The Effects of College Desegregation on Academic Achievement and Students' Social Interactions: Evidence from Turnstile Data

Tatiana Velasco*

April 29, 2024

Most recent version here

Abstract

How does socioeconomic desegregation of elite schools impact students' cross-SES interactions within their school? Does such desegregation affect academic achievement? In this paper, I study a natural experiment at an elite university in Colombia where the number of low-SES students tripled with the introduction of the financial aid program *Ser Pilo Paga*. I develop a measure of social interactions using data on students' comovements across campus captured by turnstiles located at all entrances. Increased exposure to low-SES students increased the diversity of social interactions, with no adverse effects on the academic achievement of the high-SES students traditionally attending the institution, although a bias for same-SES interactions persisted. Forty percent of the increase in interactions between low- and high-SES students is explained by high-SES students befriending very high-achieving low-SES students at higher rates than before the policy.

JEL Codes: I22, I23, I24

*Teachers College, Columbia University. 525 West, 120th Street. New York, NY. 10027. E-mail: tv2225@tc.columbia.edu. I am grateful to my dissertation advisors Peter Bergman and Jordan Matsudaira for their guidance and support throughout this research. I also thank Sarah Cohodes, Alex Eble, Judith Scott-Clayton, Tomas Rodriguez-Barraquer and Roman Andres Zarate for their feedback and comments. This paper has benefited from the many comments of seminar participants at the Association for Public Policy Analysis and Management Conference, Northeastern Universities Development Consortium, Southern Economic Association, Economics and Education Research Seminar at Columbia University, and Syracuse–Cornell Education and Social Policy Daniel Patrick Moynahan Conference. I am grateful to the National Academy of Education/Spencer Foundation Dissertation Fellowship for providing funding for this project. All errors are my own.

1 Introduction

Segregation of students by socioeconomic status, race, or ethnicity is a pervasive issue in education. At the postsecondary level, policymakers have implemented financial aid and affirmative action programs that foster access to selective institutions for low-income and underrepresented groups. These policies are the subject of considerable debate, with affirmative action rules for college admission recently overturned in the U.S. and discussions about the relevance and ramifications of fostering a diverse college environment continuously proliferating.[1]

Understanding the consequences of policies aiming to foster the access of low-income and potentially lower-achieving students to elite universities is critical. On one hand, these policies may exacerbate achievement gaps within institutions, particularly if the benefited students struggle to perform as well as their classmates. This could lead to potentially negative peer effects on the performance of students from incumbent demographic groups at these institutions (Arcidiacono, Lovenheim and Zhu, 2015). On the other hand, desegregation could diversify social interactions, which is a desirable outcome, especially if we account for the positive impacts that exposure to diversity has on privileged students (Rao, 2019; Boisjoly et al., 2006; Londoño-Vélez, 2022; Corno, La Ferrara and Burns, 2022). In this paper, I examine what the consequences of college desegregation are for academic achievement and whether desegregation can diversify students' social interactions.

To answer these questions, I use a natural experiment at a large elite college in Colombia that experienced a sharp and unexpected increase in the enrollment of students of low socioeconomic status (SES) after the introduction of a nationwide financial aid program known as *Ser Pilo Paga* (SPP, "Being Smart Pays"). To measure social interactions, I assemble a novel database of over a hundred million records of student movements across

[1] In the summer of 2023, the U.S. Supreme Court overturned *Grutter v. Bollinger*, a landmark decision that allowed colleges to consider race in their admissions for affirmative action purposes.

campus collected by the turnstiles guarding all entrances. I develop a measure to identify which students socialize with one another based on how commonly I observed them entering and exiting campus buildings together, and I validate it against a survey where students listed their friends and acquaintances. I combine these data with student-level records on course enrollment and academic achievement and persistence. I find that, although some bias for same-SES interactions persisted among students, the socioeconomic desegregation significantly increased interactions between students of different socioeconomic backgrounds, with no adverse effects on the achievement of the relatively better-off students traditionally attending this elite university. The increase in diverse interactions is partially driven by students befriending very high-achieving low-SES students at higher rates than other low-SES students.

In October 2014, the Colombian government launched SPP, a policy targeting students of the lowest SES with outstanding academic achievement to promote their attendance of high-quality universities in the country. The program consisted of a loan covering 100 percent of tuition, to be forgiven upon degree completion, plus a small stipend for living expenses. SPP induced an influx of low-SES students into high-quality private universities, effectively closing the socioeconomic enrollment gap among high achievers (Londoño-Velez, Rodriguez and Sanchez, 2020). The rollout of the program was fairly quick, with the first cohort of students benefiting from SPP enrolling in January 2015, barely three months after its announcement. This timely implementation meant that recipient universities and the high-SES students who traditionally attend these institutions had little to no time to adjust their admission and application strategies to accommodate their preferences for peers of certain backgrounds. Critical to this fact is that, in Colombia, college applications are submitted to a college and major bundle (*program*), with admitted students enrolling directly into their program of application. In this context, as a result of SPP, the number of students admitted to the university that I study sharply increased, with the number of lower-SES students tripling while the number and characteristics of students from other groups remained unchanged.

To estimate the effects of this episode of college desegregation, I leverage the plausibly random variation in the amount of exposure to low-SES students within each program and across entry cohorts. Specifically, I implement a standard difference-in-differences (DID) approach that compares high-SES students from different programs enrolling right before and after the SPP rollout (the 2014 vs. spring 2015 cohorts). A key identification assumption is that high-SES students' program choices did not change in response to the changes in student body composition in terms of the share of low-SES peers. To support this assumption, I conduct multiple pretrend and placebo tests and find no evidence of changes in the characteristics of high-SES students enrolling before and after the SPP rollout or in response to changes in the socioeconomic composition of their peer group. As a robustness check, and to avoid estimating effects involving negative-dosage treatments, I present results using discrete treatment definitions for when the share of low-SES students is above the maximum share observed in previous cohorts.

I start by analyzing the changes in academic achievement among elite university students with the introduction of SPP. The presence of more low-SES students increased the achievement gap of this group relative to their high-SES peers, driven mainly by a drop in the performance of the low-SES students enrolling in the SPP cohort. Before SPP, the cumulative GPA and credits attempted by the third term for both high- and low-SES students hovered around 3.8 points and 50 credits, respectively. In the SPP entry cohort, however, low-SES students' GPA dropped to 3.65 points and the credits attempted to 46 credits, while there were no changes among high-SES students. The increased exposure to these somewhat underachieving students did not affect the outcomes of high-SES students. I show this by using a DID design to estimate the causal effect of increased exposure to low-SES students on high-SES students' achievement, finding no effect across GPA, credits attempted, dropout or degree completion probability.

A potential explanation for the lack of effects on the achievement of high-SES students is a lack of interaction between high- and low-SES students. Such a result would be consistent with the finding in prior literature that the absence of peer effects on performance is explained by a lack of social interaction between high- and low-achieving students that leaves high achievers' performance unaffected while negatively impacting the performance of low achievers (Carrell, Sacerdote and West, 2013). Students of different socioeconomic status may not interact much with each other because of a preference for the company of others of similar backgrounds (a phenomenon known as homophily). The main part of my empirical analysis examines how social interactions between low- and high-SES students changed with the increased presence of low-SES students in the university.

I measure social interactions using data on students' comovements across the university captured by turnstiles located at the 18 campus entrances. These data are available for the period from fall 2016 to spring 2019, and I use them to measure comovements among students in the same program and entry cohorts after six and seven semesters of enrollment. I define a pair of students as linked if they passed through the turnstiles in the same direction (entering or exiting a building) within a time window of five seconds or less at least three times in a term. I describe the process by which I validated and arrived at this definition and address potential measurement error limitations. Descriptive statistics show that high-SES students had an average of five links in their program and cohort before the implementation of SPP, with an average of 0.23 links being with a low-SES peer. On the other hand, low-SES students had on average 4.75 links, and 0.32 of those were with other low-SES peers. With the onset of SPP, high-SES students increased their total links to 5.6 and had four times more links with low-SES peers, whereas low-SES students increased their total links to five but had over six times more links with other low-SES peers. I next proceed to estimate how these changes in exposure to low-SES peers affected the number of interactions across socioeconomic groups.

Interactions between low- and high-SES students significantly increased with the onset of SPP, with high-SES students substituting links among themselves with links with low-SES peers, particularly among the groups with the largest shares of exposure. I estimate an increase in the probability of a link between the two groups of 14 percentage points at the average percent of low-SES peers and a 0.70 increase in the number of low-SES links. In programs with shares of low-SES peers above the median of that in the pre-SPP period, the number of links with low-SES peers increased by one, while the links among high-SES students declined by 1.2, suggesting some link substitution between high- and low-SES links in the programs with the largest shares of low-SES peers. Further analyses suggest that approximately a third of these new links between low- and high-SES students are particularly strong and knitted, as the associated interactions tend to occur around the end of the day -suggesting students start their commute together, and are only partially mediated by students taking courses together at that time.

However, I do find evidence of a persisting bias for links within the same SES group, suggesting some lingering segregation among students. First, estimations on the percentage of low-SES links indicate that high-SES students' responses to the changes in peer group composition are not monotonic. Specifically, a one-percentage-point increase in the share of low-SES peers translates to a 0.75-point increase in the percentage of high-SES links with that group. This suggests an unrealized 25 percent reduction in segregation of social interactions that proved resistant to the increased exposure to low-SES peers. To further assess this issue, I compute a measure of friendship bias following Chetty, Jackson, Kuchler, Stroebel, Hendren, Fluegge, Gong, Gonzalez, Grondin, Jacob et al. (2022). My results suggest that, while the variation in friendship bias was reduced, many of the programs among the 2015 cohort did exhibit some bias in favor of own-SES links. I find no evidence that this bias correlates with the percentage of low-SES students in the group.

The increased interactions between high- and low-SES students may be explained by students' preferences for interacting with high-achieving peers being stronger than those for interacting with same-SES peers. To test this, I identify low-SES students with an academic performance equal to or above the average performance of high-SES students and estimate the effect of exposure to low-SES peers on the number of interactions with these low-SES but very high-achieving students. Forty percent of the increase in interactions between low- and high-SES students is explained by interactions with very high-achieving low-SES peers, with high-SES students befriending high-achieving low-SES peers at much higher rates than they do other low-SES students in general.

