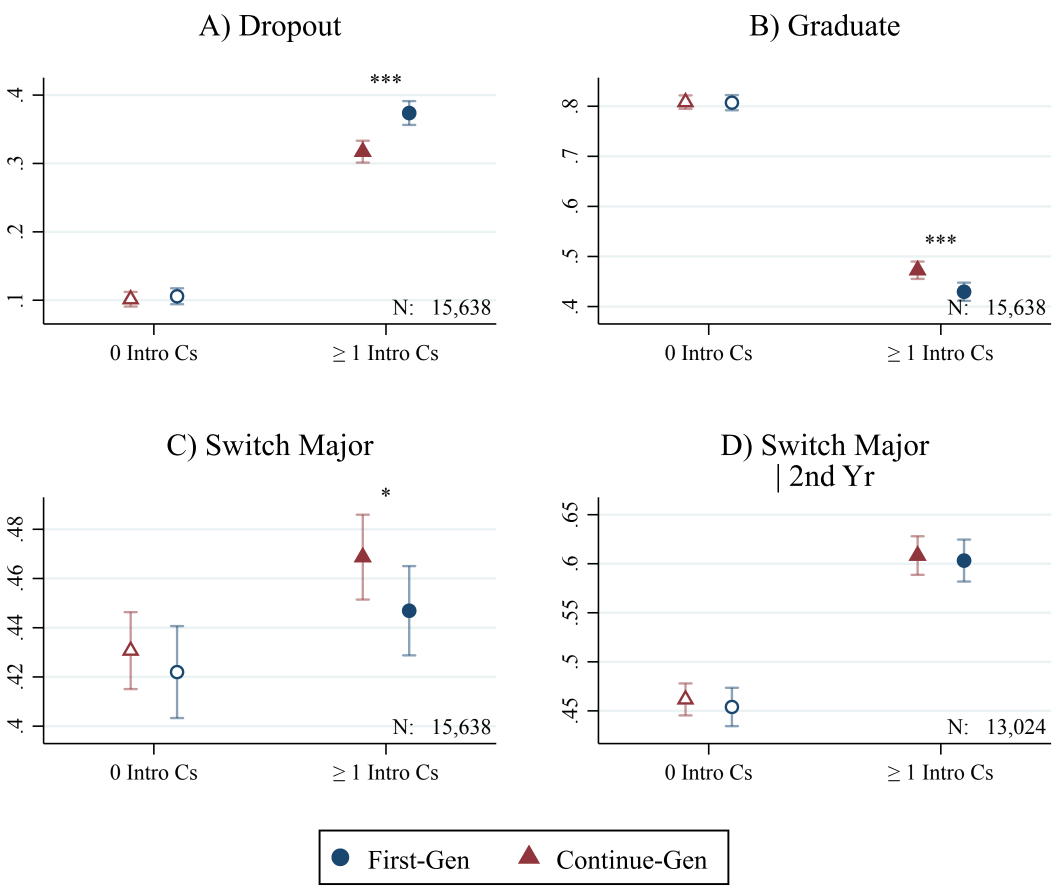
Notes: Figure displays the ranges of the grade signal θ over which first-gen agents (vertically shaded region) and continuing-gen agents (dotted region) optimally choose to switch majors to m'. At or above θ_f^2 (θ_c^2) first-gen. (continuing-gen.) agents remain in major m, and below θ_f^1 (θ_c^1) first-gen. (continuing-gen.) agents drop out of university.

Figure 4: Marginal Effects of Grade Signals on Belief Updating



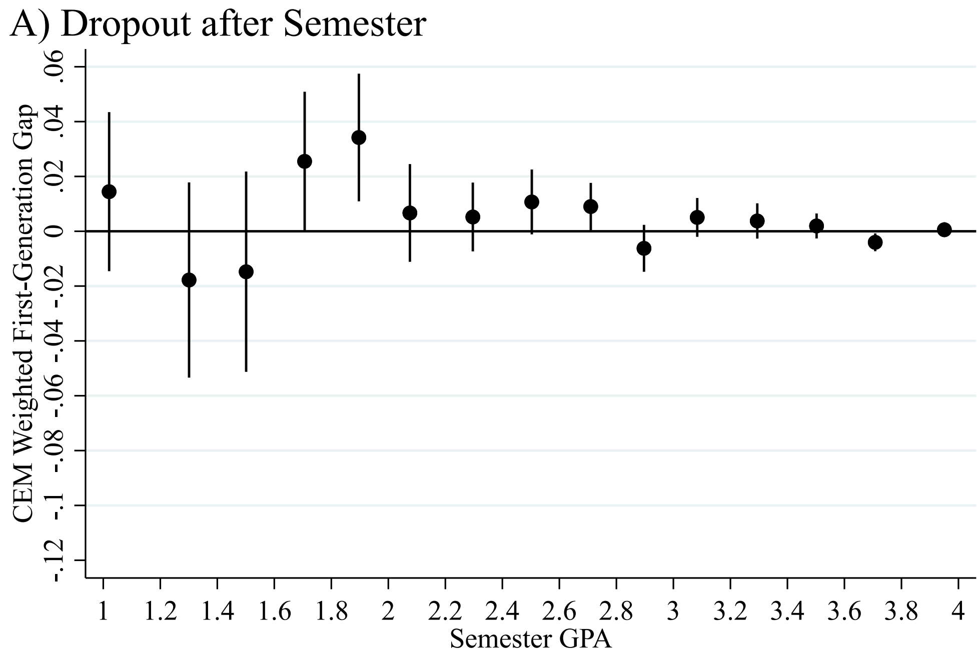
Notes: Figure plots the marginal effects of first-semester GPA on posterior (survey round 2) beliefs. Sample weighted to balance first-gen and continuing-generation samples across 10 GPA bins. First- and continuing-generation effects were jointly estimated in a fractional probit regression for each outcome. Independent variables included first-semester GPA, prior belief, log expected income at age 30 conditional on not graduating, a first-generation dummy, first-generation interaction terms with each of the previous variables, and survey round 1 GPA expectations. Table A13 displays the full set of marginal effects.

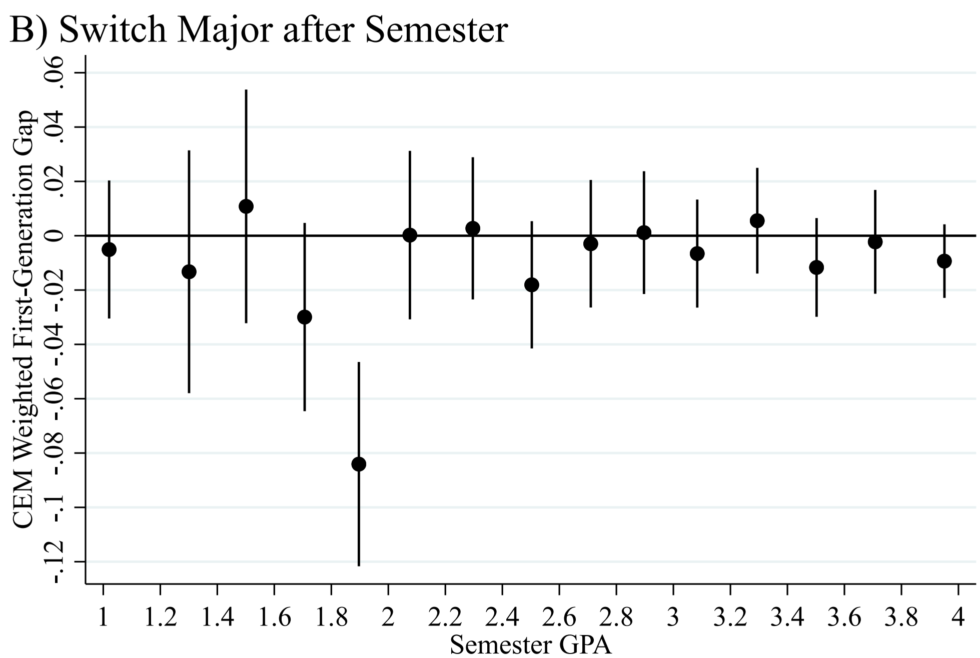
Figure 5: Reaction to Receiving at Least One C or Lower in Intro Courses



Notes: Figure plots the (weighted) average outcomes within the 2000-2017 cohorts. CEM weights were constructed to balance matching variables within first-year performance group. Poor (good) First-year performance is measured as (not) receiving at least one grade of C or worse in a student's major-specific introductory courses. Matching variables include: race indicators (non-minority, Black, Hispanic, and other), gender, initial major category (based on 2-digit CIP family group codes), Arizona residency, transfer credit status, financial aid, household income, high school GPA, ACT/SAT scores, and first-semester GPA. 95% confidence intervals indicated using I-bars. Panel A displays the share of students who drop out. Panel B displays the share of students who graduate. Panel C displays the share of students who ever switch from a declared terminal major to a different declared terminal major. Panel D displays the share of students who switch majors conditional on enrolling for strictly more than 2 semesters. Stars represent statistical differences between first-generation and continuing-generation weighted averages. ***: $p<0.01$, **: $p<0.05$, *: $p<0.10$

Figure 6: CEM Weighted First-Generation Conditional Dropout Gaps within First Year.





Notes: Panel A plots the difference in semester-specific dropout rates between first- and continuing-generation students conditional on semester GPA using the Coarsened Exact Matching (CEM) sample weighting scheme used to construct the estimates in Table 2 column 4. Drop out is measured as a one if the student drops out immediately after receiving the indicated GPA and zero otherwise. Panel B plots the difference in semester-specific major switching rates between first- and continuing-generation students conditional on semester GPA using the same CEM weights. The sample for both panels includes student-semester observations within the first year for students in the 2000-2018 cohorts. Switching is measured as a one if the student switches majors immediately after receiving the indicated GPA and zero otherwise. In both panels, observations with zero semester GPA excluded from sample (2.1% of observations). Semester GPA winsorized below 1 (3.2% of observations). Gaps were estimated separately with 0.2 GPA bins. Standard errors clustered at the student level. Dropout measured as a binary indicator taking a value of one if the observation is a student's last semester at ASU and the student dropped out and zero otherwise. Major switching measured as a binary indicator taking a value of one if a student is in a terminal major (not undecided) and their next-semester major is different from their current major and zero otherwise. Spikes represent 90% confidence intervals.

Appendix A - Tables

Table A1: Sample Selection: Known Parental Education and HH Income

	Known Parental Education (1)	Unknown Parental Education (2)	Difference (3)	Known HH Income (4)	Unknown HH Income (4)	Difference (5)
Observations	90,401	55,718		93,564	52,555	
Know HH Income (0/1)	1.00	0.06	0.93***			
Know Parental Educ. (0/1)				0.96	0.01	0.96***
Age in First Year	18.11	18.09	0.03***	18.12	18.07	0.05***
Minority (0/1)	0.37	0.21	0.16***	0.37	0.19	0.18***
US Citizen (0/1)	0.97	0.97	0.01***	0.97	0.97	0.00***
ACT (or SAT eq.)	24.52	25.53	-1.01***	24.47	25.69	-1.22***
High School GPA	3.40	3.38	0.03***	3.40	3.38	0.02***
First Maj. Stem (0/1)	0.28	0.24	0.04***	0.28	0.24	0.03***
Honors (0/1)	0.12	0.15	-0.04***	0.11	0.16	-0.04***
AZ High School (0/1)	0.66	0.62	0.04***	0.66	0.62	0.04***
First Year at ASU	2010	2008	2***	2010	2008	2***
Graduated (0/1)	0.65	0.71	-0.06***	0.64	0.73	-0.09***
Transfered (0/1)	0.14	0.11	0.03***	0.15	0.10	0.04***
Dropout (0/1)	0.20	0.18	0.03***	0.21	0.16	0.05***
Enrolled (0/1)	0.01	0.01	0.00***	0.01	0.01	0.00***

Notes: Table reports average values of observable characteristics by parental education and household income reporting status. Columns 1 and 2 show means for students with known and unknown parental education, respectively. Column 3 reports the difference in means between these two groups. Columns 4 and 5 show means for students with known and unknown household income, respectively. Column 6 reports the corresponding difference in means. Stars denote statistical significance. The sample is restricted to students in the 2000–2017 cohorts. ***: $p<0.01$, **: $p<0.05$, *: $p<0.10$.

Table A2: ASU vs Large Flagship Universities

	ASU (1)	Flagship Univ. (2)	P-value (3)
Observations	42,895	1,381,857	
Female	0.51	0.50	0.00
Black	0.04	0.06	0.00
White	0.51	0.60	0.00
Hispanic	0.28	0.12	0.00
Int. Students	0.02	0.06	0.00
First Generation	0.32	-	-
Median Family Income	74,060	-	-
Freshman	0.31	-	-
Sophomore	0.26	-	-
Junior	0.20	-	-
Senior	0.24	-	-
ACT 25th %tile	21	24	-
ACT 75th %tile	28	30	-

Notes: Columns 1 and 2 report the average of each row variable for all ASU students enrolled in Fall 2019 and for U.S. flagship universities, respectively. Column 3 reports the p-value from a test of the difference in means. Flagship data are from IPEDS 2019. Income is reported in 2019 dollars.

Table A3: ASU and U.S. First-Generation Student Comparison

	ASU (1)	BPS12/17 (2)
First-Gen. Proportion	37.55	47.70
Among First-Gen.:		
Female	57.80	60.10
White	42.56	52.30
Black	5.96	18.80
Hispanic	39.19	18.40
Asian	5.74	5.30
ACT (or SAT eq.)	23.36	19.00
First Year GPA	2.74	2.81

Notes: Columns 1 and 2 report the average of each row variable for first-generation students in the ASU Fall 2011 cohort and in the BPS 12/17 dataset, respectively. BPS data come from the Beginning Postsecondary Students Longitudinal Study (National Center for Education Statistics, 2019), which follows a cohort of students who began their first year of postsecondary education in 2011–2012. The BPS sample is restricted to students attending four-year institutions. Students are classified as first-generation if neither parent holds a college degree.

Table A4: High School GPA and first-Generation Status on ACT and Expected GPA Performance

	ACT	Exp. 1st Semester GPA
	(1)	(2)
First-gen	-0.50*	0.05
	(0.26)	(0.22)
High School GPA (HS GPA)	3.23***	0.30***
	(0.05)	(0.04)
First-Gen × HS GPA	-0.40***	-0.03
	(0.08)	(0.06)
Major FE	✓	
STEM Major FE		✓
Dep. mean	24.46	3.61
N	62,935	1,304
R2	0.28	0.11

Notes: Both specifications regress the outcome variable (ACT score in Column 1 and expected first-semester GPA in Column 2) on a first-generation indicator, high school GPA, and their interaction. Column 1 uses administrative data from the 2000 to 2017 cohorts and includes six-digit CIP major fixed effects. Column 2 uses survey data, restricted to Round 1 respondents with known parental education and high school GPA; the specification includes a STEM major indicator. Robust standard errors are reported in parentheses. ***: $p<0.01$, **: $p<0.05$, *: $p<0.10$.