This study makes four contributions to the literature. First, my paper contributes to the research examining the consequences of college diversity and desegregation for academic achievement, which overall finds mixed results. This literature has exploited quasi-experimental variation in the implementation of affirmative action rulings Bleemer (2021a) or in college cohorts Arcidiacono and Vigdor (2010) to examine the effects of exposure to minority students on white and Asian students' performance, finding null or negative effects. [2] Other researchers exploit random group (or dorm) assignment to examine the effects of interracial exposure on student achievement, finding positive effects on White and Black students (Lau, 2022; Corno, La Ferrara and Burns, 2022). My paper exploits quasi-random variation in exposure at the major–cohort level from an affirmative action–type policy to examine the effects on student academic achievement, finding null effects. Notably, this paper measures diversity along the socioeconomic, not racial, dimension, which has been much less explored in this literature.

Second, this paper contributes to the literature examining how students' social interactions change with financial policies fostering desegregation. While prior research has consistently found positive impacts on the college attainment of students benefiting from financial aid and affirmative action programs for underrepresented groups (Bleemer, 2021b; Chetty et al., 2020; Londoño-Velez, Rodriguez and Sanchez, 2020; Mello, 2022), I provide

[2]A couple of studies in K–12 settings examining a similar research question have found null results. Angrist and Lang (2004) study the effect of a desegregation program in Boston on the academic achievement of students from the groups traditionally attending the receiving schools, finding no significant impact; a similar study by Dobbie and Fryer (2014) focuses on students eligible to attend schools with high-achieving peers and finds no impacts on either group's achievement.

novel evidence on how social interactions change under a desegregation policy in light of its lack of impacts on achievement. Findings from Michelman, Price and Zimmerman (2022) and Zimmerman (2019) show that low-income and minority students tend not to take part in privileged students' social clubs even if they share the same college environment, which may explain the somewhat slower or absent social mobility among low-income students attending elite institutions. My measure extends the definition of social interaction by capturing dynamics outside clubs and classrooms, showing that high- and low-SES students do connect at the outset of desegregation, which may have other positive ramifications for the social mobility of low-income students and for prosocial behaviors among wealthy ones (Rao, 2019; Boisjoly et al., 2006; Londoño-Vélez, 2022; Chetty, Jackson, Kuchler, Stroebel, Hendren, Fluegge, Gong, Gonzalez, Grondin, Jacob, Johnston, Koenen, Laguna-Muggenburg, Mudekereza, Rutter, Thor, Townsend, Zhang, Bailey, Barberá, Bhole and Wernerfelt, 2022).[3]

Third, this paper connects to the literature examining diversity in school settings and its effects on segregation in social networks. This research has examined the process by which friendships form in college settings and has relied on proxies of social interaction such as email exchanges (Marmaros and Sacerdote, 2006) and Facebook friendships (Baker, Mayer and Puller, 2011). My study provides a finer measure of effects on social interactions by capturing the effects of desegregation on both the probability of interaction with a low-SES student and the number of low-SES peers whom students from incumbent classes connect with. Similarly, the evidence coincides in indicating that peers' proximity and race are determinants of friendship formation: namely, students assigned to the same dorm are more likely to be connected, but the chances of connecting are higher for same-race students.[4] My study uses a different dimension of proximity, namely, being in the

[3]My work is closely aligned with that of Londoño-Vélez (2022), who studies the effect of socioeconomic diversity at an elite college in Colombia on students' preferences for redistribution. In this work, Londoño-Vélez finds positive impacts of exposure on wealthy students' preferences—a result that seems to be related to an increase in their interactions with low-income peers. My work validates this latter finding while pointing out that the change in social interactions is relatively small.

[4]Marmaros and Sacerdote (2006) examine how people form social networks with their peers. They use

same major and entry cohort. My findings indicate that proximity in major and cohort groups is determinant for student interactions.

My results also connect to those of Chetty, Jackson, Kuchler, Stroebel, Hendren, Fluegge, Gong, Gonzalez, Grondin, Jacob et al. (2022), who show that increases in exposure to high-SES peers across school cohorts lead to more friendships between low- and high-SES people in schools with lower levels of *friendship bias*, a measure that I replicate in my analysis and that allows me to document a persistent bias for friendships with same-SES students at the university I study. A related substream of research has focused on measuring overall segregation in social interaction and on studying how policies can reduce within-group segregation in K–12 settings, finding no association between whom students interact with and academic achievement (Echenique, Fryer and Kaufman, 2006) and nonlinear responses of interactions to scenarios in which minorities are reallocated across schools (Mele, 2020).[5] My findings show consistently positive impacts of exposure to socioeconomic diversity in social interactions at the program level, with no evidence of impact in academic achievement.

Last, this paper contributes to the research examining the role of social networks in academic achievement and peer effects. Carrell, Sacerdote and West (2013) design a peer effects experiment aiming to optimize the assignment of high-achieving students to boost the performance of low-achieving ones. Their results indicate no effects on high achievers' performance and negative effects on that of low achievers, a result that the authors document is driven by segregation in social interactions between the two groups. Complementing Carrell, Sacerdote and West's (2013) results, I find a lack of peer effects on the

student email exchange data and find that first-year students form friendships with students in physical proximity to them and are more likely to form friendships with peers of the same race. Baker, Mayer and Puller (2011) use data from Facebook and random dorm assignment at one college and find that exposure to students of different races in dorms leads to more diverse friendships.

[5] Echenique, Fryer and Kaufman (2006) measure within-school segregation as the extent to which students interact socially with other students of the same race. Mele (2017) develops a structural model of friendship formation among students, and Mele (2020) uses it to simulate reallocation programs across schools and examine its impacts on within-school friendship formation. His findings suggest that policies that reallocate students by parental income have less impact on racial segregation within schools than those that reallocate on the basis of race.

achievement of high-SES students—the large majority of whom are high performers—but a positive effect on interactions with low-SES students—the group with lower achievement on average. My results do suggest high-SES students have some persistent bias for interactions with same-SES students and that they tend to befriend very high-achieving low-SES students at high rates, indicating academic achievement is a strong driver of the increased diversity in social interactions.

2 Background and Setting

I study the case of a large private university located in Bogotá, Colombia (from now on, *Elite University* or EU[6]), which in 2015 experienced a large and unexpected increase in the number of low-income students enrolled while the enrollment of relatively wealthy students remained constant. The increase was driven by Ser Pilo Paga—a forgivable loan program for high-achieving low-income students who wished to attend a high-quality university. Importantly, the increase in low-income students' enrollment varied across the thirty-one programs offered at EU. My research design focuses on relatively wealthy students and compares students from the entry cohorts before and after the SPP rollout (2014 vs. 2015). I use the change in the number of low-SES students across programs and cohorts as the treatment. In this section, I explain the context of SPP and EU, where the natural experiment took place.

Higher education in Colombia is strongly segregated. By 2014, the gap in gross post-secondary enrollment between low-income and wealthy youth was 51 percentage points (Arias Ortiz, Elacqua and Gonzalez-Velosa 2017). Among those enrolled in bachelor's degrees, high-ability low-income students are much less likely to be enrolled at a private university than their wealthy counterparts (Carranza and Ferreyra 2019). This can be explained by the high tuition rates of private universities relative to average salaries

[6] I do not provide the real name of the university I study for confidentiality reasons.

in the country and the limited financial aid options available for low-income students. SPP aimed to address this segregation by providing low-income students a loan that covered tuition plus a small allowance for attending a high-quality accredited institution.[7] The loan was forgiven conditional on degree completion. Eligibility for SPP required students to be classified as poor under the government's index of household wealth and to have scored in the top decile of the national high-school exit exam, SABER 11 (SB11).[8] SPP awarded loans for new cohorts of students between 2015 and 2018, benefiting approximately 40,000 students nationwide. Previous research has found that SPP increased diversity at top private universities by shifting the basis of selection more toward ability instead of income (Londoño-Velez, Rodriguez and Sanchez, 2020). Figure 1 depicts that, of all the institutions eligible for the program, EU had the largest change in the percentage of low-SES students enrolled, with over 500 new low-SES students in the 2015 entry cohort, which tripled the share of this group relative to its 2014 enrollment share.

The timing of the SPP rollout and EU's admission rules set the conditions of my research design. First, admissions to EU are open for each year's spring and fall terms and are determined by the applicant's SB11 score. Students must apply to a major and entry cohort for which admission officers predetermine a specific SB11 weighting formula and cutoff score.[9] Second, SPP was broadly unanticipated among students and higher education institutions. SPP was launched in October 2014, and only students who had taken that October's test were SPP-eligible. Candidates had to apply for enrollment in the following spring (2015), for which 10,000 forgivable loans were offered. Thus, students from the demographic groups that traditionally apply to EU had very little time to change their application portfolio, and university officers could not adjust the admission

[7]High-quality accreditation is granted to higher education institutions by the National Council of Accreditation after a detailed review by a panel formed by the institution, the academic community, and the council. By 2014, the year of the first round of SPP, 32 universities held high-quality accreditation.

[8]The index of household wealth is known as SISBEN, and it is based on the census survey targeted to household previously screened as potentially poor. Londoño-Velez, Rodriguez and Sanchez (2020) provide more details about how SISBEN was used to screen SPP-eligible students.

[9]Higher education applicants in Colombia must apply to both a major program and a college. SB11 is composed of five modules, which are given different weights depending on the major of application.

criteria to limit the influx of admitted and eventually enrolled students. As a result, the number of high-SES students enrolled in 2015 remained similar to that in 2014, but the number of low-SES students increased significantly.

Figure 2 depicts the first-term enrollment trends by SES at EU. Between 2012 and 2014, fewer than 150 first-term students were of low-SES backgrounds. When the first cohort of SPP beneficiaries enrolled, the number of low-SES students tripled to 541, while the number of students of other socioeconomic backgrounds remained almost the same. Figure 3 compares the number of low-SES students across programs in the entry cohorts before and after SPP. The gray- and blue-lined bars depict the number of low-SES students in the cohorts right before SPP (i.e., 2014-1 or spring 2014 and 2014-2 or fall 2014), whereas the gray filled bars depict the number of low-SES students in the first cohort under SPP (i.e., 2015-1 or spring 2015). The variation in the number of low–SES students is important. Majors such as business and music experienced virtually no change in the number of low-SES students, while others such as civil engineering and psychology experienced a notable increase.

Table 2 examines the relationship between the number of low- and high-SES students by program and cohort. Column (1) displays the unconditional correlation, and Column (2) controls for average program–cohort student characteristics. Both indicate that increases in the number of low-SES students are positively associated with the presence of more high-SES students in the program and cohort, suggesting that traditionally large programs enrolled more of the incoming low-SES students. Once program fixed effects are included in Column (3), the correlation between the number of low- and high-SES students is not statistically different from zero. The size of the estimated correlation also becomes much smaller in magnitude. I add entry cohort fixed effects in Column (4) to address shocks common to all programs in a given entry cohort and find no changes in the relationship. These results suggest that the positive correlation between the numbers of low- and high-SES students is a feature of certain programs and did not change with the increase in the number of low-SES students brought by SPP. Moreover, I find no evidence that the influx of low-SES students crowded out high-SES students from certain programs.

The influx of low-SES students did make classrooms busier—albeit not over capacity. Figure 4 provides descriptive statistics of the courses taken by first-term students from the 2012–2016 entry cohorts. In 2015, classroom occupation peaked but remained below 100 percent (i.e., 84 percent on average), suggesting that classrooms, on average, did not have crowding issues that could have hampered learning. These findings also mean that, in the year of SPP implementation, the university did not create more sections per course or increase the number of seats available per classroom. For EU, I find no significant increases in the number of sections per course or seats per section in the 2016 cohort, either, suggesting that the university already had the capacity to accommodate the extra students by 2015.

3 Data

The data for this paper come from two sources: EU administrative records and detailed records from the turnstiles located at each of the 18 access points to the EU campus.

EU administrative records. I use records from all students enrolled at EU between 2012 and 2018, which contain student–course-level data on student characteristics (i.e., gender, age, mother's education, high-school ID), SB11 scores, SPP recipient status, selected major, entry cohort and term of enrollment. For each semester, I observe each of the courses in which the student is enrolled and her course GPA. More importantly, I observe the student's household social stratum indicator. This six-category indicator, used to designate households eligible for utility subsidies, is a widely recognized proxy of social status in the country. I use the household social stratum at the time of college application to classify students into two SES groups: high SES (strata 3–6) and low SES (strata 1 and 2). The students benefiting from SPP mostly fall in the low-SES category. As depicted in Figure 2, most students at EU are classified as high SES.

Turnstile records. I use records on student movements on the EU campus to identify students' social interactions. The EU campus is guarded by turnstiles located at 18 entrances to main buildings and campus areas. To enter or exit through any of these entrances, students and university staff must swipe their university ID. Security officers at EU provided me individual-level records of university ID swipes at the turnstiles from February 1, 2016, to November 1, 2019. These records include the student ID number, entrance, action (IN or OUT of campus), and date, hour, minute and second of the swipe. Appendix Figure 8 displays a heat map of the average frequency of student ID swipes at three of the busiest campus entrances by 20-minutes block. Yellow and blue cells indicate peak and off-peak hours, respectively. The figure documents the constant flow of students across the campus entrances throughout the day, with peak hours at times of class changes and during lunch hours.

I define a pair of students as linked when their IDs are swiped at a turnstile in a time window of five seconds or less, at the same entrance and in the same direction (either entering or exiting campus), and when I observe the same pair of IDs comoving at least three times in a semester.

Validation of student links definition. I define a time window and frequency thresholds by comparing the links identified from the turnstile data with survey-elicited links among first-term undergraduate economics students in the fall 2017 cohort. The survey was conducted online with Qualtrics between December 7, 2017, and January 5, 2018, and elicited the network information of 106 economics students from the fall 2017 cohort. Students who completed the survey received a free lunch voucher. Cárdenas et al. (2022) provide a detailed description of the survey. The survey inquired about two types of links: friendships and acquaintances. Table 3 shows the results. The time windows tested in Table 3 were selected based on in-person observations of different entrances conducted between August 26 and 30, 2019. Because there are multiple turnstiles at each entrance, students walking together can essentially swipe their IDs simultaneously using different scanners—hence the choice of short time windows. I select the time window and comovement frequency combination that minimizes the measurement error in two steps: I look for windows and frequencies with a rate of false positives below five percent; among these, I look for the combination with the lowest false negatives rate. In this table, the false positives rate (or type I error) represents turnstile-elicited links not matched to survey links over the survey-unlinked dyads. The false negatives rate (or type II error) is the share of survey links not found in the turnstile-elicited links. The underlying assumption is that the actual numbers of links are those captured by the survey, and thus I aim to identify the turnstile-elicited links that most closely mimic those that would be obtained with the survey.

To interpret the results in Table 3, let us focus on the time window of five seconds and the survey-reported ″are friends″ links. The numbers in bold indicate the combinations of time windows and frequencies that minimize the false positives and negatives rates. In this case, comovements occurring at least three times in a semester have a false positives rate below five percent and the lowest rate of false negatives at 33 percent. Notice this is also the case for ″acquaintances″. When I constrain the turnstile-elicited comovements to happen in a two- or three-seconds window, then the frequency of comovements that reduces the false positives to below five percent while minimizing the rate of false negatives is twice in the semester. In the appendix, I discuss in detail the validation process for these definition - including an assessment of how representative turntile-elicited interactions are of survey-elicited interactions, and present the estimated effects on social interactions for the turnstile-elicited interactions derived under alternative definitions.

Sample. My analytic sample consists of all first-term students in the entry cohorts before and after the SPP rollout (i.e., fall and spring 2014 and spring 2015). I search for their interactions during the 6th and 7th calendar semesters after their first term of enrollment and among students in the same entry cohort and major. For example, I match students in the spring 2014 entry cohort with their interactions as captured by the turnstiles during fall 2016 and spring 2017. I merge administrative records and pairwise student interaction data using the student ID number, which is available in both data sources. My final sample consists of 4,027 students across 31 majors and three entry cohorts. This sample captures the universe of students enrolled in these majors and cohorts, except two programs (government and directed studies) that started after SPP and that I therefore exclude from my study.

Student characteristics. Table 1 provides descriptive statistics of high- and low-SES students in the pre- and post-SPP entry cohorts (i.e., 2014 vs. spring 2015). Both groups are similar in their observed characteristics, including gender shares, age, and mother's education level. Approximately a quarter of them graduated from high schools outside Bogotá, suggesting that they are internal migrants. The 2015 cohort of high-SES students has fewer numbers of high-school peers in the cohort (11.54 vs. 8.81 students) and slightly higher SB11 test scores (0.00 vs. 0.05 standard deviations). Both the 2014 and 2015 cohorts have similar ID swipes at the turnstiles (1,340.19 vs. 1,311.73) and similar numbers of turnstile-elicited links (5.21 vs. 5.62). In Table 9, I further document the similarities between the 2014 and 2015 cohorts and show that the high-SES student characteristics do not significantly change with the changes in the share of low-SES peers.

Low- and high-SES students differ significantly in the 2014 and the 2015 cohorts, but the differences between the two widen among the 2015 students. In 2014, low-SES students were more likely than their high-SES peers to have a mother with no college degree (24 vs. 8 percent), to be internal migrants (35 vs. 23 percent), and to have a scholarship or loan (37 vs. 7 percent). Low-SES students also had fewer high-school peers in their cohort than their high-SES peers (3.16 vs. 11.54 peers) and lower SB11 test scores (-0.10 vs. 0.00 standard deviations from the mean). These differences widen for the 2015 cohort. On average, 40 percent of 2015 low-SES students had a mother with no college degree, vs. 11

percent of their high-SES peers; 57 percent were internal migrants, relative to 24 percent of high-SES students; and 87 percent had either SPP or other forms of loans or scholarships, relative to 16 percent of high-SES students. Both low- and high-SES students had fewer high-school peers, although the gap between the two groups in favor of high-SES students persisted (8.81 vs. 1.96 high-school peers). In terms of academic achievement, the gap in SB11 test scores between low- and high-SES students also widened in 2015 to 0.21 standard deviations, relative to the 0.05 standard deviations among the 2014 students. Notably, the increase in the share of low-SES students whose mother had no college education increased from 24 to 40 percent, suggesting that the drop in performance among low-SES students could be due to those students coming from particularly disadvantaged backgrounds. I examine this divergence in achievement between the two SES groups and its implications in more detail in the following sections.

Table 1 shows that high-SES students increased their number of links with low-SES peers to an average of one (from 0.23 in 2014). Their links also became more dispersed in terms of the difference in age and SB11 test scores in 2015 but remained very similar in other characteristics such as gender or the share of links with high-school peers. The differences between high- and low-SES link characteristics also remain fairly similar between the two cohorts, except the SB11 test scores, as high-SES students had a greater distance in test scores from those of their links than low-SES students (0.79 in 2015 vs. 0.70 in 2014). Importantly, while low-SES students had similar numbers of turnstile-elicited links in the 2014 and 2015 cohorts (approximately 4.9), the 2015 low-SES students had significantly fewer ID swipes at the turnstiles than the high-SES students (1,099.8 vs. 1,311.73 swipes). These differences in ID swipes have implications for identification of the effects on social interactions, which I examine in the empirical strategy section.

Student achievement and gaps between low- and high-SES students. I characterize the differences in academic achievement between the high- and low-SES students in Figures 5 and 6. For these figures, I take advantage of the administrative data availability and plot the trends in academic achievement across the entry cohorts enrolling since 2012.

The cohort of high- and low-SES students who enrolled at EU at the onset of SPP (i.e., in the 2015 entry cohort) exhibits significant achievement gaps, particularly in GPA and cumulative credits attempted, with low-SES students having on average a lower cumulative GPA and fewer attempted credits than their high-SES peers. For example, the GPA of pre-SPP cohorts is relatively constant and close to 3.85 for both high- and low-SES students. For the SPP cohort, however, the GPA of low-SES students drops to 3.75 in the first term of college and to 3.6 by the third term, while the GPA of high-SES students remains the same. Regarding the cumulative number of credits attempted, the pre-SPP cohorts of high- and low-SES students attempted on average 50 and 48 credits by the third term, respectively. In the SPP cohort, however, low-SES students on average attempted 45.7 credits, while high-SES students continued to attempt on average 50 credits. A course at EU usually bears three credits. This means that low-SES students enrolling in 2015 had attempted on average one fewer class than their high-SES peers by the third term of college and had a cumulative GPA 0.25 points lower. Nevertheless, the differences in achievement do not pair with differences in dropout or graduation rates, suggesting that the lower achievement of low-SES students did not translate into diminished persistence.[10]

4 Identification Strategy: Effects on Academic Achievement

First, I examine the effects of increased exposure to low-SES peers on the achievement of students from the high-SES demographic groups who traditionally attend EU. I use a standard DID with a continuous treatment approach that exploits the variation in the share of low-SES peers within programs and across entry cohorts before and in the first

[10]Importantly, graduation in fewer than eight terms is very uncommon at EU across all groups, as many students take extra semesters to complete minor degrees or double major in other programs. Low-income students benefiting from SPP and other financial aid programs tend to be constrained in that they are not financed for terms beyond those scheduled for their major curriculum, which explains their slightly higher likelihood of on-time graduation.

cohort with SPP students (2014 vs. 2015-1).

$$Y_i^{mc} = \beta_l R_{mc}^l + \mathbf{X_i'} B + \beta_m + \beta_c + \varepsilon_{imc} \quad (1)$$

Equation 1 describes the econometric model. Y_i^{mc} represents the academic outcome of student i enrolled in major m and entry cohort c. R_{mc}^l represents the percentage of student i's peers who are of low SES, and $\mathbf{X_i}$ is a matrix of female, mother with no college education, and intermediate-SES indicators, SB11 scores, and age in years at the start of college. β_m and β_c capture major and entry cohort fixed effects. ε_{imc} represents robust standard errors clustered at the program and cohort level. I estimate Equation 1 using ordinary least squares (OLS). The estimated effect β_l captures the average treatment effect on the treated (ATT) of increased exposure to low-SES peers on student achievement. Figure 3 describes the variation exploited for causal identification. The percentage of low-SES students relative to that in 2014 increased at different rates across programs. Table 2 shows that the increased number of low-SES peers did not crowd out wealthy students.