Table A5: Achievement Gaps by Income

	Lower-Income (1)	Higher-Income (2)	Basic Difference (3)	CEM Weighted Difference (4)
Observations	48,811	44,753	93,564	44,822
Panel A) Baseline Characteristics				
Household Income (1,000s $2019)[†] [m]	31.94	169.62	-137.68***	-116.52***
In State (0/1)[m]	0.77	0.58	0.18***	0.00
Race: Black (0/1)[m]	0.08	0.04	0.04***	0.00
Race: Latino/Hispanic (0/1)[m]	0.30	0.17	0.13***	-0.00
Race: White/Asian (0/1)[m]	0.54	0.73	-0.19***	-0.00
Female (0/1)[m]	0.55	0.53	0.02***	0.00
ACT (or SAT eq.)[m]	23.59	25.39	-1.80***	-0.08
HS GPA[m]	3.37	3.44	-0.07***	-0.01
Financial Aid (0/1)[m]	0.94	0.89	0.04***	-0.00
Transfer Credits (0/1)[m]	0.27	0.28	-0.01***	-0.00
Panel B) Educational Outcomes				
Graduated (0/1)	0.62	0.65	-0.03***	0.01
Graduated in 4 (0/1)	0.43	0.53	-0.11***	-0.06***
Transferred (0/1)	0.12	0.17	-0.05***	-0.03***
Dropped-out (0/1)	0.24	0.17	0.08***	0.02***
Years before dropout | dropout	1.81	1.53	0.28***	0.38***
Graduated in STEM major (0/1)	0.13	0.14	-0.01***	0.01***
Panel C) Intermediate Outcomes				
Ever Switch Major	0.47	0.43	0.04***	0.03***
Ever Switch Major | dropout	0.34	0.31	0.03***	0.04***
Ever Switch Major | grad	0.55	0.50	0.04***	0.02***
ASU semesters before in grad major | grad	2.10	1.77	0.33***	0.22***
Move into Undec.	0.05	0.04	0.01***	0.00**
Started in STEM major (0/1)	0.28	0.28	0.00	0.00
Twice Off Track (0/1)	0.26	0.22	0.05***	0.02***
Ever Academic Probation (0/1)	0.29	0.19	0.10***	0.02***
Num W per W/F[††]	0.53	0.58	-0.05***	0.00
First semester GPA[m]	2.81	3.02	-0.21***	-0.01

Notes: Sample includes all in-person first-time-freshmen students first enrolling in the 2000-2017 semesters. Columns 1 and 2 report the average of each row variable for low and high-income students, respectively. Low-income is defined as household income below the sample median; high-income is above the median. Column 3 shows the unweighted mean difference between the two groups, and Column 4 presents the mean difference using the CEM-weighted sample. Stars denote statistical significance. ***: p<0.01, **: p<0.05, *: p<0.10 ***: p<0.01, **: p<0.05, *: p<0.10

[†] Income winsorized below 1st percentile ($0) and above 99th percentile (∼ $680,000)

[††] Number of withdrawn courses divided by number of courses withdrawn or failed.

[m] Matching variables.

Table A6: Achievement Gaps by Other Demographic Groups

	Minority (1)	Non-minority (2)	Diff. (3)	Female (4)	Male (5)	Diff. (6)
Observations	32,899	57,141		48,850	41,190	
Panel A) Baseline Characteristics						
Household Income (1,000s $2019)†	75.41	112.40	-36.98***	95.44	102.96	-7.52***
In State (0/1)	0.73	0.65	0.08***	0.70	0.65	0.04***
Race: Black (0/1)	0.16	0.00	0.16***	0.06	0.06	0.00***
Race: Latino/Hispanic (0/1)	0.65	0.00	0.65***	0.24	0.23	0.02***
Race: White/Asian (0/1)	0.00	1.00	-1.00	0.62	0.65	-0.03***
Female (0/1)	0.56	0.53	0.03***	1.00	0.00	1.00
ACT (or SAT eq.)	22.96	25.47	-2.51***	23.95	25.19	-1.25***
HS GPA	3.34	3.44	-0.10***	3.46	3.34	0.13***
Financial Aid (0/1)	0.95	0.90	0.05***	0.92	0.91	0.01***
Transfer Credits (0/1)	0.25	0.30	-0.04***	0.31	0.25	0.06***
Panel B) Educational Outcomes						
Graduated (0/1)	0.60	0.67	-0.08***	0.68	0.61	0.07***
Graduated in 4 (0/1)	0.43	0.52	-0.09***	0.53	0.43	0.10***
Transferred (0/1)	0.14	0.14	-0.01***	0.13	0.15	-0.02***
Dropped-out (0/1)	0.25	0.17	0.08***	0.18	0.23	-0.05***
Years before dropout | dropout	1.75	1.77	-0.02	1.70	1.82	-0.12***
Graduated in STEM major (0/1)	0.12	0.15	-0.04***	0.10	0.19	-0.09***
Panel C) Intermediate Outcomes						
Ever Switch Major	0.46	0.45	0.00	0.46	0.44	0.02***
Ever Switch Major | dropout	0.35	0.34	0.01**	0.34	0.35	-0.01
Ever Switch Major | grad	0.53	0.52	0.01**	0.53	0.52	0.00
ASU semesters before in grad major | grad	1.95	1.92	0.03	1.88	2.00	-0.12***
Move into Undec.	0.05	0.04	0.01***	0.04	0.05	-0.02***
Started in STEM major (0/1)	0.28	0.28	0.00	0.20	0.38	-0.18***
Twice Off Track (0/1)	0.28	0.22	0.06***	0.22	0.27	-0.05***
Ever Academic Probation (0/1)	0.31	0.20	0.11***	0.20	0.28	-0.07***
First semester GPA	2.75	3.03	-0.28***	3.01	2.83	0.18***
Num W per W/F††	0.50	0.59	-0.09***	0.57	0.53	0.04***

† Income winsorized below 1st percentile ($0) and above 99th percentile (∼ $680,000)
†† Number of withdrawn courses divided by number of courses withdrawn or failed.
Notes: Sample includes all in-person first-time freshmen from the 2000–2017 entering cohorts. Columns 1 and 2 report the average of each row variable for minority and non-minority students, and column 3 reports the difference in means. Minority is defined as students who are neither White nor Asian. Columns 4 and 5 report averages for female and male students, respectively, and column 6 reports the corresponding difference. Stars denote statistical significance. ***: p<0.01, **: p<0.05, *: p<0.10

Table A7: Top 10 Majors by First-Generation Status

	First-Gen (1)	Continue-Gen (2)
Panel A) First Semester		
1	Business, Management, Marketing, and Related Support Services, Other.	Business, Management, Marketing, and Related Support Services, Other.
2	Biology/Biological Sciences, General.	Biology/Biological Sciences, General.
3	Psychology, General.	Psychology, General.
4	Registered Nursing/Registered Nurse.	Journalism, Other.
5	Biochemistry.	Mechanical Engineering.
6	Criminal Justice/Law Enforcement Administration.	Registered Nursing/Registered Nurse.
7	Architecture.	Computer Science.
8	Computer Science.	Bioengineering and Biomedical Engineering.
9	Mechanical Engineering.	Architecture.
10	Journalism, Other.	Biochemistry.
Panel B) At Graduation		
1	Psychology, General.	Biology/Biological Sciences, General.
2	Biology/Biological Sciences, General.	Psychology, General.
3	Business, Management, Marketing, and Related Support Services, Other.	Business, Management, Marketing, and Related Support Services, Other.
4	Multi-/Interdisciplinary Studies, Other.	Multi-/Interdisciplinary Studies, Other.
5	Speech Communication and Rhetoric.	Journalism, Other.
6	Criminal Justice/Law Enforcement Administration.	Speech Communication and Rhetoric.
7	Elementary Education and Teaching.	Marketing/Marketing Management, General.
8	Political Science and Government, General.	Finance, General.
9	Registered Nursing/Registered Nurse.	Political Science and Government, General.
10	Journalism, Other.	Registered Nursing/Registered Nurse.

Notes: Top 10 most common majors based on six-digit CIP codes during the first semester and at graduation, by first-generation status. Listed in descending order of frequency. Based on data from the 2000–2017 entering cohorts.

Table A8: Matching Variables by Parental Education for Full and Weighted CEM Samples

	Full Sample (1)	CEM Weighted Sample (2)	Difference (3)
Observations	90,401	19,411	
Household Income (1,000s \$2019)†	98.88	76.73	22.15***
In State (0/1)	0.68	0.79	-0.11***
Race: Black (0/1)	0.06	0.03	0.03***
Race: Latino/Hispanic (0/1)	0.24	0.20	0.04***
Race: White/Asian (0/1)	0.63	0.75	-0.12***
Female (0/1)	0.54	0.54	0.00
ACT (or SAT eq.)	24.52	24.33	0.19
HS GPA	3.40	3.40	0.00
Financial Aid (0/1)	0.92	0.96	-0.04***
Transfer Credits (0/1)	0.28	0.22	0.06***
First semester GPA	2.93	2.83	0.10***

† Income winsorized below 1st percentile (\$0) and above 99th percentile (∼ \$680,000)

Notes: Table reports the mean of each observable characteristic used in the CEM matching procedure. The sample includes all in-person first-time freshmen from the 2000–2017 entering cohorts. Column 1 reports averages for the full sample, Column 2 for students with positive CEM weights, and Column 3 shows the difference in means. Stars denote statistical significance. ***: p<0.01, **: p<0.05, *: p<0.10

Table A9: Expectation Revision by Realized GPA

	GPA Expectation Gap		Revision: GPA by Spring 2023		Revision: GPA by Graduation	
Sample:	gpa < $\mathbb{E}$	gpa $\geq$ $\mathbb{E}$	gpa < $\mathbb{E}$	gpa $\geq$ $\mathbb{E}$	gpa < $\mathbb{E}$	gpa $\geq$ $\mathbb{E}$
	(1)	(2)	(3)	(4)	(5)	(6)
First gen. (0/1)	-0.058	-0.052	-0.160**	0.123**	-0.198***	0.118**
	(0.086)	(0.044)	(0.063)	(0.055)	(0.064)	(0.051)
Demographic/STEM Controls	✓	✓	✓	✓	✓	✓
GPA Expectation Gap Control			✓	✓	✓	✓
Dep. mean	-0.482	0.278	-0.186	0.097	-0.109	0.071
N	157	262	157	262	157	262
R^2	0.081	0.097	0.169	0.206	0.150	0.128
Adjusted R^2	0.025	0.065	0.112	0.174	0.091	0.094

Notes: Table reports coefficients from regressions of the outcome indicated at the top of each column on a first-generation indicator. Controls include the logarithm of household income, high school GPA, gender, a non-white indicator, a STEM major indicator, and indicators for missing values in control variables. The sample includes students who completed both survey rounds and have known parental education. The sample is split by whether realized first-semester GPA is lower than expected (Columns 1, 3, and 5) or greater than or equal to expected (Columns 2, 4, and 6), based on Round 1 GPA expectations. Robust standard errors are reported in parentheses. ***: p<0.01, **: p<0.05, *: p<0.10

Table A10: Lower Income Students and Subjective GPA Uncertainty

	Dep Variable: Variance in Subj. GPA Expectations			
	Distriubtional Assumptions			
	Uniform Mass Within Bin		Fitted Beta Distribution	
	(1)	(2)	(3)	(4)
Low Income (0/1)	0.024***	0.007	0.015*	0.003
	(0.008)	(0.010)	(0.008)	(0.009)
Controls		✓		✓
Dep. mean	0.118	0.118	0.075	0.075
N	1,316	1,316	1,316	1,316
R^2	0.007	0.049	0.003	0.020
Adjusted R^2	0.006	0.041	0.002	0.013

Notes: Table shows estimates from regressions of the variance of the distribution of round 1 GPA beliefs onto a lower-income indicator (below median income). Column (2) includes controls for a non-white indicator, gender, the logarithm of household income, high school GPA, and major fixed effects. Robust standard errors are in parentheses. Variance winsorized above the 99th percentile. ***: p<0.01, **: p<0.05, *: p<0.10

Table A11: Correlation of Subjective Beliefs about Major Earnings: 2023 Survey

	First-Gen	Continue-Gen	p-value (1)=(2)
	(1)	(2)	(3)
obs.	187	421	
		Subjective Beliefs	
Dep Var: Business Earn			
STEM Earn	0.755	0.559	0.009
	(0.048)	(0.036)	
Health Earn	0.562	0.369	0.030
	(0.060)	(0.037)	
Social Sci. Earn	1.065	0.831	0.080
	(0.090)	(0.064)	
Dep Var: STEM Earn			
Health Earn	0.653	0.539	0.142
	(0.055)	(0.036)	
Social Sci. Earn	1.020	0.774	0.071
	(0.092)	(0.072)	
Dep Var: Health Earn			
Social Sci. Earn	0.811	0.879	0.631
	(0.105)	(0.079)	
		Subjective Beliefs: Residualized	
Dep Var: Business Earn			
STEM Earn	0.748	0.534	0.004
	(0.048)	(0.036)	
Health Earn	0.572	0.358	0.013
	(0.059)	(0.035)	
Social Sci. Earn	1.061	0.779	0.036
	(0.093)	(0.064)	
Dep Var: STEM Earn			
Health Earn	0.675	0.543	0.071
	(0.054)	(0.035)	
Social Sci. Earn	1.037	0.746	0.026
	(0.096)	(0.073)	
Dep Var: Health Earn			
Social Sci. Earn	0.820	0.921	0.500
	(0.107)	(0.080)	

Notes: Data come from a separate survey of ASU freshman students conducted during the Fall 2023 semester. Each cell reports the coefficient from a separate regression of the indicated dependent variable (expected earnings at age 35 if one were to graduate with the major as denoted by the dependent variable) on the row-specific independent variable (expected earnings at age 35 if one were to graduate with the row major). The exact wording of the survey question that elicited earnings expectations was: *"Below is a list of major categories at ASU. You are asked to report how much you expect YOUR annual income at age 35 to be, if you graduate with a major in that category. When answering, assume a dollar when you are 35 is worth a dollar today, i.e., ignore the effects of inflation."* Column 1 presents estimates for first-generation students, Column 2 for continuing-generation students, and Column 3 reports the p-value from a test of equality between the two coefficients. Robust standard errors are reported in parentheses. The upper panel presents correlations within raw expectations. The lower panel presents correlations within residualized expectations derived from regressions of expectations onto log household income, race (a binary minority variable), gender, age, high school GPA, and ACT (or SAT equivalent) scores. Expectations residualized separately by first-generation status. ***: p<0.01, **: p<0.05, *: p<0.10

Table A12: Expectation Revision by Realized GPA: Network Size Effect

	Revision: GPA by Spring 2023		Revision: GPA by Graduation	
Sample:	gpa $<\mathbb{E}$	gpa $\geq \mathbb{E}$	gpa $<\mathbb{E}$	gpa $\geq \mathbb{E}$
	(1)	(2)	(3)	(4)
First gen. (0/1)	-0.108	0.070	-0.169**	0.061
	(0.071)	(0.067)	(0.076)	(0.062)
Network Size Bottom Tercile (0/1)	-0.051	0.021	-0.042	-0.009
	(0.067)	(0.047)	(0.069)	(0.047)
First gen. × Network Size Bottom Tercile	-0.102	0.123	-0.055	0.138*
	(0.113)	(0.080)	(0.104)	(0.074)
Dep. mean	-0.186	0.097	-0.109	0.071
Obs.	157	262	157	262
R2	0.195	0.221	0.161	0.142
Adjusted R2	0.128	0.184	0.091	0.101