Unbiased identification of β_l requires the standard parallel trends assumption: namely, that in the absence of the treatment, the outcomes for the treated and control groups would have exhibited the same trends. To test this, I estimate the placebo effects of the share of low-SES students on high-SES student achievement using data from the 2012 and 2013 entry cohorts. The results are displayed in Table 5. I find no evidence that changes in the percentage of low-SES students were associated with student outcomes in periods before SPP, suggesting that changes observed in the SPP period can be attributed to the increase in low-SES peers.

Identification would be compromised if high-SES students' allocation across majors and entry cohorts changed with SPP. At the outset of SPP, high-SES students might have self-selected into programs and entry cohorts on the basis of their preferences regarding the proximity of low-SES students; if they did, this would be reflected by changes in high-SES students' characteristics. I test whether this is the case by estimating Equation

1, without the matrix of student characteristics, for observed student sociodemographics. I display the results in Table 6. I find no evidence that the characteristics of high-SES students changed in response to the changes in the share of low-SES peers.

Recent literature has documented potential identification issues arising with the use of continuous treatments in a DID setting (Callaway, Goodman-Bacon and Sant'Anna, 2024). Estimation issues can arise from units with a negative treatment dosage, yielding biased estimated effects. To account for such potential bias, I offer alternative estimates that use as treatment a discrete variable for when the percentage of low-SES students in a program exceeds its maximum in the previous terms. The 50th percentile of the distribution of this variable for all programs is a 24 percent share of low-SES peers, which indicates that the treatment dosage is positive for all programs. I also include estimates derived with a dummy for the 75th percentile of the distribution in 2015-1, equivalent to low-SES shares of over 36 percent. Figures 10 to 13 show that the pretrends in the outcomes and observed characteristics remain parallel under these alternative, discrete variables for treatment assignment.

5 Results on the Effects of Desegregation on Achievement

Table 7 displays the estimated effect of increased exposure to low-SES peers on wealthier students' academic achievement and persistence. Panel A displays OLS estimates of β_l for the outcomes in Figures 5 and 6. In Panel A, the estimated effects on cumulative GPA by the first, third, and sixth terms and on dropout and graduation are imprecise and not statistically different from zero. The point estimates for the number of credits attempted by the first and third terms are positive and statistically significant, but their magnitudes are small. The share of low-SES students increased by 18 percentage points on average from 2014 to 2015. This yielded an increase in the number of credits taken by the first term of 0.27 credits and of 0.52 in the number of credits by the third term. Considering that the average course at EU bears three credits, this effect on courses taken is small. Overall, I do not find conclusive evidence that exposure to low-SES peers impacted the academic performance of high-SES students.

Panels B and C display results based on the discrete variables for treatment assignment for high-SES students exposed to shares of low-SES peers greater than 24 and 36 percent, respectively. The estimated effects for all outcomes continue to be imprecise and not statistically different from zero. Importantly, the positive effects on credits attempted found under the continuous treatment variable disappear. As an extra test, Panel D displays estimates that use the percentage of SPP students instead of the percentage of low-SES ones, showing no difference in findings.[11] Overall, these results are consistent with prior literature finding no effects of desegregation on traditionally privileged students' GPA in K–12 settings (Angrist and Lang, 2004; Dobbie and Fryer, 2014) and complement findings of positive effects in similar higher education settings (Bleemer, 2021b).

One hypothesis to explain the lack of effects on academic achievement is that segregation between high- and low-SES students persisted within program–cohorts. This would be consistent with findings from Carrell, Sacerdote and West (2013), which suggest that assigning low-achieving students to high-achieving classrooms can lead to segregation between the two groups. In the next section, I estimate the effects of increases in the number of low-SES peers on the social interactions of high-SES students.

6 Estimating the Effects of Desegregation on Social Interactions

I estimate the effect of the increased exposure to low-SES students on high-SES students' social interactions using Equation 1. Additionally, I control for the number of peers

[11] Unlike the number of low-SES peers, the number of SPP peers is zero for 2014 and takes positive values for 2015.

from the same high school in the student cohort, as this is likely to confound the social interactions that high-SES students have with other socioeconomic groups. I link pairs of students when I observe them swiping their student IDs at the same entrance and going in the same direction within a window of five seconds or less and at least three times in a semester. The appendix shows the results when the turnstile-elicited interactions are defined in an alternative way, documenting the consistency of my estimates.

My use of turnstile-elicited interactions makes the DID setup susceptible to another possible source of bias: the observed interactions could be the result of coincidental co-movements across the turnstiles of pairs not socially interacting—i.e., they could suffer from measurement error. If these comovements are nonrandom, they could be falsely attributed to the effects of exposure to more low-SES peers.

To build testable implications, I rely on a potential outcomes framework in a simplified 2×2 DID. Define t as a treated group (i.e., a group with a large R^{l}_{mc}) and u as an untreated group:

$$\hat{\alpha_P}^{2x2} = (E[L_t|Post] - E[L_t|Pre]) - (E[L_u|Post] - E[L_u|Pre]) \tag{2}$$

In Equation 2, the estimated $\hat{\alpha_P}^{2x2}$ is written as the difference between the expected post- and pretreatment values of the outcome L for the treated group t $(E[L_t|Post] - E[L_t|Pre])$ minus the difference between the expected post- and pretreatment values of the outcome L for the untreated group u $(E[L_u|Post] - E[L_u|Pre])$. Equation 2 can be rewritten in potential outcome terms. Define as L^0 the potential outcome had no treatment been assigned and as L^1 the potential outcome had the treatment been assigned:

$$
\begin{aligned}
\hat{\alpha_P}^{2x2} &= \underbrace{E[L_t^1|Post] - E[L_t^0|Post]}_{\text{ATT}} \\
&+ \underbrace{\underbrace{(E[L_t^0|Post] - E[L_t^0|Pre])}_{\text{treatment counterfactual}} + (E[L_u^0|Post] - E[L_u^0|Pre])}_{\text{nonparallel trend bias==0}}
\end{aligned} \tag{3}
$$

Equation 3 implies that $\hat{\alpha_P}^{2x2}$ is composed of the ATT and the bias from nonparallel trends. I showed in section 4 that there is no evidence of the latter. However, if measurement error in the outcome L is associated with the treatment, the estimated ATT may differ from the true ATT.

I define the number of links I aim to measure as $L^{true} = L^{obs} - L^{F(+)} + L^{F(-)}$. That is, true links are defined as the number of observed links L^{obs} minus the links falsely defined as such—the false positive turnstile-elicited links $L^{F(+)}$—plus the number of true links not captured by the turnstile-elicited measure $L^{F(-)}$, the false negatives:

$$
\begin{aligned}
ATT^{estimated} &= E[L_t^{1,obs} - L_t^{1,F(+)} + L_t^{1,F(-)}|Post] - E[L_t^{0,obs} - L_t^{0,F(+)} + L_t^{0,F(-)}|Post] \\
&= \underbrace{E[L_k^{1,obs}|Post] - E[L_k^{0,obs}|Post]}_{\text{observed ATT}} + \\
&\underbrace{E[L_t^{1,F(-)} - L_t^{1,F(+)}|Post] - E[L_t^{0,F(-)} - L_t^{0,F(+)}|Post]}_{\text{measurement error bias}}
\end{aligned} \tag{4}
$$

Thus, the estimated ATT can be rewritten as:

$$
ATT^{estimated} = ATT^{obs} + \underbrace{E[L_t^{1,F(-)} - L_t^{0,F(-)}|Post]}_{\text{ATT on F(-)}} - \underbrace{E[L_t^{1,F(+)} - L_t^{0,F(+)}|Post]}_{\text{ATT on F(+)}} \tag{5}
$$

Equation 5 implies that if the treatment has no impact on $L^{F(-)}$ or $L^{F(+)}$, then $ATT^{estimated} = ATT^{obs}$. Ideally, I would have data on the measurement error variables $L^{F(-)}$ and $L^{F(+)}$

across different majors and cohorts such that I could use variation in the treatment R^l_{mc} to assess its effects. Since I do not have data of this nature, I rely on proxy variables that can help me assess the extent to which the treatment R^l_{mc} may lead to measurement error bias. I use two variables for this: the number of ID swipes at the turnstiles for each student, and the number of courses a student took with a turnstile-elicited link. I measure both proxies for the same enrollment terms for which I measure interactions (i.e., the sixth and seventh terms after first enrollment).

Intuitively, if the treatment led to more ID swipes at the turnstiles, the chances that I am capturing false positives $L^{F(+)}$ in the treated group increases. If the treatment led to fewer ID swipes, the chances of missing true links $L^{F(-)}$ in the treated group increases. If the treatment led to more classes being taken with links, the turnstile-elicited links may reflect casual comovements with classmates rather than actual connections, thus increasing the rate of false positives $L^{F(+)}$.

Table 4 displays the results of regressing R^l_{mc} on the measurement error proxies under each time window considered. The estimation follows the same structure as that of Equation 1 but uses the proxy variables on the left-hand side. The estimated coefficients are small in magnitude relative to the pretreatment averages—approximately 0.1 standard deviations—and statistically insignificant. I conclude there is no evidence that measurement error in the turnstile-elicited interactions biases my causal estimations.

6.1 Results

Table 8 displays the results from Equation 1 showing the effects of increased exposure to low-SES peers on students' social interactions. Panel I displays the estimated effects on the probability of interaction with high- and low-SES peers, Panel II displays the estimated impacts on the number of links (i.e., unique connections), and Panel III displays the impacts on the percentage of links with low-SES students. This last measure describes how much a student's connections diversify in response to desegregation in her group.

Panel A shows the estimates for the continuous treatment variable, following Equation 1, whereas Panels B and C display the estimates based on the discrete variables for treatment assignment.

Increased exposure to low-SES peers significantly changed the social interactions of high-SES students. Focusing on Panel A of Table 8, the average increase of 18 percentage points increased the probability of interaction between high- and low-SES students by 14.4 percentage points and the number of links among them by 0.70 links. I also find evidence of a reduction in interactions among high-SES students of 0.65 links (significant at 95 percent). The increase in the share of low-SES peers almost doubled the probability of interaction between the two SES groups from its pre-SPP level and almost tripled the number of links among them.

Segments B and C of Panels I and II in Table 8 show that the largest changes in the number of links occurred in programs where the percentage of low-SES peers exceeded 36 percent. In these cases, the number of unique links between low- and high-SES students increased by one to four times the average number of links before SPP. I also see a reduction in the number of links among high-SES students of 1.3, suggesting that in programs with the largest shares of low-SES enrollees, students substituted high- for low-SES interactions. Notably, the probability of links with low-SES students did not increase beyond the initial average of 14.4 percentage points. This might be explained by students mechanically responding to the compositional changes in their group rather than by higher levels of exposure significantly changing students' preferences in favor of interacting with low-SES peers.

The last panel of Table 8 displays the estimates of how sensitive high-SES students' friendships were to changes in the socioeconomic composition of their group. Specifically, I estimate the effect of changes in the share of low-SES peers on the percentage of high-SES students' low-SES friends. If high-SES students experienced merely compositional responses to the changes of peers in their group, the estimated effect would be one:

a one-percentage-point increase in low-SES peers would translate to a one-percentage-point increase in the share of low-SES friends. Estimates over one would suggest that high-SES students were even more welcoming of low-SES students among their friends, and estimates below one would suggest some reluctance to accept lower-SES links. My estimates bear out the latter effect. For every additional percentage point of the low-SES peer share, the percentage of links with low-SES students increases by only three-quarters (0.75) of a percentage point, suggesting some aversion among high-SES students to changing their friend group in the same ratio as their peer group changed.

To further assess the aversion to forming links with low-SES students, I follow Chetty, Jackson, Kuchler, Stroebel, Hendren, Fluegge, Gong, Gonzalez, Grondin, Jacob et al. (2022) and apply their concept of friendship bias.[12] In the context of this paper, friendship bias is the tendency of high-SES students to befriend low-SES peers at lower rates than high-SES peers. It is mathematically defined as one minus the percentage of low-SES friends over the share of low-SES peers in the program and entry cohort. Values close to one suggest a high friendship bias that favors links with other high-SES peers, whereas values close to zero suggest no friendship bias. Similarly, values below zero suggest a bias that favors links with low-SES students beyond their representation in the group.

The friendship bias analysis suggests that in the 2015-1 cohort, there is a pattern of favoring same-SES links, regardless of the share of students in the program and cohort. Figure 7 plots the estimated friendship bias for programs in the 2014 and 2015-1 cohorts relative to the percentage of low-SES peers in each group. These results suggest large variation in friendship bias in the cohorts enrolling before SPP and no relation with the percentage of low-SES students in the group. The red dots, which plot the estimated friendship bias in the 2015-1 entry cohort, suggest that the variation in the bias diminished, but with no visible relationship with the percentage of low-SES students. The

[12] Chetty, Jackson, Kuchler, Stroebel, Hendren, Fluegge, Gong, Gonzalez, Grondin, Jacob et al. (2022) define friendship bias as "the tendency for people with low SES to befriend people with high SES at lower rates even conditional on exposure".

estimated friendship bias does seem to be consistently over one, with only a few exceptions. I regress Equation 1 using the student-level friendship bias as outcome and find no significant relationship. Coupling these results with those in Panel III of Table 8, I conclude that interactions between low- and high-SES students did increase but that the response was not proportional to the new shares of low-SES peers in the group and a bias for friendships among high-SES peers persisted.

7 Role of Academic Achievement in the Diversity of Social Interactions

My results show significant performance gaps between low- and high-SES students enrolling after SPP, with low-SES students underperforming their high-SES peers. I find no decreases in high-SES students' performance with greater exposure to low-SES peers, but I do find that the two groups were socially interacting. Arcidiacono and Vigdor (2010) and Epple and Romano (2011) find that exposure to low-achieving peers has a negative impact on student performance, and Carrell, Sacerdote and West (2013) suggest that such exposure effects are shaped by the social interactions among the students. Could the increased interactions between low- and high-SES students be driven by interactions among the high-achievers?

While incoming low-SES students are on average lower performers than their high-SES counterparts, there is variation in the distribution of scores among the two groups: 27 percent of the low-SES students had SB11 test scores equal to or above the average scores of their high-SES peers. I use this variation to identify high-achieving low-SES students and measure how much of the change in social interactions was driven by interactions with students of this type. I use three measures of performance: SB11 test scores, first-term GPA, and credits attempted in the first term. I flag a low-SES student as a high achiever if her performance is equal to or above the average performance of the high-SES students in the program and cohort. The SB11 exam is taken before college enrollment, which makes it independent of first-term social interactions, but outcomes on this exam might be harder for the students to observe.

Table 9 displays the results. Overall, my estimates suggest that the increased interactions between low- and high-SES students are partially driven by links formed with high-achieving low-SES students. According to Panel II, when the share of low-SES peers increases by 18 percentage points and when the achievement measure is SB11 test scores, the estimated effect on the number of links with low-SES high-achieving students is 0.31 links. This is approximately 40 percent of the increase in low-SES links estimated in Table 8. However, while the increase in the probability of a link with a low-SES student relative to pretreatment levels is 80 percent overall, for links with high-achieving low-SES peers, the increase is 140 percent—from a pretreatment level of 9 percent. This suggest that high-SES students befriend high-achieving low-SES peers at higher rates that they do low-SES students in general. These effects are more pronounced when I use the (possibly endogenous) achievement measures potentially observed by students such as the number of credits that students attempted in their first term and their GPA.[13]

These results complement previous findings examining diversity in social interactions—namely, the findings of positive impacts of increased diversity in schools on interaction intensity as measured by email exchanges (Marmaros and Sacerdote, 2006) and survey questions about the willingness to interact with racially and ethnically diverse groups (Boisjoly et al., 2006; Rao, 2019). My results provide a finer disaggregation by distinguishing the impacts on the probability and number of interactions with peers in the same group. The findings complement those of Mayer and Puller (2008) and Baker, Mayer and Puller (2011), who examine changes in the composition of friendships with a measure of interactions bounded to peers in the same college group.[14] My findings add to this

[13]Notably, I find no evidence that the increased exposure changed academic achievement among high-SES students. This suggests that endogeneity is less of an issue in this particular context.

[14]Mayer and Puller (2008) and Baker, Mayer and Puller (2011) use data from Facebook to study whether students' friendships on that social media platform become more diverse when the students are exposed

literature by showing that diversity in group composition increases interactions among students from different socioeconomic groups despite the persistent bias toward same-group friendships.

8 How Close Are the Newly Diverse Interactions?

My results so far indicate that social interactions did diversify at the onset of SPP, though a bias for same-SES interactions persisted. They also suggest that this diversity in interactions is largely driven by high-SES students befriending high-achieving low-SES students. How close are these new interactions between low- and high-SES students? In this section, I leverage the richness of the turnstile data to proxy for the closeness of interactions between students. I focus on interactions that happened during off-peak hours of traffic through the turnstiles (see Appendix Figure 8) and look for comovements happening early in the morning (before 7 am) and at the end of the day (after 5:30pm). Interactions during these hours would suggest students at least partially commute together to or out of campus, indicating their interactions extent to other out-of-classes activities. In this case, I relax my definition of social interactions and flag a link when there is at least one comovement through the turnstile within five seconds and during the given time block. The results are displayed in Table 10.

Some of the new interactions between students of diverse SES are particularly close and are only partially mediated by their taking courses together. Panel II shows that approximately 28 percent of the 0.039 increase estimated in Table 8 corresponds to interactions happening after 5:30pm. These interactions could be driven by the pair of student taking a course together that either starts or finishes after 5:30pm. To account for this, I control for the number of courses the linked students took together. The size of the point

to diverse peers in their dorms. The authors argue that the effects on the diversity of friendships are small. In contrast to my measure of social interactions, their measure of social networks is not bounded to peers from college.

estimate declines but continues to be positive and significant, accounting for 23 percent of the estimated initial effect. Overall, these results indicate that between a quarter to a third of the new cross-SES interactions are fairly close.

9 Conclusion

In this paper, I study whether changes in the socioeconomic composition of students at an elite university diversified social interactions and explore the role of academic achievement as both outcome and mechanism. I exploit variation in the share of low-SES students driven by Ser Pilo Paga, a financial aid program in Colombia targeting high-achieving low-income students. To measure social interaction, I leverage records on turnstiles located across the 18 college campus entrances and develop a measured based on students' comovements across campus.

I find that the increased exposure to low-SES peers led to more interactions between high- and low-SES students, albeit with some persistent bias for interactions among high-SES students. Importantly, close to half of the increase in interactions with low-SES peers is explained by interactions with high-achieving low-SES peers, and in general, students are more likely to befriend high-achieving low-SES students than before SPP. Notably, the increased diversity at EU had no impact on the overall academic achievement of low-SES students—as measured by their GPA, credits attempted and retention rates.

These findings provide evidence of how socioeconomic desegregation of elite colleges can impact students within the institution. Similarly to Angrist and Lang (2004) and Bleemer (2021a), I show that there are no adverse impacts on the achievement of students from the privileged demographic groups traditionally attending these institutions. However, the diversity of social interactions can increase for students traditionally attending this elite institution. High academic achievement among low-SES students is an important contributor to the diversification of interactions, as high-SES students befriend high achieving low-SES students at higher rates. Notably, my findings suggest around a third of these interactions are fairly close, in that they occur around commuting hours and are only partially mediated by students coming and going from classes together.