Notes: Table reports coefficients from regressions of the outcome indicated at the top of each column on a first-generation indicator, a bottom-tercile network size indicator, and their interaction. All specifications control for the logarithm of household income, high school GPA, gender, a non-white indicator, a STEM major indicator, indicators for missing values in control variables, and the size of the GPA expectations gap from Round 1. The sample includes students who completed both survey rounds and have known parental education. The sample is split by whether realized first-semester GPA is lower than expected (Columns 1 and 3) or greater than or equal to expected (Columns 2 and 4), based on Round 1 GPA expectations. Robust standard errors are in parentheses. ***: p<0.01, **: p<0.05, *: p<0.10

Table A13: Belief Updating: Marginal Effects of Grade Signal and Priors on Posterior Beliefs

<table>
<tr><th></th><th colspan="4">R2 Belief:
Likelihood NOT Return Y2</th><th colspan="4">R2 Belief:
Likelihood NOT Grad in 4 Years</th></tr>
<tr><th></th><th>ME |
FG=1</th><th>ME |
FG=0</th><th>ME |
FG=1</th><th>ME |
FG=0</th><th>ME |
FG=1</th><th>ME |
FG=0</th><th>ME |
FG=1</th><th>ME |
FG=0</th></tr>
<tr><th></th><th>(1)</th><th>(2)</th><th>(3)</th><th>(4)</th><th>(5)</th><th>(6)</th><th>(7)</th><th>(8)</th></tr>
<tr><td>R1 Prior</td><td>0.27***</td><td>0.15**</td><td>0.25***</td><td>0.16**</td><td>0.32***</td><td>0.30***</td><td>0.32***</td><td>0.31***</td></tr>
<tr><td></td><td>(0.07)</td><td>(0.06)</td><td>(0.06)</td><td>(0.07)</td><td>(0.09)</td><td>(0.06)</td><td>(0.09)</td><td>(0.06)</td></tr>
<tr><td></td><td colspan="2">p=0.267</td><td colspan="2">p=0.274</td><td colspan="2">p=0.828</td><td colspan="2">p=0.829</td></tr>
<tr><td>First-Semester GPA</td><td>-0.06***</td><td>-0.02</td><td>-0.05***</td><td>-0.02</td><td>-0.09***</td><td>-0.02</td><td>-0.08***</td><td>-0.01</td></tr>
<tr><td></td><td>(0.01)</td><td>(0.02)</td><td>(0.01)</td><td>(0.02)</td><td>(0.02)</td><td>(0.01)</td><td>(0.02)</td><td>(0.02)</td></tr>
<tr><td></td><td colspan="2">p=0.059</td><td colspan="2">p=0.088</td><td colspan="2">p=0.013</td><td colspan="2">p=0.011</td></tr>
<tr><td>Log Exp. Inc | No Grad</td><td>-0.01</td><td>0.01</td><td>-0.00</td><td>0.01</td><td>0.01</td><td>-0.01</td><td>0.01</td><td>-0.01</td></tr>
<tr><td></td><td>(0.02)</td><td>(0.02)</td><td>(0.02)</td><td>(0.02)</td><td>(0.02)</td><td>(0.02)</td><td>(0.02)</td><td>(0.02)</td></tr>
<tr><td></td><td colspan="2">p=0.431</td><td colspan="2">p=0.611</td><td colspan="2">p=0.627</td><td colspan="2">p=0.439</td></tr>
<tr><td>R1: Expected Sem1 GPA</td><td>-0.05**</td><td>-0.05**</td><td>-0.05**</td><td>-0.05**</td><td>-0.05**</td><td>-0.05**</td><td>-0.05**</td><td>-0.05**</td></tr>
<tr><td></td><td>(0.02)</td><td>(0.02)</td><td>(0.02)</td><td>(0.02)</td><td>(0.02)</td><td>(0.03)</td><td>(0.03)</td><td>(0.03)</td></tr>
<tr><td>Controls</td><td></td><td></td><td>✓</td><td>✓</td><td></td><td></td><td>✓</td><td>✓</td></tr>
<tr><td>Obs.</td><td>417</td><td>417</td><td>417</td><td>417</td><td>417</td><td>417</td><td>417</td><td>417</td></tr>
<tr><td>Y mean</td><td>0.060</td><td>0.060</td><td>0.060</td><td>0.060</td><td>0.105</td><td>0.105</td><td>0.105</td><td>0.105</td></tr>
</table>

Notes: Table reports marginal effects (ME) of prior beliefs, first-semester grades, and grade expectations on posterior beliefs on the outcomes indicated at the top of the columns, estimated separately for first-generation (FG = 1) and continuing-generation (FG = 0) students. Effects are estimated using a fully saturated fractional probit regression for each outcome. The sample includes students who participated in both survey rounds and have known parental education, and is weighted to balance the first- and continuing-generation samples across ten GPA bins. Controls include the logarithm of household income, high school GPA, gender, a non-white indicator, a STEM major indicator, and indicators for missing values in control variables. Robust standard errors are reported in parentheses. ***: $p<0.01$, **: $p<0.05$, *: $p<0.10$

Table A14: CEM Selection on Observables

	Dropout (0/1)	Dropout (0/1)	Graduate (0/1)	Graduate (0/1)	Swtich Majors (0/1)	Switch Majors (0/1)
	(1)	(2)	(3)	(4)	(5)	(6)
First-Gen	0.057***	0.057***	-0.043***	-0.044***	-0.022*	-0.024**
	(0.012)	(0.012)	(0.013)	(0.012)	(0.013)	(0.012)
Controls	N	Y	N	Y	N	Y
δ (Min. Selection on Unobs. : $\beta_{FGxEvent} = 0$)	6.114		9.232		-3.017	
N	7,414	7,414	7,414	7,414	7,414	7,414
r^2	0.004	0.096	0.002	0.120	0.000	0.165

Notes: Each column reports coefficients from a regression of the outcome variable indicated at the top on a first-generation indicator. The sample is the CEM-matched sample for the 2000–2017 entering cohorts who received at least one C or lower grade in their first year. Controls include gender, honors indicator, household income, highest standardized test score (ACT or SAT equivalent), high school GPA, high school Math GPA, high school English GPA, number of transfer credits, athlete indicator, age, military veteran indicator, U.S. citizenship status, and fixed effects for major, cohort, and high school state. Robust standard errors are reported in parentheses. δ represents the minimum degree of selection on unobservables relative to selection on observables needed to explain away the first-generation coefficient, following the method in Oster (2019).

Table A15: LEAD Regression Discontinuity Estimates: Placebo Cohorts

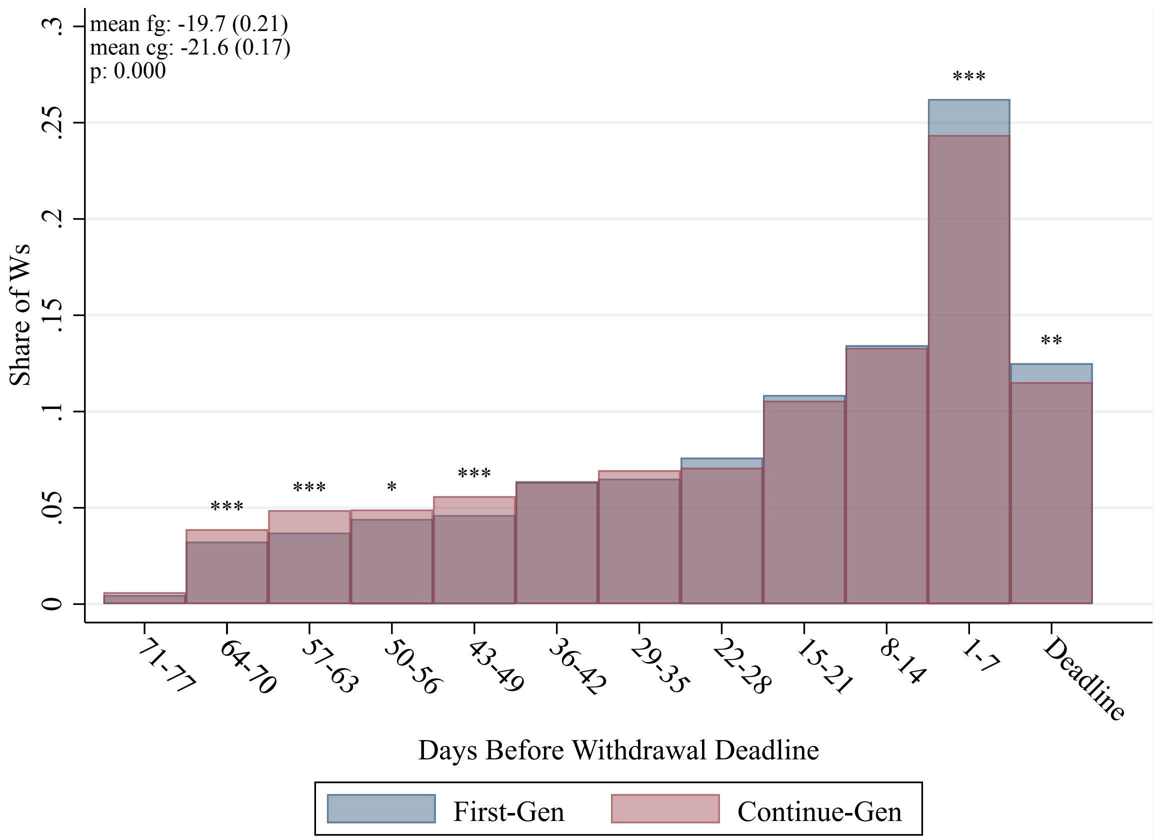
	1st sem GPA	Cum GPA in Last Sem.	Cum GPA in Last Sem. (excl. 1st yr)	Finish Year 2 (0/1)	Grad. (0/1)	Grad. in 4 Years (0/1)	Ever Switch Major (0/1)	Ever Switch into Undec. (0/1)	Median Grad. Maj. Salary (1,000s)†
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Nudged (ITT)	-0.06	-0.07	-0.09*	-0.02	0.01	-0.04	-0.02	0.02	-0.67
	(0.06)	(0.06)	(0.05)	(0.01)	(0.01)	(0.03)	(0.01)	(0.01)	(0.99)
Observations	19,411	19,518	19,667	17,889	19,667	14,776	19,667	12,832	7,288
Within BW Mean	2.67	2.63	2.17	0.74	0.57	0.31	0.47	0.07	48.10
Optimal BW	8	8	8	7	8	6	8	5	5

† Missing major salaries estimated by LASSO on the course content of the major.

Notes: Table reports robust bias-corrected sharp RD (ITT) estimates of the effect of LEAD participation on key academic outcomes. Outcomes are indicated at the top of each column. Estimates use coverage-error optimal bandwidths following Calonico et al. (2020). All specifications control for cohort fixed effects, high school GPA, and transfer credits. The sample includes students in the 2003–2015 cohorts with reported high school GPA and standardized test scores in the administrative data, and with 12 or fewer transfer credits. Standard errors, clustered at the CI score (running variable) level, are reported in parentheses.***: $p<0.01$, **: $p<0.05$, *: $p<0.10$

Appendix A - Figures

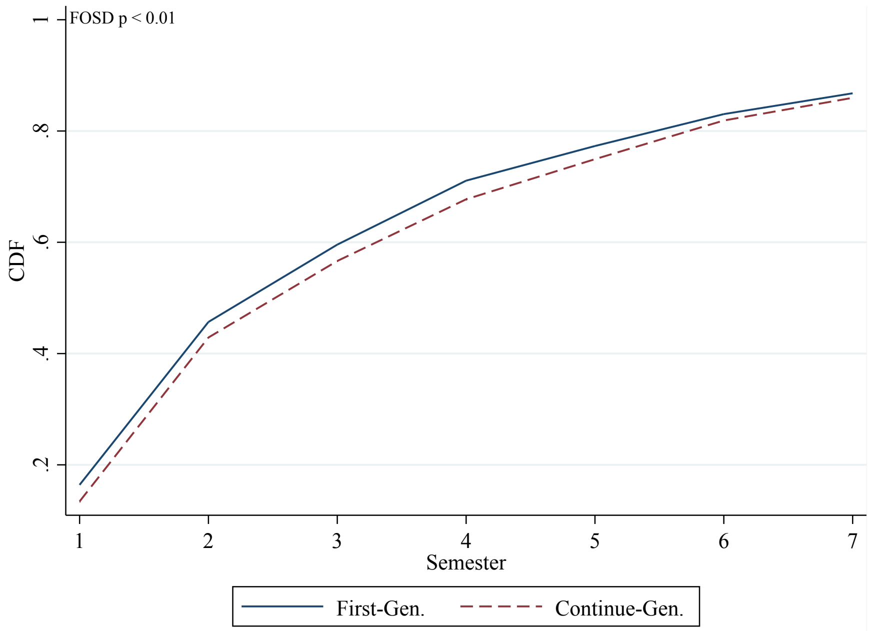
Figure A1: Timing of Withdrawal



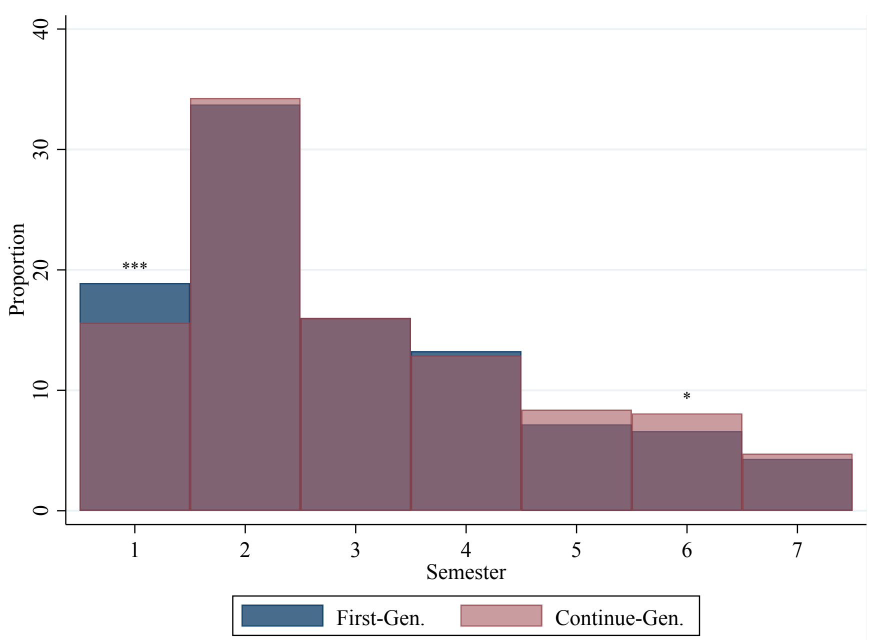
Notes: Figure plots the share of all course withdrawals during the first semester at ASU that occurred by week to the withdrawal deadline separately by parental education. The sample includes all first-semester course observations for the 2000-2017 cohorts. Stars denote statistical significance between groups. ***: p<0.01, **: p<0.05, *: p<0.10

Figure A2: Dropout Timing Conditional on Dropout

Panel A) CDF

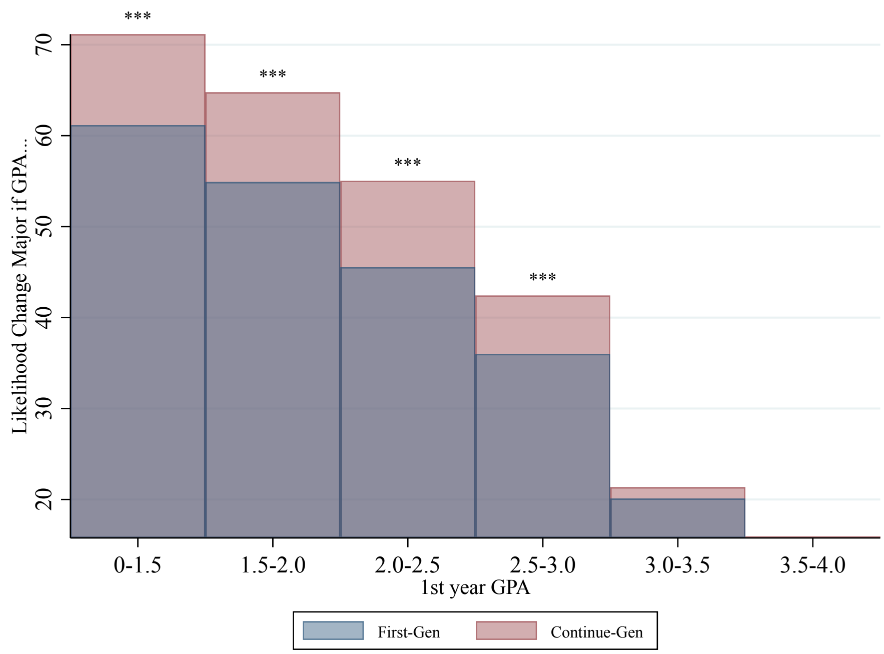


Panel B) Histrogram



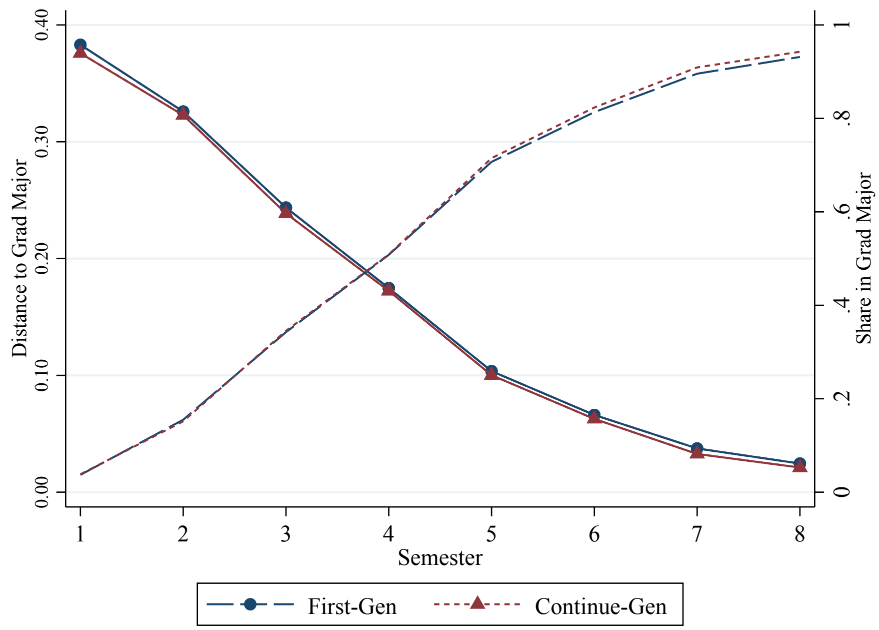
Notes: Figure plots two panels based on the CEM (weighted) sample. Panel A shows the cumulative distribution function (CDF) of the semester in which students dropped out, conditional on dropout, for first-generation and continuing-generation students (up to semester 7). FOSD p is the p-value from a first-order stochastic dominance test following Davidson and Duclos (2013). The test rejects the null in favor of the hypothesis that the distribution for first-generation students FOSD that of continuing-generation students. Panel B displays the proportion of first-generation and continuing-generation students who dropped out in each semester, conditional on dropping out before their 8th semester. Stars indicate the significance level of the difference between the two groups. ***: $p<0.01$, **: $p<0.05$, *: $p<0.10$

Figure A3: Baseline Intent to Change Major Conditional on First Year GPA



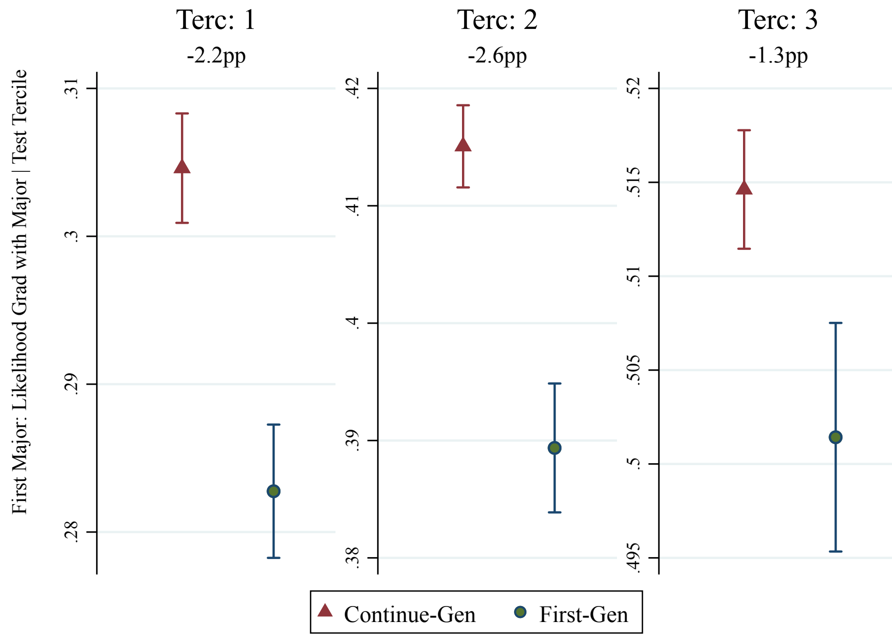
Notes: Figure plots the average intended likelihood of changing major reported by first- and continuing-generation students at the start of their first semester. Each student was asked to report their perceived likelihood of switching majors if their GPA at the end of their first year fell into each of the labeled bins.

Figure A4: Difference in Major Content Across Major Switches Conditional on Graduation



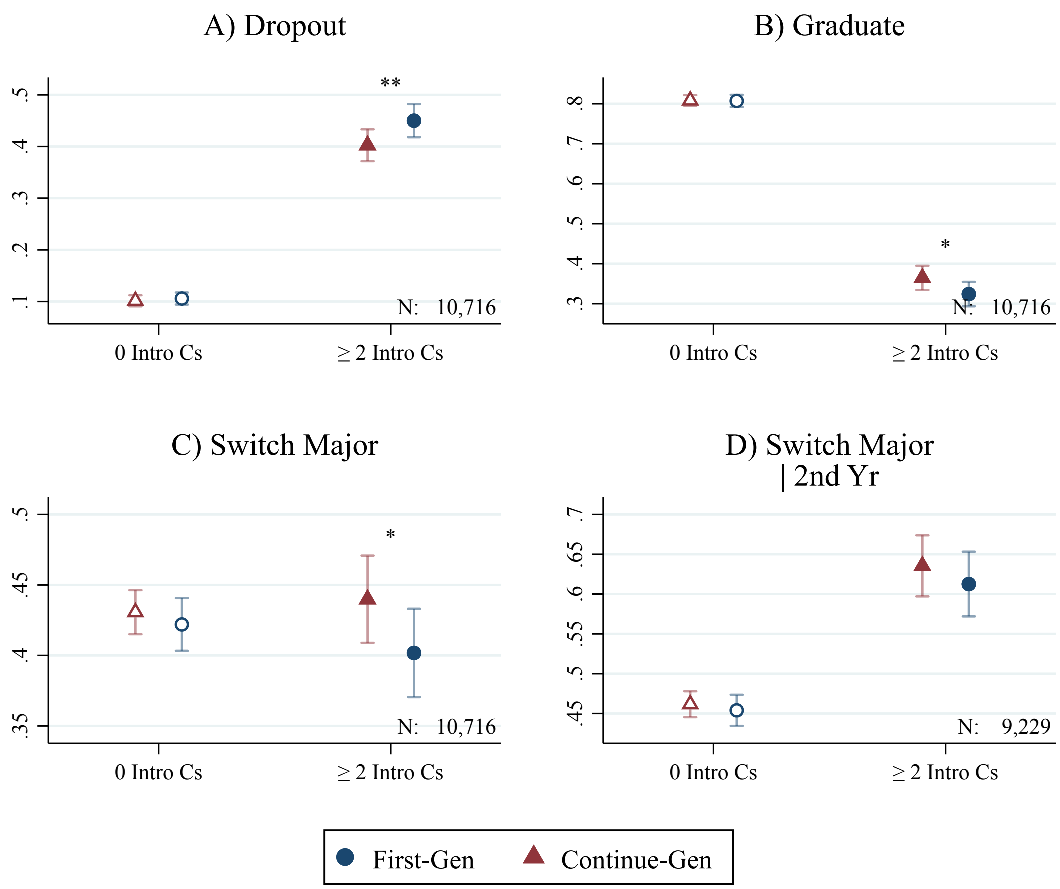
Notes: Connected scatterplot lines represent the distance in course content between initial and terminal major for college graduates, conditional on ever switching majors (left axis). The sample includes all students in the 2000-2017 cohorts. Course content for a major is measured as the average share of credit hours in each academic department for never-switching graduates. Distances are measured as Euclidean distances between Course content vectors. Economics to Finance is a distance of 0.23 while Economics to Social Work is a distance of 0.56. Dashed lines report the share of students in their terminal graduating major within the same sample (right axis)

Figure A5: Sorting into Initial Major by Levels of Math Preparation



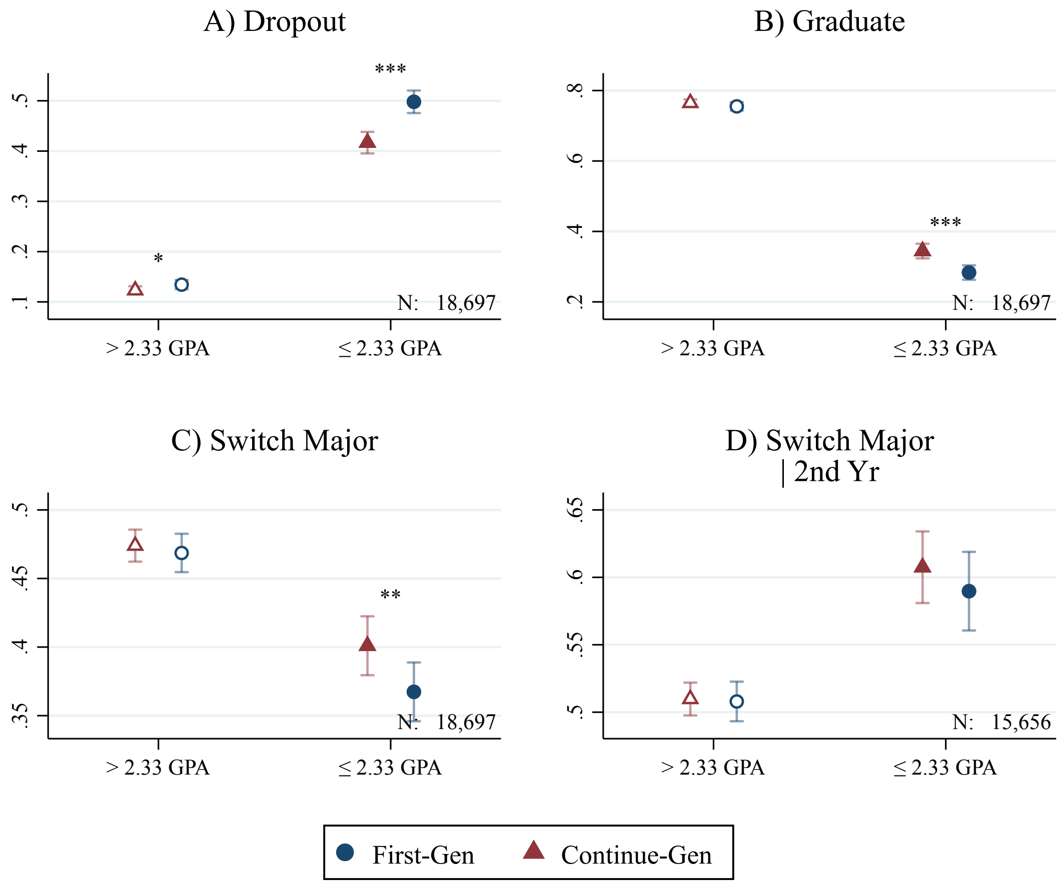
Notes: Figure plots the sorting into initial majors by students' math proficiency for the 2017-2021 cohorts. Major sorting is measured as the average major graduation rates for students in previous cohorts (2012-2016) with the same initial major and math tercile.