Coupled with the findings from Londoño-Velez, Rodriguez and Sanchez (2020) and Londoño-Vélez et al. (2023) that SPP effectively increased higher education access and completion among the low-income students that it aimed to benefit, my findings contribute to the evidence on the positive effects of affirmative action and financial aid targeted to low-income students, showing that this policy did not harm the achievement of students from the demographic groups that traditionally concentrate in elite institutions while it did diversify their interactions.

Tatiana Velasco. Teachers College, Columbia University.

References

Angrist, Joshua D. and Kevin Lang. 2004. "Does school integration generate peer effects? Evidence from Boston's metco program." American Economic Review 94(5):1613–1634.

Arcidiacono, Peter and Jacob L. Vigdor. 2010. "Does the river spill over? estimating the economic returns to attending a racially diverse college." Economic Inquiry 48(3):537–557.

Arcidiacono, Peter, Michael Lovenheim and Maria Zhu. 2015. "Affirmative Action in Undergraduate Education." Annual Review of Economics 7(1):487–518.

Arias Ortiz, Elena, Gregory Elacqua and Carolina Gonzalez-Velosa. 2017. "Making the Most of Higher Education.".

Baker, Sara, Adalbert Mayer and Steven L. Puller. 2011. "Do more diverse environments increase the diversity of subsequent interaction? Evidence from random dorm assignment." Economics Letters 110(2):110–112.

Bleemer, Zachary. 2021a. "Affirmative Action, Mismatch, and Economic Mobility After California's Proposition 209." The Quarterly Journal of Economics 137(1):1–46.

Bleemer, Zachary. 2021b. "Top Percent Policies and the Return to Postsecondary Selectivity." Manuscript .

Boisjoly, Johanne, Greg J. Duncan, Michael Kremer, Dan M. Levy and Jacque Eccles. 2006. "Empathy or antipathy? The impact of diversity." American Economic Review 96(5):1890–1905.

Cárdenas, Juan Camilo, Danisz Okulicz, Davide Pietrobon, Tomás Rodríguez and Tatiana Velasco. 2022. "Propensity to Trust and Network Formation.".

Carranza, Juan Esteban and María Marta Ferreyra. 2019. "Increasing higher education access: Supply, sorting, and outcomes in colombia." Journal of Human Capital 13(1):95–136.

Carrell, Scott E., Bruce Sacerdote and James E. West. 2013. "From Natural Variation to Optimal Policy? The Importance of Endogenous Peer Group Formation." Econometrica 81(3):855–882.

Chetty, Raj, John N Friedman, Emmanuel Saez, Nicholas Turner and Danny Yagan. 2020. "Income Segregation and Intergenerational Mobility Across Collegres in the United States." Quarterly Journal of Economics pp. 1567–1633.

Chetty, Raj, Matthew O. Jackson, Theresa Kuchler, Johannes Stroebel, Nathaniel Hendren, Robert B. Fluegge, Sara Gong, Federico Gonzalez, Armelle Grondin, Matthew Jacob, Drew Johnston, Martin Koenen, Eduardo Laguna-Muggenburg, Florian Mudekereza, Tom Rutter, Nicolaj Thor, Wilbur Townsend, Ruby Zhang, Mike Bailey, Pablo Barberá, Monica Bhole and Nils Wernerfelt. 2022. "Social capital I: measurement and associations with economic mobility." Nature 608(7921):108–121.
URL: *https://www.nature.com/articles/s41586-022-04996-4*

Chetty, Raj, Matthew O Jackson, Theresa Kuchler, Johannes Stroebel, Nathaniel Hendren, Robert B Fluegge, Sara Gong, Federico Gonzalez, Armelle Grondin, Matthew Jacob et al. 2022. "Social capital II: determinants of economic connectedness." Nature 608(7921):122–134.

Corno, Lucia, Eliana La Ferrara and Justine Burns. 2022. "Interaction, stereotypes, and performance: Evidence from South Africa." American Economic Review 112(12):3848–75.

Dobbie, Will and Roland G. Fryer. 2014. "The impact of attending a school with high-achieving peers: Evidence from the New York City exam schools." American Economic Journal: Applied Economics 6(3):58–75.

Echenique, Federico, Roland G. Fryer and Alex Kaufman. 2006. "Is school segregation good or bad?" American Economic Review 96(2):265–269.

Epple, Dennis and Richard E Romano. 2011. Peer Effects in Education: A Survey of the Theory and Evidence. In Hanbook of Social Economics. Elsevier B.V. chapter 20.

Lau, Yan. 2022. "Does racial diversity improve academic outcomes? A natural experiment in higher education classrooms." Labour Economics 78:102254.

Londoño-Vélez, Juliana. 2022. "The impact of diversity on perceptions of income distribution and preferences for redistribution." Journal of Public Economics 214:104732.

Londoño-Velez, Juliana, Catherine Rodriguez and Fabio Sanchez. 2020. "Upstream and Downstream Impacts of College Merit-Based Financial Aid for Low-Income Students: Ser Pilo Paga in Colombia." American Economic Journal: Economic Policy 12(2):193–227.

Londoño-Vélez, Juliana, Catherine Rodríguez, Fabio Sánchez and Luis Esteban Álvarez. 2023. "Financial Aid and Social Mobility: Evidence From Colombia's Ser Pilo Paga." Mimeo .

Marmaros, David and Bruce Sacerdote. 2006. "How Do Friendships Form?" The Quarterly Journal of Economics 121(1):79–119.

Mayer, Adalbert and Steven L. Puller. 2008. "The old boy (and girl) network: Social network formation on university campuses." Journal of Public Economics 92(1-2):329–347.

Mele, Angelo. 2017. "A Structural Model of Dense Network Formation." Econometrica 85(3):825–850.

Mele, Angelo. 2020. "Does school desegregation promote diverse interactions? An equilibrium model of segregation within schools." American Economic Journal: Economic Policy 12(2).

Mello, Ursula. 2022. "Centralized Admissions, Affirmative Action and Access of Low-income Students to Higher Education." American Economic Journal: Economic Policy 14:3.

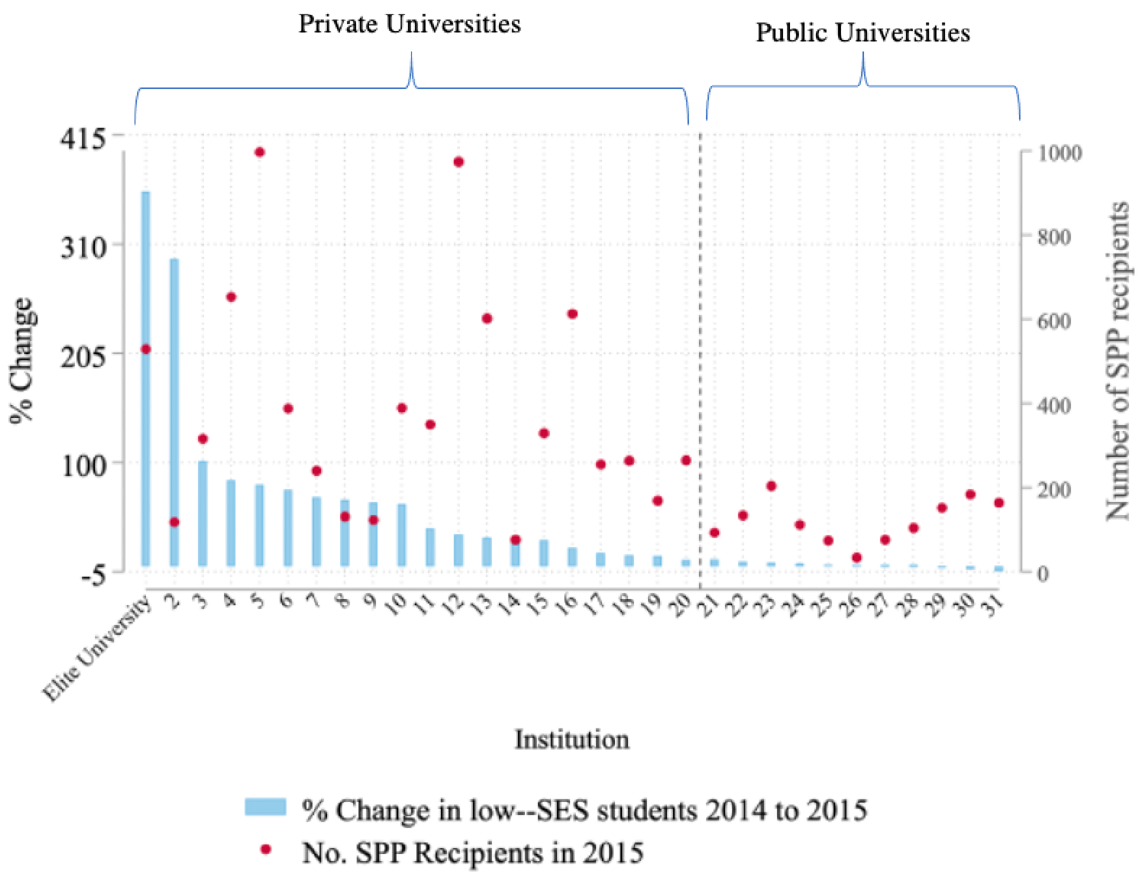
Michelman, Valerie, Joseph Price and Seth D Zimmerman. 2022. "Old boys' clubs and upward mobility among the educational elite." The Quarterly Journal of Economics 137(2):845–909.

Rao, Gautam. 2019. "Familiarity Does Not Breed Contempt: Generosity, Discrimination and Diversity in Delhi Schools." American Economic Review 109(3):774–809.

Zimmerman, Seth D. 2019. "Elite colleges and upward mobility to top jobs and top incomes." American Economic Review 109(1):1–47.