Figure A6: Reaction to low Grades: 2 or More C or Lower in Intro Courses



Notes: Figure plots the (weighted) average outcomes within the 2000-2017 cohorts. CEM weights were constructed to balance matching variables within first-year performance group. Poor (good) First-year performance is measured as (not) receiving at least two grades of C or worse in a student's major-specific introductory courses. Matching variables include: race indicators (non-minority, Black, Hispanic, and other), gender, initial major category (based on 2-digit CIP family group codes), Arizona residency, transfer credit status, financial aid, household income, high school GPA, ACT/SAT scores, and first-semester GPA. 95% confidence intervals indicated using I-bars. Panel A displays the share of students who drop out. Panel B displays the share of students who graduate. Panel C displays the share of students who ever switch from a declared terminal major to a different declared terminal major. Panel D displays the share of students who switch majors conditional on enrolling for strictly more than 2 semesters. Stars represent statistical differences between first-generation and continuing-generation weighted averages. ***: $p<0.01$, **: $p<0.05$, *: $p<0.10$

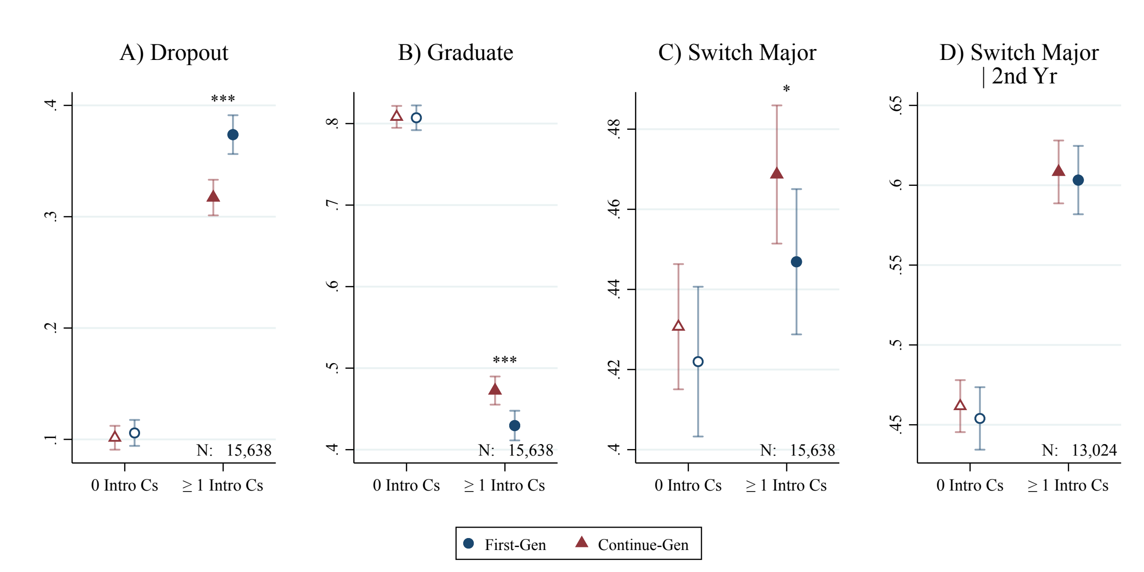
Figure A7: Reaction to Low Grades: First Semester GPA $\leq$ 2.33



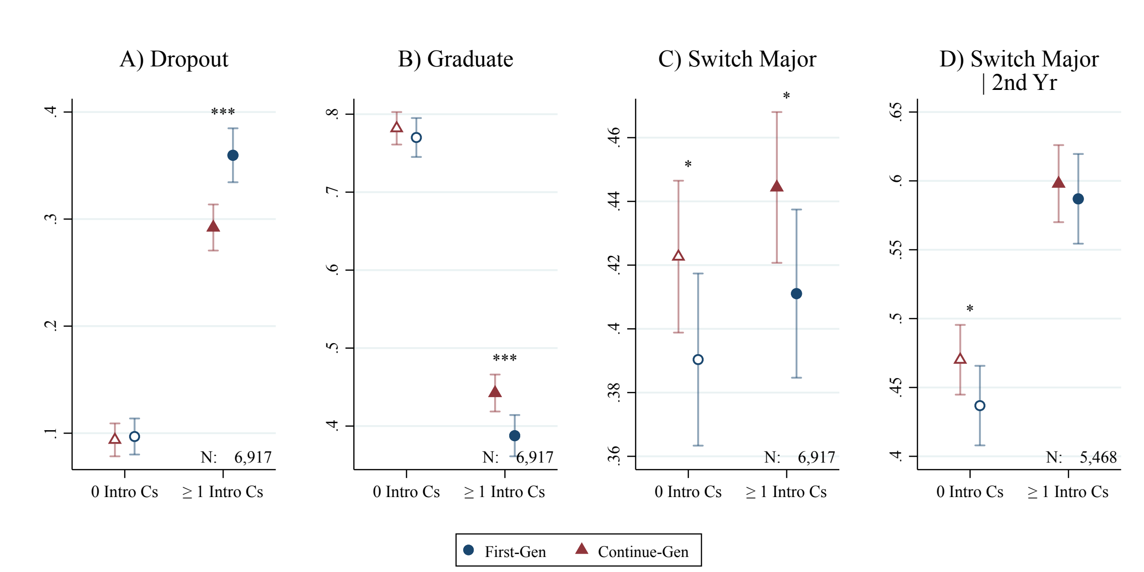
Notes: Figure plots the (weighted) average outcomes within the 2000-2017 cohorts. CEM weights were constructed to balance matching variables within first-year performance group. Poor (good) First-year performance is measured as (not) receiving a cumulative GPA of 2.33 or less in the first semester. Matching variables include: race indicators (non-minority, Black, Hispanic, and other), gender, initial major category (based on 2-digit CIP family group codes), Arizona residency, transfer credit status, financial aid, household income, high school GPA, ACT/SAT scores, and first-semester GPA. 95% confidence intervals indicated using I-bars. Panel A displays the share of students who drop out. Panel B displays the share of students who graduate. Panel C displays the share of students who ever switch from a declared terminal major to a different declared terminal major. Panel D displays the share of students who switch majors conditional on enrolling for strictly more than 2 semesters. Stars represent statistical differences between first-generation and continuing-generation weighted averages. ***: p<0.01, **: p<0.05, *: p<0.10

Figure A8: Reaction to Low Grades: One or More C or Lower in Intro Courses - Matching on Neighborhood Income

In-State Sample: Standard Matching Variables

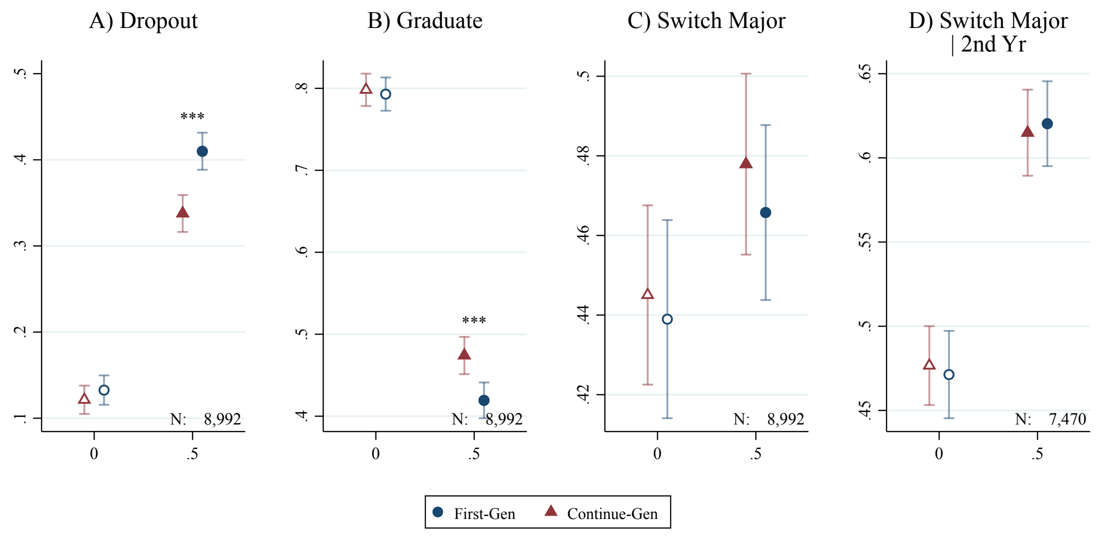


In-State Sample: Including Neighborhood Income Match



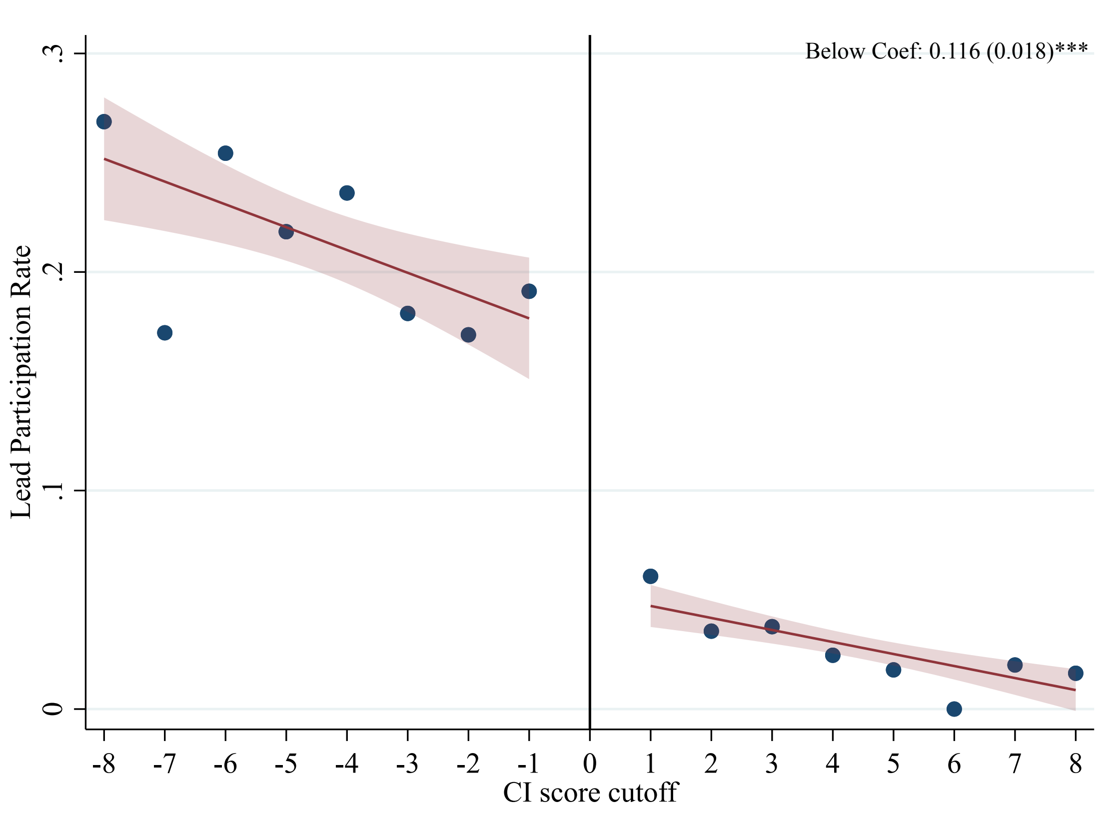
Notes: Figure plots the (weighted) average outcomes for In-state students with known high-school locations matched using two different sets of CEM-match variables. The upper panel matches on the standard set of matching variables in Figure 5. The lower panel matches on the same variables plus high school zipcode income (measured in 2014-2015). The measure of poor first-year performance corresponds to one or more C or D in intro courses. 95% confidence intervals indicated using I-bars. Stars represent statistical differences between first-generation and continuing-generation weighted averages. ***: p<0.01, **: p<0.05, *: p<0.10

Figure A9: Reaction to Low Grades: One or More C or D in Intro - Lower Income Sample



Notes: Figure plots the (weighted) average outcomes within sub-sample of below-median household income students in the 2000-2017 cohorts. Median household income is $74,280 in 2019 dollars. CEM weights were constructed to balance matching variables within first-year performance group. Poor (good) First-year performance is measured as (not) receiving a cumulative GPA of 2.33 or less in the first semester. Matching variables include: race indicators (non-minority, Black, Hispanic, and other), gender, initial major category (based on 2-digit CIP family group codes), Arizona residency, transfer credit status, financial aid, household income, high school GPA, ACT/SAT scores, and first-semester GPA. 95% confidence intervals indicated using I-bars. Panel A displays the share of students who drop out. Panel B displays the share of students who graduate. Panel C displays the share of students who ever switch from a declared terminal major to a different declared terminal major. Panel D displays the share of students who switch majors conditional on enrolling for strictly more than 2 semesters. Stars represent statistical differences between first-generation and continuing-generation weighted averages. ***: p<0.01, **: p<0.05, *: p<0.10

Figure A10: LEAD First Stage



Notes: Figure plots the share of students participating in at least one semester of the LEAD program by distance to the CI score nudge cut-off. The sample includes 2016-2018 cohorts.

B Appendix: CEM Implementation

B CEM Implementation

In the CEM approach, the discretized matching variables are used to group observations into unique bins based on their combinations of coarsened values. CEM weights are assigned to ensure that the total weight of each sub-sample matches the total weight of the continuing-generation/good grades observations within each bin. These weights are then adjusted so that the weighted sample matches the empirical joint distribution of coarsened variables for first-generation students. Let n_c^0, n_c^1 represent the number of continuing-generation and first-generation students, respectively in coarsened variable bin c. Then, the share of all sample observations in bin c is $s_c = \frac{n_c^0 + n_c^1}{\sum_k n_k^0 + n_k^1}$. Sample weights are assigned to observations in bin c such that the proportion of total weight in bin c matches the share of the ASU population sample in bin c (among all bins with non-zero weight):

$$w_c^0 = \frac{s_c}{n_c^0}$$
$$w_c^1 = \frac{s_c}{n_c^1}$$

In practice, weights are scaled up to avoid floating-point issues. When matching separate sub-populations, such as in section 6 where the sample is stratified by good and poor first-year grades, the weights for the good-grade (g) and poor-grade (p) observations are:

$$w_c^{0g} = \frac{s_c}{n_c^{0g}}$$
$$w_c^{1g} = \frac{s_c}{n_c^{1g}}$$
$$w_c^{0p} = \frac{s_c}{n_c^{0p}}$$
$$w_c^{1p} = \frac{s_c}{n_c^{1p}}$$

Assuming the level of coarsening is fine enough to exclude economically significant differences in match variables within a strata, differences in educational outcomes between the first- and continuing-generation groups can be interpreted as caused by factors other than those associated with the matched variables and measured relative to the average ASU student. The sample effectively narrows the set of potential causal mechanisms to only those that do not operate through observable factors, such as household resource constraints, racial bias, academic preparation, or differences in educational preferences associated with gender.

To test whether the baseline degree of coarsening remains fine enough to eliminate observable differences in the underlying variables, Table B1 reports the balance of our CEM matched sample across first-generation and continuing-generation students with good and poor grades in their first semester for targeted and untargeted characteristics. In each panel, the upper section reports the average for the (uncoarsened) targeted variables. There are no meaningful differences between the populations across those characteristics, indicating that the level of coarsening used is sufficient to eliminate differences in continuous measures. The second part reports the averages for characteristics not used in the matching algorithm (untargeted variables). Our four subsamples are the same in terms of age, the likelihood of being a US citizen, and the likelihood of military affiliation. There are also no economically relevant differences across the subsamples regarding census track characteristics like poverty rate, mean income percentile, and share of single-headed households.

Table B1: CEM Balance

	N	First-Gen Good Grades	Continue-Gen. Good Grades	First-Gen. Poor Grades	Continue-Gen. Poor Grades
	(1)	(2)	(3)	(4)	(5)
Panel A: 2 or More Cs or Ds in Intro Courses					
Targeted					
High School GPA	10,489	3.64	3.64	3.03	3.01
ACT (or SAT eq.)	7,556	26.19	26.30	21.72	21.87
Proportion instate	10,716	0.83	0.83	0.76	0.76
Female	10,716	0.60	0.60	0.42	0.42
Household hhincW (1ks)	10,716	81131.86	82211.90	65572.77	67038.31
Black	10,716	0.01	0.01	0.05	0.05
Latino/Hispanic	10,716	0.13	0.13	0.27	0.27
White/Asian	10,716	0.86	0.86	0.65	0.65
First semester GPA	10,575	3.49	3.49	1.63	1.65
Transfer Credits (0/1)	10,716	0.29	0.29	0.08	0.08
Financial Aid (0/1)	10,716	0.98	0.98	0.94	0.94
Untargeted					
Age	10,715	18.07	18.06	18.07	18.11
Proportion US citizen	10,716	0.97	0.97*	0.98	0.97*
Proportion military affiliation	10,716	0.00	0.01	0.01	0.01
2000 Poverty Rate†	8,358	0.09	0.08***	0.11	0.09***
Mean Pctile hhincW Dist. 2014-2015†	8,358	0.52	0.54***	0.51	0.52***
Share Single-Headed HH in 2000†	8,358	0.24	0.23***	0.27	0.25***
Panel B: One or More C or D in Intro Courses					
Targeted					
High School GPA	15,302	3.64	3.64	3.14	3.14
ACT (or SAT eq.)	10,620	26.19	26.30	22.25	22.33
Proportion instate	15,638	0.83	0.83	0.75	0.75
Female	15,638	0.60	0.60	0.46	0.46
Household hhincW (1ks)	15,638	81131.86	82211.90	70459.49	71506.79
Black	15,638	0.01	0.01	0.05	0.05
Latino/Hispanic	15,638	0.13	0.13	0.25	0.25
White/Asian	15,638	0.86	0.86	0.67	0.67
First semester GPA	15,459	3.49	3.49	2.07	2.11
Transfer Credits (0/1)	15,638	0.29	0.29	0.11	0.11
Financial Aid (0/1)	15,638	0.98	0.98	0.94	0.94
Untargeted					

Table B1: (continued)

	N	First-Gen Good Grades	Continue-Gen. Good Grades	First-Gen. Poor Grades	Continue-Gen. Poor Grades
	(1)	(2)	(3)	(4)	(5)
Age	15,637	18.07	18.06	18.09	18.14**
Proportion US citizen	15,638	0.97	0.97*	0.97	0.97
Proportion military affiliation	15,638	0.00	0.01	0.01	0.01
2000 Poverty Rate†	11,811	0.09	0.08***	0.11	0.10***
Mean Pctile hhincW Dist. 2014-2015†	11,811	0.52	0.54***	0.51	0.52***
Share Single-Headed HH in 2000†	11,811	0.24	0.23***	0.26	0.25***
Panel C: First Semester GPA $\leq$ 2.33					
Targeted					
High School GPA	18,294	3.55	3.55	3.00	3.00
ACT (or SAT eq.)	12,668	25.18	25.28	21.20	21.35
Proportion instate	18,697	0.79	0.79	0.77	0.77
Female	18,697	0.57	0.57	0.46	0.46
Household hhincW (1ks)	18,697	81399.25	82305.21	63000.95	63771.42
Black	18,697	0.01	0.01	0.07	0.07
Latino/Hispanic	18,697	0.16	0.16	0.28	0.28
White/Asian	18,697	0.81	0.81	0.60	0.60
First semester GPA	18,557	3.33	3.33	1.40	1.47***
Transfer Credits (0/1)	18,697	0.26	0.26	0.10	0.10
Financial Aid (0/1)	18,697	0.98	0.98	0.92	0.92
Untargeted					
Age	18,696	18.06	18.05	18.14	18.17
Proportion US citizen	18,697	0.97	0.97**	0.97	0.97
Proportion military affiliation	18,697	0.01	0.01	0.01	0.01
2000 Poverty Rate†	13,991	0.09	0.08***	0.12	0.10***
Mean Pctile hhincW Dist. 2014-2015†	13,991	0.52	0.53***	0.50	0.52***
Share Single-Headed HH in 2000†	13,991	0.24	0.23***	0.27	0.26***

† Based on high school census track. It is only available for students who attended high school in Arizona.

Notes: Table reports CEM-weighted averages of observable characteristics used in the matching analysis, based on the CEM-matched sample for the 2000–2017 entering cohorts. The table presents means separately for explicitly matched (targeted) and unmatched (untargeted) variables. Column 2 corresponds to first-generation students with good first-semester grades, Column 3 to continuing-generation students with good grades, Column 4 to first-generation students with poor grades, and Column 5 to continuing-generation students with poor grades. Definitions of 'poor grades' vary across panels and are indicated at the top of each panel. Stars in Columns 3 and 5 indicate statistical significance from a test of the difference in means relative to Columns 2 and 4, respectively. Non-minority include White and Asian students.***: p<0.01, **: p<0.05, *: p<0.10

C Appendix: Model Details

C.1 Deriving Signal Thresholds

The model of student enrollment under uncertainty presented in Section 5.1 generates two relevant signal thresholds that characterize the agent's optimal decision. In this appendix section, we derive those thresholds based on the agent's posterior beliefs. We first begin with two propositions that characterize the signal values under which it is optimal for agent to dropout (Proposition 3) and switch majors (Proposition 4).

Proposition 3. *An agent will drop out if and only if both of the following hold:*

1. $\theta < v_o + \frac{\sigma_\theta^2}{\sigma_m^2}\left[v_o - \mu_{m,1}\right]$
2. $\theta < \mu_{m,1} + \frac{1}{\rho}\left(1 + \frac{\sigma_\theta^2}{\sigma_m^2}\right)\left[v_o - \mu_{m',1}\right]$

Proposition 4. *An agent will switch majors if and only if both of the following hold:*

1. $\theta < \mu_{m,1} + (\mu_{m',1} - \mu_{m,1})\left(\frac{1}{1-\rho}\right)\left(1 + \frac{\sigma_\theta^2}{\sigma_m^2}\right)$
2. $\theta \geq \mu_{m,1} + \frac{1}{\rho}\left(1 + \frac{\sigma_\theta^2}{\sigma_m^2}\right)\left[v_o - \mu_{m',1}\right]$

The values of θ that satisfy Proposition 4 fall between condition 2 (lower bound) and condition 1 (upper bound). Our assumptions that (i) $\mu_{m',1} < \mu_{m,1}$ and (ii) $0 < \rho < 1$ imply that Proposition 4 condition 1 holds only for signals lower than the agent's prior about the returns to major m. In other words, the agent will only switch majors if they receive a bad signal (one that implies a return lower than their prior).

The range of θ which satisfies Proposition 4 may be degenerate, with a lower bound that is larger than the upper bound, making major switching suboptimal for any grade signal. This occurs if $\mu_{m',1} < \rho\mu_{m,1} + v_o(1-\rho)$.[48] However, for fixed values of σ_m, ρ, and $\mu_{m,1}$, there will always be a value of $\mu_{m',1}$ sufficiently close to $\mu_{m,1}$ that results in a non-degenerate range of θ values for which major switching is optimal. Since major switching is empirically common, we assume a model parametrization in which major switching will be optimal with a non-zero probability.

Notice that both conditions in Proposition 3 are upper bounds on the values of θ for which dropping out is optimal. However, if we assume major switching is ever optimal (*i.e.*, there exists a grade signal which induces the agent to switch majors), then Proposition 3 condition 1 will never bind, leading to the following corollary:

Corollary 2. *If there exists a grade signal θ such that switching majors is optimal after observing θ, then an agent will drop out if and only if* $\theta < \mu_{m,1} + \frac{1}{\rho}\left(1 + \frac{\sigma_\theta^2}{\sigma_m^2}\right)[v_o - \mu_{m',1}]$.

Corollary 2 and Proposition 4 imply that an agent for whom major switching is potentially optimal will have two relevant θ thresholds that characterize her policy function. A lower threshold between dropping out and switching majors and a higher threshold between switching majors and remaining in the initial major.

C.2 Expectations over Future Outcomes

To align the model with our beliefs data, it is necessary for the model to characterize how agents' expectations about future academic decisions evolve over time. This requires extending the model to account for uncertainty after grade signals are observed. To parsimoniously achieve this, we split model period 1 into two subperiods: 1a and 1b. In period 1a, the agent observes a grade signal θ and updates their

[48] This inequality can be derived by setting the right-hand side of Proposition 4 condition 1 less than the right-hand side of Proposition 4 condition 2 and simplifying.

expected returns as previously modeled. In period 1b, they experience a new random shock to their expected returns to re-enrollment. Only after observing this second shock do they decide whether to remain in major m, switch to m', or drop out. This second shock may represent a subsequent grade signal, information from peers, or other factors that influence the opportunity cost of switching majors or dropping out, and which are unknown to the agent until the end of period 1.

By the end of period 1a, the agent will have updated their expectations as outlined in earlier sections. However, they cannot yet determine their optimal re-enrollment decision with certainty, as the period 1b shock remains unobserved. This framework aligns more closely with the timing of our survey data, where respondents completed the second-round survey midway through their second semester. By this time, they had observed their first-semester grades but had not yet decided on actions for the following academic year, and so reported probabilistic re-enrollment beliefs.

Let δ be a mean 0, standard deviation σ_δ normal random variable independent of ν_m, $\nu_{m'}$, and ϵ. Define the returns to major $k \in \{m, m'\}$ as $\nu_k^* = \nu_k + \delta$, where ν_k is as previously defined, and δ represents a separate component of an agent's returns, unrelated to ν_m. The shock δ is observed by the agent in period 1b, prior to the agent making their re-enrollment decision. Because $\mathbb{E}\left[\nu_k^*\right] = \mathbb{E}\left[\nu_k\right]$, belief updating over ν_k^* mirrors the process previously described for ν_k.

Under this extended model, we can describe how the agent's probabilistic posterior beliefs, which are observable in our survey data, respond to grade signals. Unlike posterior beliefs, the grade signals, θ, are not directly unobservable in our data. Instead, we observe realized first-semester grades, which we will denote as ϕ. Within the model, θ is a noisy signal of ν_m, which is measured in utility units. This may be monotonically related to ϕ, but is not actually a grade point average. To overcome this, we assume that θ is some function of realized grades, *i.e.* there exists some function g such that $\theta = g(\phi)$.

Proposition 5. *At the end of period 1a in the extended model with non-degenerate second-best major m', the agent's subjective probability of dropout changes with first-semester GPA, ϕ, according to the following relationship:*

$$\frac{\partial^- \mathbb{P}\left(dropout|\theta\right)}{\partial \phi} = -f(v_o - \bar{\mu})\frac{\partial^- \bar{\mu}}{\partial \theta}\frac{\partial^- g(\phi)}{\partial \phi},$$

where $f(\cdot)$ is the pdf of $\delta \sim \mathcal{N}(0, \sigma_\delta^2)$ and $\bar{\mu} = max\left(\mu_{m,2}, \mu_{m',2}\right)$.

Intuitively, $\frac{\partial^- \bar{\mu}}{\partial \theta}$ is directly related to the Bayesian weights agents use when updating their priors. Figure C1 plots $\bar{\mu}$ against θ for first- and continuing-generation agents to help visualize this relationship. In particular, if θ is above the threshold θ_k^2 for $k \in \{f, c\}$ then $\frac{\partial^- \bar{\mu}}{\partial \theta}|_\theta = \frac{\sigma_{m,k}^2}{\sigma_\theta^2 + \sigma_{m,k}^2}$. If θ is below the threshold θ_k^2 then $\frac{\partial^- \bar{\mu}}{\partial \theta}|_\theta = \rho_k \frac{\sigma_{m,k}^2}{\sigma_\theta^2 + \sigma_{m,k}^2}$. Note that $\bar{\mu}$ is differentiable everywhere except at the kink point θ_k^2, which is why we use left derivatives in our formulation of Proposition 5.

C.2.1 Deriving Testable Implication

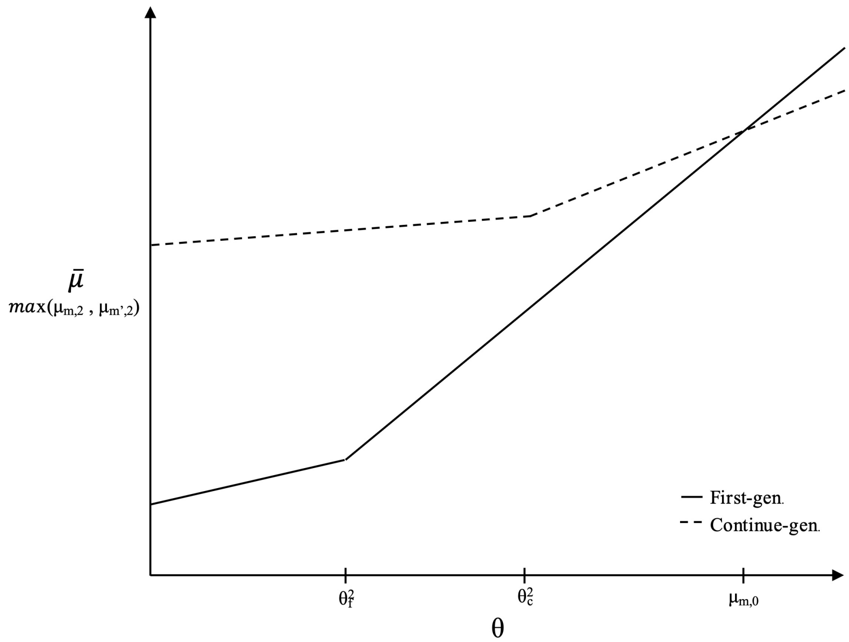
Under some assumptions, Proposition 5 provides the testable implication that first-generation students have larger marginal effects of first-semester grades on their posterior dropout expectations.

In particular, we must assume (i) first- and continuing-generation students have identical grade distributions then the fact that for any grade signal θ $\frac{\partial^- \bar{\mu}}{\partial \theta}|_\theta$ is larger for first-generation agents implies $\mathbb{E}\left[\frac{\partial^- \bar{\mu}}{\partial \theta}|_\theta\right]$ will be larger for first-generation students. In practice, we empirically impose assumption (i) in our analysis using sample weights to balance the distribution of θ across first- and continuing generation samples. We must futher assume (ii) that first- and continuing-generation students share the same $g(\cdot)$ function, then $\frac{\partial^- g(\phi)}{\partial \phi}$ will be the same for all students under assumption (i).

These two assumptions imply $\mathbb{E}\left[\frac{\partial \bar{\mu}}{\partial \theta}\frac{\partial g(\phi)}{\partial \phi}\right]$ is greater for first-generation students. However, the $f(v_o - \bar{\mu})$ term in Proposition 5 complicates our prediction about the relative size of $\frac{\partial^- \mathbb{P}(dropout|\theta)}{\partial \phi}$ between first- and continuing-generation students. Nothing we've assumed so far pins down the relative size of this object, even conditional on θ. However, note that $\frac{\partial \mathbb{P}(dropout|\theta)}{\partial v_o} = f(v_o - \bar{\mu})$. Therefore, the marginal effect of the value of an agent's outside option on their posterior dropout expectations provides a direct estimate of the value in question. Therefore, if there is no difference in the marginal effect of the value of an agent's outside option on their posterior dropout expectations, then we can can confirm there is no difference in $f(v_o - \bar{\mu})$. This leads us to the final assumption: (iii) first- and continuing-generation students have the same marginal effects of the net utility of continued enrollment on their posteriors.