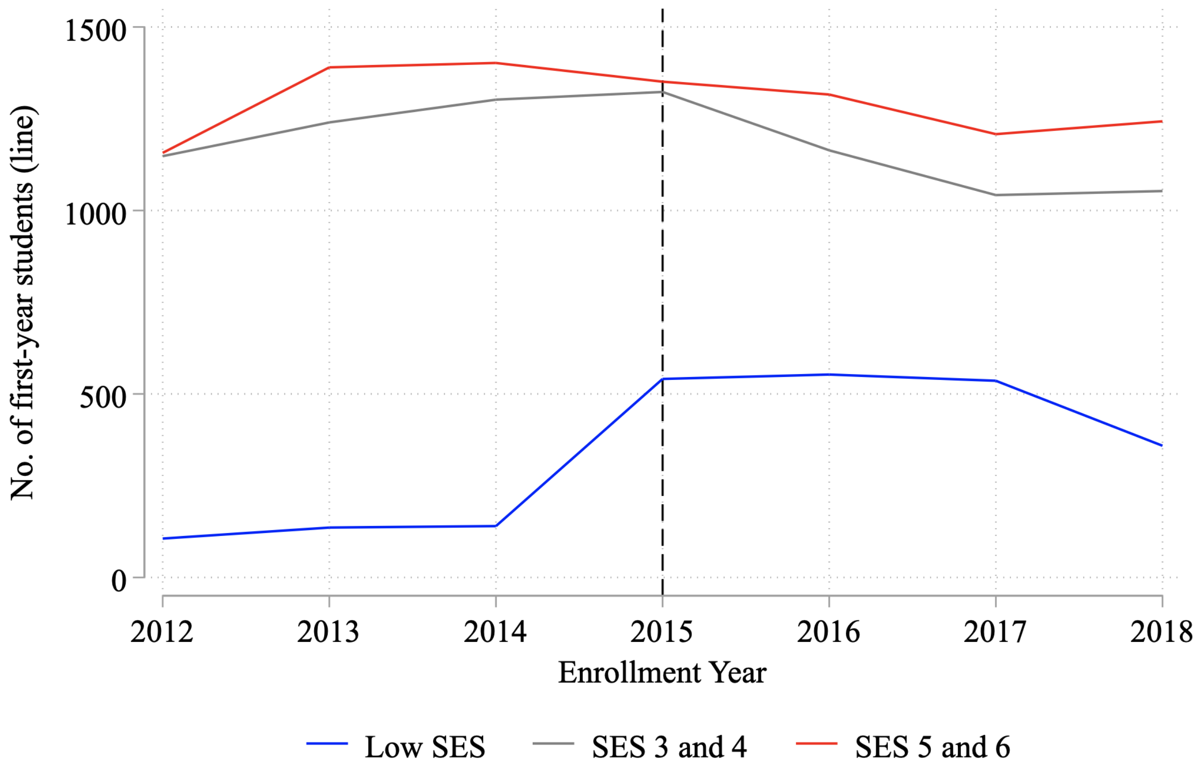
Figures

Figure 1: Change in the percentage of low-SES students across SPP-eligible universities



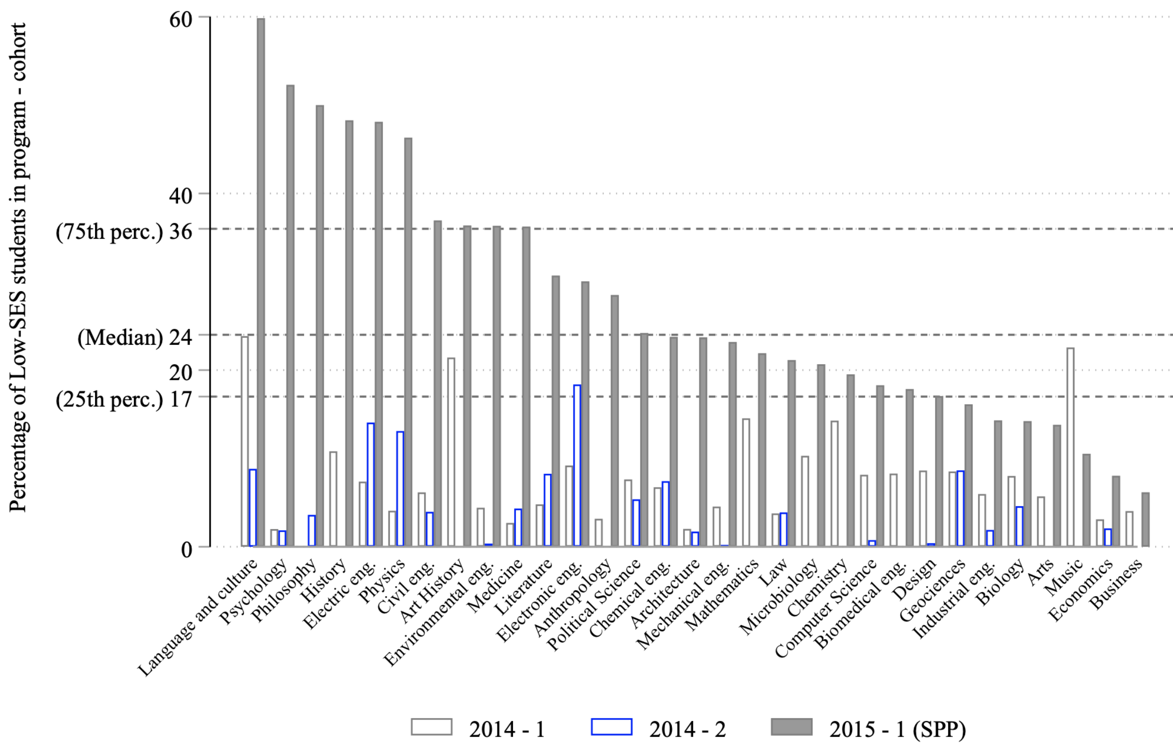
Note: This figure displays the percentage change in the number of low-SES students enrolling between the 2014 and 2015 entry cohorts at each of the SPP-eligible universities in Colombia. Calculations are based on publicly available data from the Ministry of Education.

Figure 2: Number of first-term students by SES



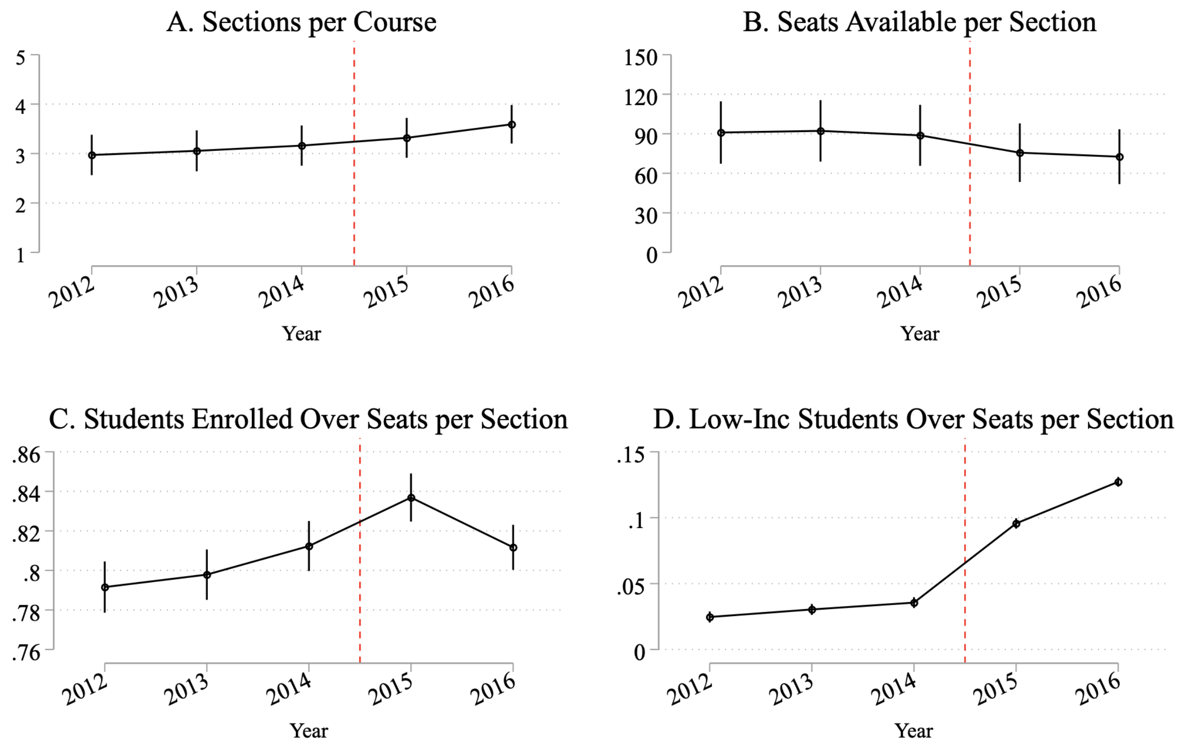
Note: This figure displays the total number of first-term students by SES group. Students are classified into three SES groups based on their housing stratum indicator. High-SES students are those from socioeconomic strata three to six, while low-SES students are those from socioeconomic strata one and two. I included both the spring and fall enrollments per year. The dotted vertical line marks the start of SPP.

Figure 3: Percentage of low–SES students by major and before and after SPP



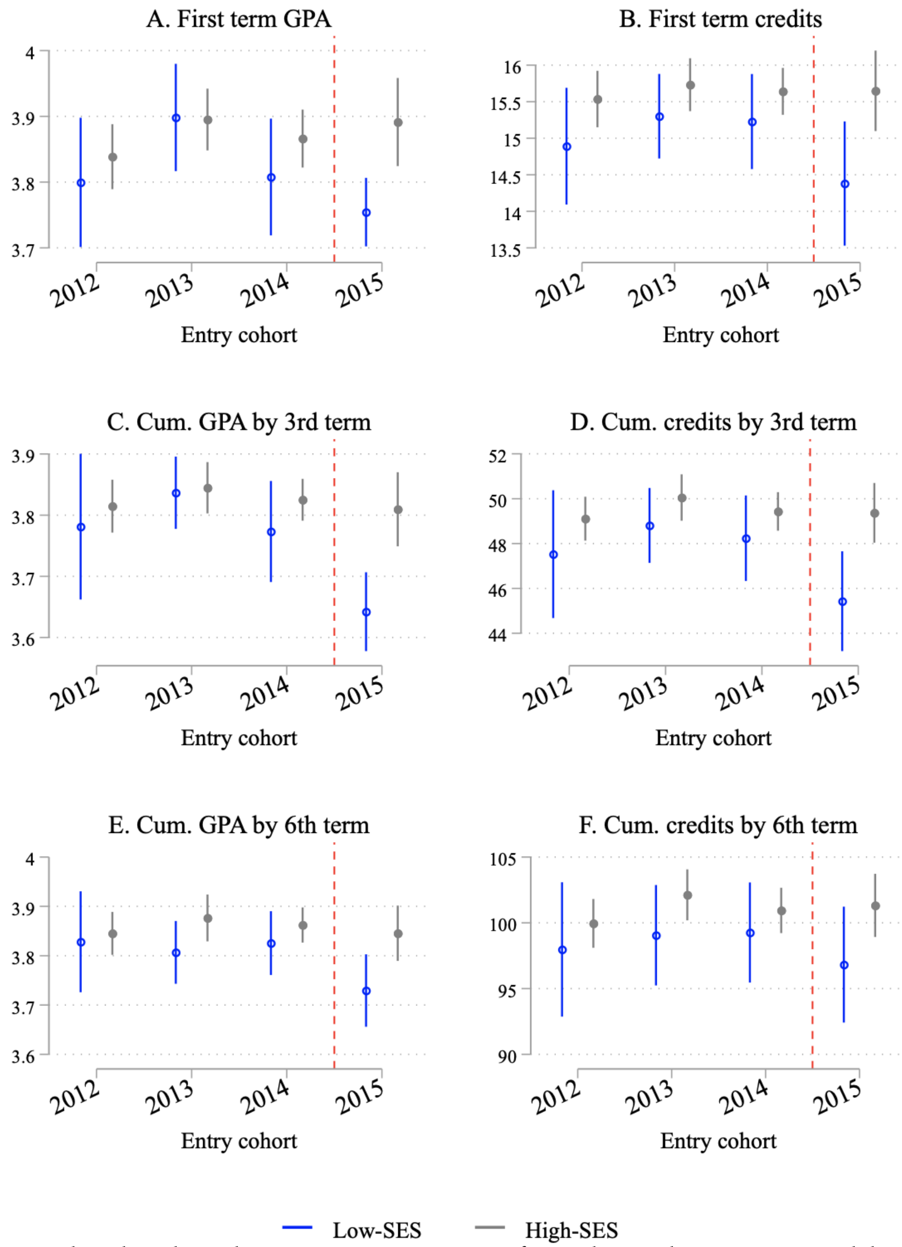
Note: This figure displays the percentage of low-income students by program and entry cohort period. The dotted horizontal lines indicate the 25th, 50th (median), and 75th percentile values of the distribution of low-SES students during the SPP period, 2015-1.