If we ensure assumptions (i), (ii), and (iii) hold in our data, then the model predicts first-generation students have larger marginal effects of first-semester grades on their posterior dropout expectations

Figure C1: Posterior Beliefs and Grade Signals



Notes: Figure plots the largest posterior belief $\bar{\mu} = max\,(\mu_{m,2}, \mu_{m',2})$ against signal θ separately for first- and continuing-generation agents.

C.3 Proofs

Proposition 1

Continuing-generation agents have a lower drop out signal threshold compared to first-generation agents: $\theta_c^1 < \theta_f^1$.

Proof.

$$\begin{aligned}
\theta_c^1 &= \mu_{m,1} + \frac{\sigma_\theta^2 + \sigma_{m.c}^2}{\rho_c \sigma_{m,c}^2} \left[v_o - \mu_{m',1} \right] \\
&= \mu_{m,1} + \frac{\sigma_\theta^2}{\rho_c \sigma_{m,c}^2} \left[v_o - \mu_{m',1} \right] + \frac{\sigma_{m,c}^2}{\rho_c \sigma_{m,c}^2} \left[v_o - \mu_{m',1} \right] \\
v_o < \mu_{m',1} \text{ and (A2)} \implies \quad &< \mu_{m,1} + \frac{\sigma_\theta^2}{\rho_c \sigma_{m,c}^2} \left[v_o - \mu_{m',1} \right] + \frac{\sigma_{m,f}^2}{\rho_f \sigma_{m,f}^2} \left[v_o - \mu_{m',1} \right] \\
v_o < \mu_{m',1} \text{ and Cor. (3.2)} \implies \quad &< \mu_{m,1} + \frac{\sigma_\theta^2}{\rho_f \sigma_{m,f}^2} \left[v_o - \mu_{m',1} \right] + \frac{\sigma_{m,f}^2}{\rho_f \sigma_{m,f}^2} \left[v_o - \mu_{m',1} \right] \\
&= \theta_f^1
\end{aligned}$$

□

Proposition 2

Continuing-generation agents have a higher major switching signal threshold compared to first-generation agents: $\theta_c^2 > \theta_f^2$.

Proof.

$$\begin{aligned}
\theta_c^2 &= \mu_{m,1} + (\mu_{m',1} - \mu_{m,1}) \left(\frac{\sigma_{m,c}^2 + \sigma_\theta^2}{\sigma_{m,c}^2 - \rho_c \sigma_{m,c}^2} \right) \\
\mu_{m',1} < \mu_{m,1} \text{ and Cor. (3.1)} \implies \quad &> \mu_{m,1} + (\mu_{m',1} - \mu_{m,1}) \left(\frac{\sigma_{m,c}^2 + \sigma_\theta^2}{\sigma_{m,f}^2 - \rho_f \sigma_{m,f}^2} \right) \\
\mu_{m',1} < \mu_{m,1} \text{ and (A1)} \implies \quad &> \mu_{m,1} + (\mu_{m',1} - \mu_{m,1}) \left(\frac{\sigma_{m,f}^2 + \sigma_\theta^2}{\sigma_{m,f}^2 - \rho_f \sigma_{m,f}^2} \right) \\
&= \theta_f^2
\end{aligned}$$

□

Proposition 3

An agent will drop out if and only if both of the following hold:

1. $\theta < v_o + \frac{\sigma_\theta^2}{\sigma_m^2}\left[v_o - \mu_{m,1}\right]$
2. $\theta < \mu_{m,1} + \frac{1}{\rho}\left(1 + \frac{\sigma_\theta^2}{\sigma_m^2}\right)\left[v_o - \mu_{m',1}\right]$

Proof. First, we prove 1. and 2. dropout. An agent drops out if and only if both $v_o > \mu_{m,2}$ and $v_o > \mu_{m',2}$ hold. We proceed by showing 1. $\implies v_o > \mu_{m,2}$ and 2. $\implies v_o > \mu_{m',2}$.

$$\begin{aligned}
\text{Condition 1.} \implies \quad & \theta < v_o + \frac{\sigma_\theta^2}{\sigma_m^2}\left[v_o - \mu_{m,1}\right] \\
& \theta < v_o + \frac{\sigma_\theta^2}{\sigma_m^2} v_o - \frac{\sigma_\theta^2}{\sigma_m^2}\mu_{m,1} \\
& \theta + \frac{\sigma_\theta^2}{\sigma_m^2}\mu_{m,1} < v_o\left(\frac{\sigma_m^2 + \sigma_\theta^2}{\sigma_m^2}\right) \\
& \left(\frac{\sigma_m^2}{\sigma_m^2 + \sigma_\theta^2}\right)\theta + \left(\frac{\sigma_\theta^2}{\sigma_m^2 + \sigma_\theta^2}\right)\mu_{m,1} < v_o \\
2 \implies \quad & \mu_{m,2} < v_o
\end{aligned}$$

$$\begin{aligned}
\text{Condition 2.} \implies \quad & \theta < \mu_{m,1} + \frac{1}{\rho}\left(1 + \frac{\sigma_\theta^2}{\sigma_m^2}\right)\left[v_o - \mu_{m',1}\right] \\
& \theta - \mu_{m,1} < \left(\frac{\sigma_m^2 + \sigma_\theta^2}{\rho\sigma_m^2}\right) v_o - \left(\frac{\sigma_m^2 + \sigma_\theta^2}{\rho\sigma_m^2}\right)\mu_{m',1} \\
& \frac{\rho\sigma_m^2}{\sigma_m^2 + \sigma_\theta^2}\left(\theta - \mu_{m,1}\right) + \mu_{m',1} < v_o \\
3 \implies \quad & \mu_{m',2} < v_o
\end{aligned}$$

Next, we prove ¬ 1. $\implies$ ¬ dropout. By the argument above, Condition 1. $\implies$ $\mu_{m,2} < v_o$. Then, ¬ Condition 1. $\implies$ $\mu_{m,2} \geq v_o$, and thus dropout is not optimal for the agent.

Finally, we prove ¬ 2. $\implies$ ¬ dropout. By the argument above, Condition 2. $\implies$ $\mu_{m',2} < v_o$. Then, ¬ Condition 2. $\implies$ $\mu_{m',2} \geq v_o$, and thus dropout is not optimal for the agent. □

Proposition 4

An agent will switch majors if and only if both of the following hold:

1. $\theta < \mu_{m,1} + (\mu_{m',1} - \mu_{m,1})\left(\frac{1}{1-\rho}\right)\left(1 + \frac{\sigma_\theta^2}{\sigma_m^2}\right)$
2. $\theta \geq \mu_{m,1} + \frac{1}{\rho}\left(1 + \frac{\sigma_\theta^2}{\sigma_m^2}\right)[v_o - \mu_{m',1}]$

Proof. First, we prove 1. and 2. $\implies$ switch major. An agent switches majors if and only if both $\mu_{m',2} > \mu_{m,2}$ and $\mu_{m',2} > v_o$ hold. The proof for Proposition 3 above shows 2. $\implies$ $v_o > \mu_{m',2}$, so are only left to show 1. $\implies$ $v_o > \mu_{m,2}$.

$$\begin{aligned}
\text{Condition 1.} \implies \qquad & \theta < \mu_{m,1} + \left(\mu_{m',1} - \mu_{m,1}\right)\left(\frac{1}{1-\rho}\right)\left(1 + \frac{\sigma_\theta^2}{\sigma_m^2}\right) \\
& \theta - \mu_{m,1} < \left(\mu_{m',1} - \mu_{m,1}\right)\left(\frac{1}{1-\rho}\right)\left(1 + \frac{\sigma_\theta^2}{\sigma_m^2}\right) \\
& (\theta - \mu_{m,1}) - \rho\,(\theta - \mu_{m,1}) < \left(\mu_{m',1} - \mu_{m,1}\right)\left(\frac{\sigma_m^2 + \sigma_\theta^2}{\sigma_m^2}\right) \\
& \left(\frac{\sigma_m^2}{\sigma_m^2 + \sigma_\theta^2}\right)(\theta - \mu_{m,1}) - \left(\frac{\rho\sigma_m^2}{\sigma_m^2 + \sigma_\theta^2}\right)(\theta - \mu_{m,1}) < \left(\mu_{m',1} - \mu_{m,1}\right) \\
& \left(\frac{\sigma_m^2}{\sigma_m^2 + \sigma_\theta^2}\right)\theta - \left(\frac{\sigma_m^2}{\sigma_m^2 + \sigma_\theta^2}\right)\mu_{m,1} + \mu_{m,1} < \mu_{m',1} + \left(\frac{\rho\sigma_m^2}{\sigma_m^2 + \sigma_\theta^2}\right)(\theta - \mu_{m,1}) \\
& \left(\frac{\sigma_m^2}{\sigma_m^2 + \sigma_\theta^2}\right)\theta + \left(\frac{\sigma_\theta^2}{\sigma_m^2 + \sigma_\theta^2}\right)\mu_{m,1} < \mu_{m',1} + \left(\frac{\rho\sigma_m^2}{\sigma_m^2 + \sigma_\theta^2}\right)(\theta - \mu_{m,1}) \\
\text{2 and 3} \implies \qquad & \mu_{m,2} < \mu_{m',2}
\end{aligned}$$

□

Next, we prove $\neg$ 1. $\implies$ $\neg$ switch majors. By the argument in , Condition 2. $\implies$ $\mu_{m',2} > \mu_{m,2}$. Then, $\neg$ Condition 2. $\implies$ $\mu_{m,2} \geq \mu_{m',2}$, and thus switching majors is not optimal for the agent.

Finally, we prove $\neg$ 2. $\implies$ $\neg$ switch. By the argument in Proposition 3, Condition 2. $\implies$ $\mu_{m',2} > v_o$. Then, $\neg$ Condition 1. $\implies$ $v_o \geq \mu_{m',2}$, and thus switching majors is not optimal for the agent.

Proposition 5

At the end of period 1a in the extended model with non-degenerate second-best major m', the agent's subjective probability of dropout changes with first-semester GPA, ϕ, according to the following relationship:

$$\frac{\partial^{-}\mathbb{P}\left(dropout|\theta\right)}{\partial\phi} = -f(v_o - \bar{\mu})\frac{\partial^{-}\bar{\mu}}{\partial\theta}\frac{\partial^{-}g(\phi)}{\partial\phi}$$

where $f(\cdot)$ is the pdf of $\delta \sim \mathcal{N}(0, \sigma_{\delta}^2)$ and $\bar{\mu} = max\left(\mu_{m,2}, \mu_{m',2}\right)$.

Proof. At the end of period 1a, before observing δ, the agent believes they will drop out with probability $\mathbb{P}\left(dropout|\theta\right) = \mathbb{P}\left(\delta < v_o - \bar{\mu}\right) = F\left(v_o - \bar{\mu}\right)$ where $F(\cdot)$ is the CDF of $\mathcal{N}(0, \sigma_{\delta}^2)$

$$\begin{aligned}
&\frac{\partial^{-}\mathbb{P}\left(dropout|\theta\right)}{\partial\phi} \\
=&\frac{\partial^{-}F\left(v_o - \bar{\mu}\right)}{\partial\phi} \\
=&\frac{\partial^{-}}{\partial\phi}\int_{-\infty}^{v_o-\bar{\mu}} f(x)\,dx \\
=&f(v_o - \bar{\mu})\frac{\partial^{-}}{\partial\phi}\left[v_o - \bar{\mu}\right] \\
=&-f(v_o - \bar{\mu})\frac{\partial^{-}}{\partial\phi}\bar{\mu} \\
=&-f(v_o - \bar{\mu})\frac{\partial^{-}\bar{\mu}}{\partial\theta}\frac{\partial^{-}\theta}{\partial\phi}
\end{aligned}$$

□

Corollary 1

Upon receiving identical grade signals, a first-generation student will update their beliefs weakly more than a continuing-generation student. *I.e.* $|\mu_{m,2,f} - \mu_{m,1}| \geq |\mu_{m,2,c} - \mu_{m,1}|$. If the signal is non-degenerate ($\theta \neq \mu_{m,1}$), this will strongly hold.

Proof.

$$
\begin{aligned}
(2) \implies \qquad |\mu_{m,2,f} - \mu_{m,1}| &= \left| \frac{\sigma_\theta^2}{\sigma_{m,f}^2 + \sigma_\theta^2} \mu_{m,0} + \frac{\sigma_{m,f}^2}{\sigma_{m,f}^2 + \sigma_\theta^2} \theta - \mu_{m,1} \right| \\
&= \left| \frac{\sigma_{m,f}^2}{\sigma_{m,f}^2 + \sigma_\theta^2} \theta - \frac{\sigma_{m,f}^2}{\sigma_{m,f}^2 + \sigma_\theta^2} \mu_{m,0} \right| \\
&= \frac{\sigma_{m,f}^2}{\sigma_{m,f}^2 + \sigma_\theta^2} \left| \theta - \mu_{m,0} \right| \\
(A2) \implies \qquad &\geq \frac{\sigma_{m,c}^2}{\sigma_{m,c}^2 + \sigma_\theta^2} \left| \theta - \mu_{m,0} \right| \\
&= |\mu_{m,2,c} - \mu_{m,1}|
\end{aligned}
$$

□

Corollary 2

If there exists a grade signal θ such that switching majors will be optimal after observing this signal, then an agent will drop out if and only if $\theta < \mu_{m,1} + \frac{1}{\rho}\left(1 + \frac{\sigma_\theta^2}{\sigma_m^2}\right)[v_o - \mu_{m',1}]$

Proof. First, note that the existence of a grade signal θ that induces major switching implies that the right-hand side of Proposition 4 condition 1 is greater than the right-hand side of condition 2. Simplifying this inequality yields the following: $\mu_{m',1} > \rho\mu_{m,1} + v_o(1 - \rho)$.

Next, let T_1 be the upper bound on θ implied by Proposition 3 condition 1 and T_2 be the upper bound on θ implied by Proposition 3 condition 2. We will proceed to show that the inequality derived above implies $T_1 > T_2$, and thus $\theta < T_2 \implies \theta < T_1$. Then, the corollary follows directly from Proposition 3.

$$
\begin{aligned}
T_2 - T_1 &= \left[\mu_{m,1} + \frac{1}{\rho}\left(1 + \frac{\sigma_\theta^2}{\sigma_m^2}\right)[v_o - \mu_{m',1}]\right] - \left[v_o + \frac{\sigma_\theta^2}{\sigma_m^2}[v_o - \mu_{m,1}]\right] \\
&= \left[\mu_{m,1} + \frac{1}{\rho}\left(\frac{\sigma_m^2 + \sigma_\theta^2}{\sigma_m^2}\right) v_o - \frac{1}{\rho}\left(\frac{\sigma_m^2 + \sigma_\theta^2}{\sigma_m^2}\right)\mu_{m',1}\right] \\
&- \left[\left(\frac{\sigma_m^2 + \sigma_\theta^2}{\sigma_m^2}\right) v_o - \frac{\sigma_\theta^2}{\sigma_m^2}\mu_{m,1}\right] \\
&= \left(\frac{\sigma_m^2 + \sigma_\theta^2}{\sigma_m^2}\right)\mu_{m,1} + \frac{1-\rho}{\rho}\left(\frac{\sigma_m^2 + \sigma_\theta^2}{\sigma_m^2}\right) v_o - \frac{1}{\rho}\left(\frac{\sigma_m^2 + \sigma_\theta^2}{\sigma_m^2}\right)\mu_{m',1} \\
&= \left(\frac{\sigma_m^2 + \sigma_\theta^2}{\sigma_m^2}\right)\left[\mu_{m,1} + \frac{1-\rho}{\rho} v_o - \frac{1}{\rho}\mu_{m',1}\right] \\
&= \left(\frac{\sigma_m^2 + \sigma_\theta^2}{\sigma_m^2}\right)\frac{1}{\rho}\left[\rho\mu_{m,1} + (1-\rho) v_o - \mu_{m',1}\right] \\
\mu_{m',1} > \rho\mu_{m,1} + v_o(1-\rho) \implies &< \left(\frac{\sigma_m^2 + \sigma_\theta^2}{\sigma_m^2}\right)\left[\mu_{m',1} - \mu_{m',1}\right] \\
&= 0
\end{aligned}
$$

□

Corollary 3. *Assumptions (A1) and (A2) imply:*

1. $\rho_f \sigma_{m,f}^2 > \rho_c \sigma_{m,c}^2$
2. $\sigma_{m,f}^2 - \rho_f \sigma_{m,f}^2 < \sigma_{m,c}^2 - \rho_c \sigma_{m,c}^2$

Corollary 3.1 $\rho_f \sigma_{m,f}^2 > \rho_c \sigma_{m,c}^2$

Proof. Follows directly from assumptions A1 and A2 and $0 < \rho_k < 1$ for $k \in \{f, c\}$ □

Corollary 3.2 $\sigma_{m,f}^2 - \rho_f \sigma_{m,f}^2 < \sigma_{m,c}^2 - \rho_c \sigma_{m,c}^2$

Proof.

$$
\begin{aligned}
(\text{A2}) \implies & \quad \frac{\rho_f \sigma_{m,f}^2}{\sigma_{m,f}^2} > \frac{\rho_c \sigma_{m,c}^2}{\sigma_{m,c}^2} \\
\implies & \quad \frac{\rho_f \sigma_{m,f}^2 - \sigma_{m,f}^2}{\sigma_{m,f}^2} > \frac{\rho_c \sigma_{m,c}^2 - \sigma_{m,c}^2}{\sigma_{m,c}^2} \\
(\text{A1}) \implies & \quad \rho_f \sigma_{m,f}^2 - \sigma_{m,f}^2 > \rho_c \sigma_{m,c}^2 - \sigma_{m,c}^2 \\
\implies & \quad \sigma_{m,f}^2 - \rho_f \sigma_{m,f}^2 < \sigma_{m,c}^2 - \rho_c \sigma_{m,c}^2
\end{aligned}
$$

□

D Appendix: LEAD Program

D.1 LEAD Program: Details

The LEAD program offers several distinctive features that make it different from the regular academic experience at ASU:

- Course content: Many classes are designed to help students navigate their first year of college and develop critical thinking/writing, creative problem-solving, and soft skills. A growing literature documents the increasing importance of multiple skills (i.e., soft/non-cognitive, cognitive skills) in success (for example, see Heckman and Kautz, 2012).

- Small class sizes: While regular introductory classes at ASU have, on average, 250 students enrolled, LEAD courses only have 30-40 students. Class size has been shown to be an important determinant of student academic performance (for example, see Bettinger and Long, 2018).

- Academically homogeneous classrooms with the same students across classes: Students enrolled in LEAD have similar academic preparation when starting college, and many share several courses. The boutique model of peer effects (i.e., emphasizing gains from peer group homogeneity) supports this type of classroom composition (for example, Lu and Anderson, 2015).

- Academic success coordinators and mentors: LEAD has been designed so students receive frequent advice from LEAD mentors (students who have already completed the program) and academic success coordinators (paid full-time staff). Mentoring programs have been successful in higher education (for example, Carrell and Sacerdote, 2017).

- Concerned Professors: LEAD faculty members have experience teaching LEAD classes and interacting with students who may face more academic challenges than the typical student at ASU.

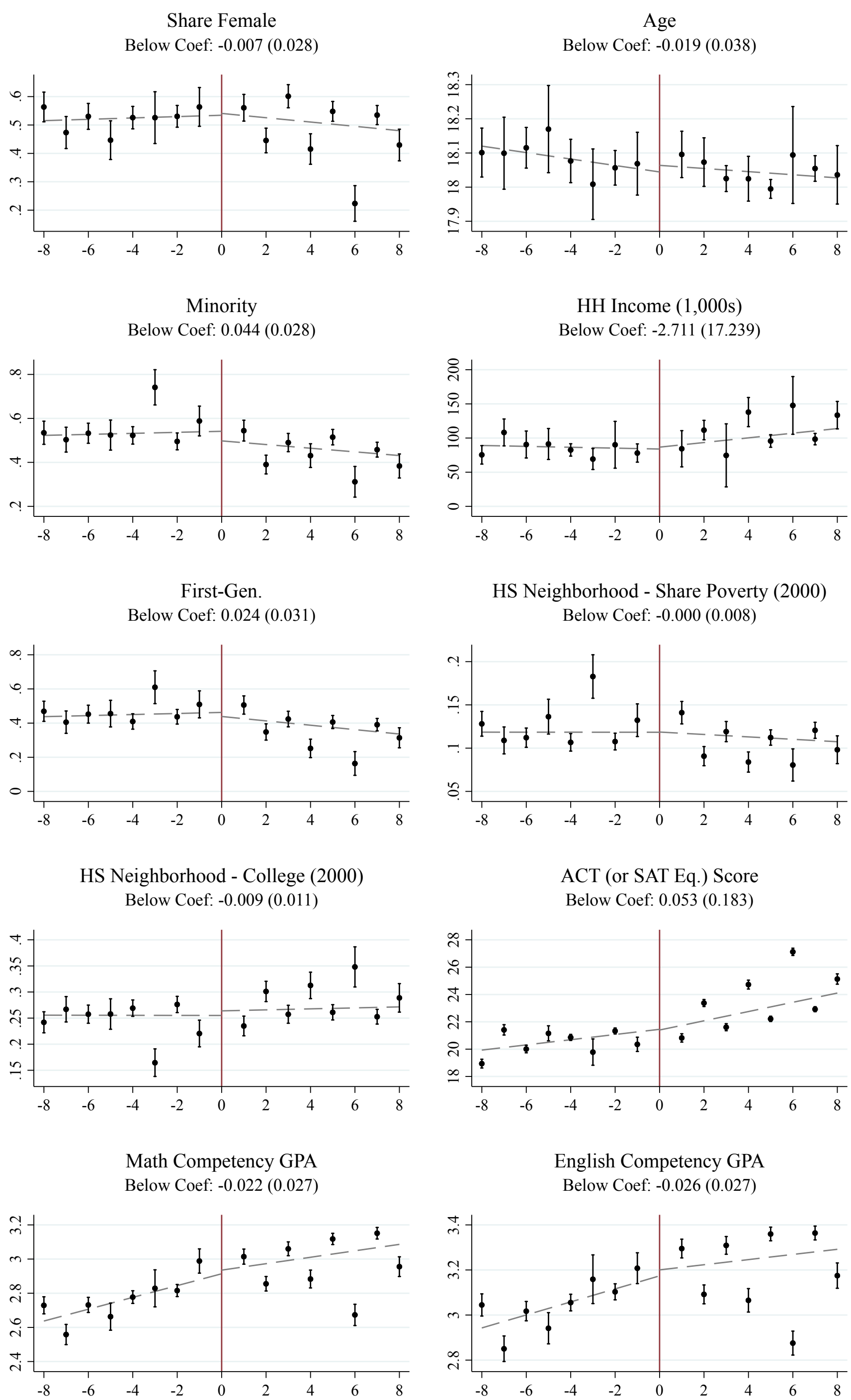
D.2 LEAD Program: RDD Identification Strategy

Student transcript data confirms the presence of a discontinuity in LEAD enrollment, indicating a strong first-stage for the fuzzy RD estimates. More specifically, Figure A10 shows that students below the cutoff value are (statistically and economically) significantly more likely to enroll in the program. While strongly suggestive that ASU's academic counselors were following a policy of nudging students below the threshold to enroll in LEAD, a discontinuity in LEAD enrollment is not sufficient for a causal interpretation of the fuzzy RD estimates. The key identification assumption in our setting is that students immediately above and below the cutoff only differ in their propensity to enroll in LEAD (and the subsequent effects of this enrollment). This assumption could fail for at least two reasons. First, the CI score may be so coarse that, even near the cutoff, students above and below could differ in some observable or unobservable way. Second, CI scores near the cutoff may be correlated with college experiences in ways other than LEAD.

While the key identification assumption is not directly testable, we find no evidence that pre-enrollment observables are discontinuous at the cutoff score, suggesting that students near the cutoff are likely very similar. In particular, Appendix Figure D1 plots binscatters of several pre-enrollment observables across the CI score cutoff and finds no evidence of a discontinuity in any of these factors.[49]

[49]Figure D1 does not rule out the possibility that some unobservable factor related to academic outcomes differs at the cutoff score. As discussed in McCrary (2008), one common way RD designs may fail is if the experimental units are able to manipulate the running variable (CI score in our context) and suggests that a discontinuity in the sample density at the cutoff is indicative of manipulation. This test is not particularly useful in our setting because CI scores, by construction, are not smoothly distributed (see appendix Figure D2 panel C for the overall distribution of CI scores). Further, as an aggregation of standardized test scores and high school grades, there is little scope or incentive for either the university or the student to manipulate scores specifically at the cutoff. With these caveats in mind, Appendix Figure D2 panels A and B plot the distribution of CI scores at the cutoff along with a placebo cutoff 10 points above the actual cutoff. We see the same distributional pattern in both panels.

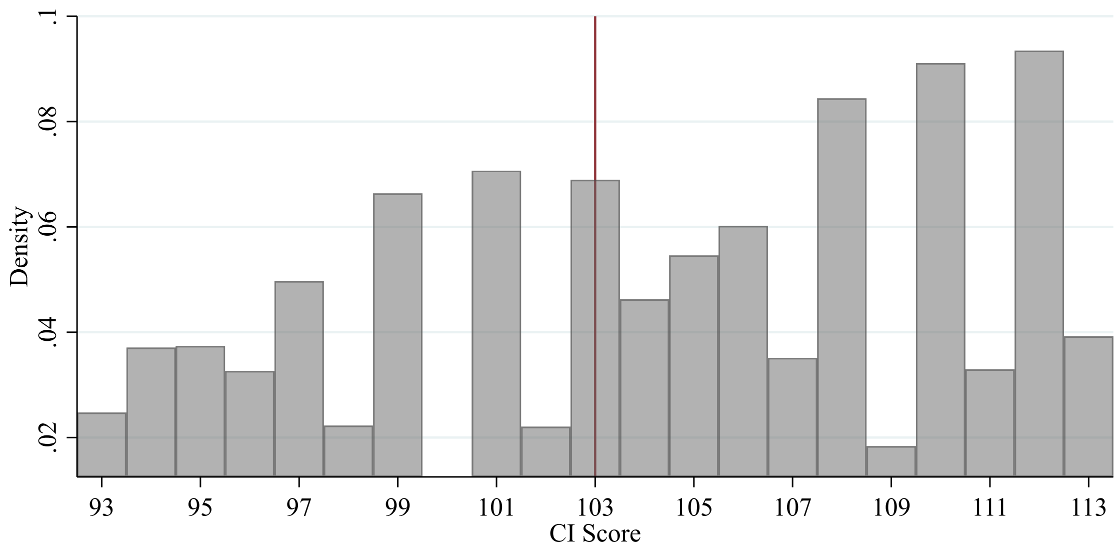
Figure D1: Smoothness of Covariates at LEAD CI Score Cutoff



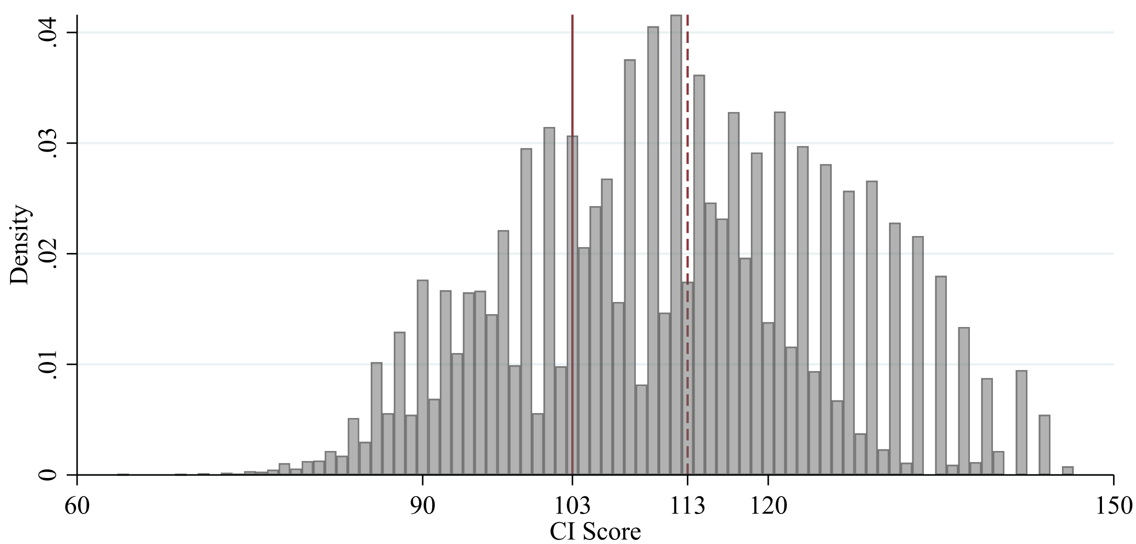
Notes: Figure plots the average covariate values for students in the 2016-2018 cohorts within +/- 8 CI score points from the LEAD cutoff score. Dashed lines represent separate linear trends above/below the cutoff. Coefficients (along with standard errors) for the magnitude of the discontinuity in linear trends are presented below each covariate label. Household income (HH Income) and First-generation status (First-Gen.) panels exclude observations missing FAFSA information. Neighborhood characteristics for high school census tract (HS Neighborhood) measured in the year 2000 and include only students from Arizona high schools. Competency GPAs are composite scores based on qualifying high school courses. A small number of observations (fewer than 300) are missing standardized test scores or one or more competency.

Figure D2: CI Score Distribution

(a) Distribution Near CI Score Cutoff



(b) Distribution of CI Scores



Notes: Figure plots the distribution of CI scores for the 2016-2018 cohorts. The LEAD nudge cutoff (103) is highlighted by a solid red line in both panels. The average score is highlighted by a dashed line in panel (b).