Figure 4: Composition of courses taken by first-term students in each entry cohort



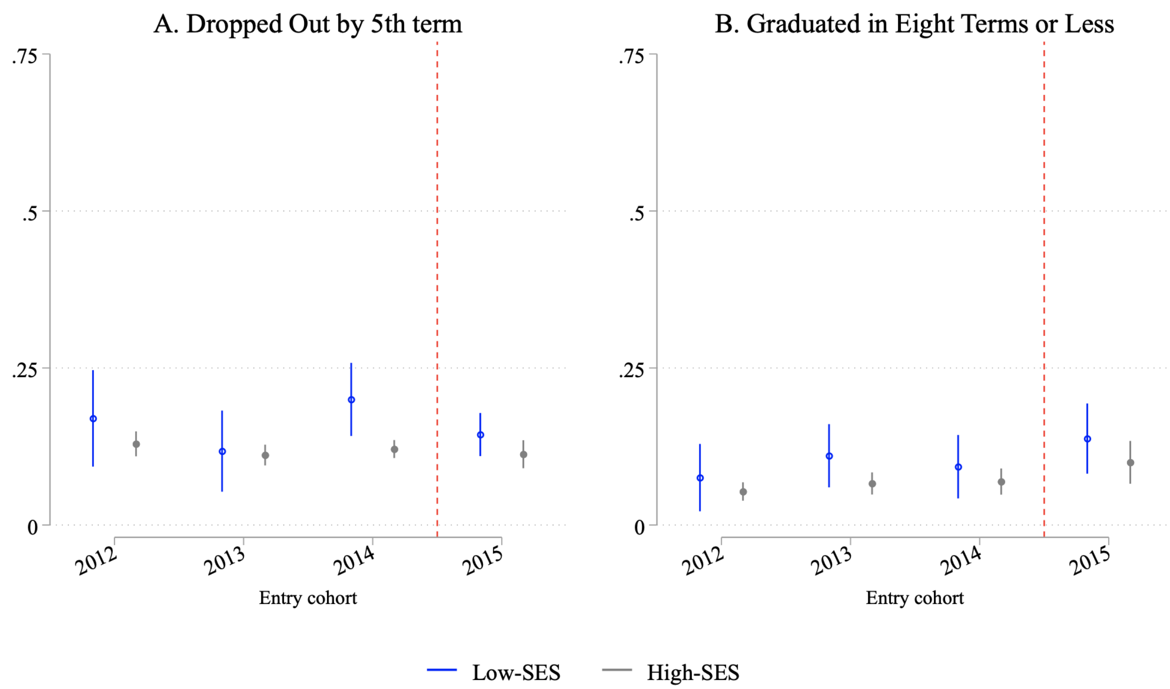
Note: This figure illustrates the composition of courses and classrooms attended by first-term students in each entry cohort. Panel A shows the average number of sections per course, Panel B shows the average number of available spots or seats per section, Panel C shows the ratio of enrolled students to the total number of available seats, and Panel D shows the ratio of low-SES students enrolled to the total number of available seats. In each panel, every point represents results from an OLS regression with no constant and dummies by entry year; 95 percent confidence intervals are plotted as a vertical line on each point. The data were previously aggregated at the section–course–term level. The dotted red line separates the cohorts that enrolled before the start of SPP (2014 and earlier) from those that enrolled during SPP (2015 onward).

Figure 5: Achievement gaps between high- and low-SES students by entry cohort



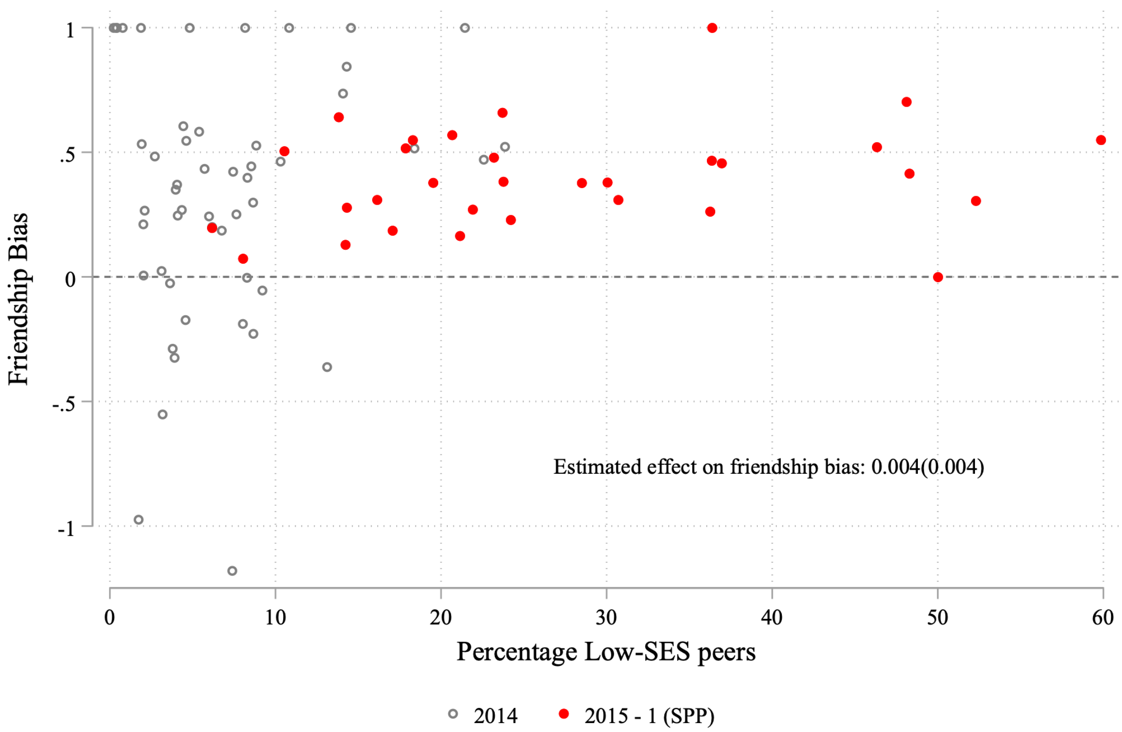
Note: These graphs display the point estimates of a cohort dummy variable from an OLS regression with no intercept, where the dependent variable is the student's GPA, ranging from one to five, with five being the highest grade, or the number of credits attempted. The 95 percent confidence intervals are shown as vertical lines on each dot and are based on clustered standard errors at the program–cohort level. Each yearly entry cohort includes the spring and fall cohorts of the respective calendar year. The dotted red line separates the cohorts that enrolled before the start of SPP (2014 and before) from the first SPP cohort, 2015-1.

Figure 6: Persistence gap between high- and low-SES students by entry cohort



Note: These graphs display the point estimates of a cohort dummy variable from an OLS regression with no intercept, where the dependent variable is an indicator equal to one if the student dropped out by the 5th term or graduated in fewer than eight terms. A student is labeled as a dropout if she does not show up as enrolled for two consecutive terms after the fifth term of college. I label a student as graduating if she completed the degree in eight terms or fewer. The 95 percent confidence intervals are shown as vertical lines on each dot and are based on clustered standard errors at the program–cohort level. Each yearly entry cohort includes the spring and fall cohorts of the respective calendar year. The dotted red line separates the cohorts that enrolled before the start of SPP (2014 and before) from the first SPP cohort, 2015-1.

Figure 7: Estimated friendship bias among programs and entry cohorts before and after SPP



Note: Graph displays the average friendship bias for low-SES friendships among high-SES students. Friendship bias follows the definition in Chetty, Jackson, Kuchler, Stroebel, Hendren, Fluegge, Gong, Gonzalez, Grondin, Jacob et al. (2022) and is calculated as one minus the average percentage of low-SES links over the percentage of low-SES peers in the program and entry cohort. Thus, values close to one indicate a bias for same-SES friendships, and values below zero suggest a bias in favor of friendships with students from low-SES backgrounds. Values of zero suggest no bias, as the percentage of friendships with low-SES peers would equal the size of their presence in the program–cohort.

Table 1: Descriptive statistics

	2014 entry cohort			2015 entry cohort		
	High SES	Low SES		High SES	Low SES	
	Mean	Mean	t-test	Mean	Mean	t-test
Composition of Peers						
Number of links	5.00	4.75	0.65	5.55	4.97	1.90
Low-income links	0.23	0.32	1.59	1.02	1.96	5.44
Student Characteristics						
Female	0.43	0.34	2.15	0.45	0.41	0.91
Age	17.59	17.24	3.94	17.59	17.13	10.54
Mother with no college degree	0.08	0.24	5.74	0.11	0.40	14.14
SB11 standardized test score	0.00	–0.10	1.17	0.05	–0.16	2.90
SPP recipient	0.00	0.00	N.A.	0.09	0.84	39.09
Other scholarship or loan	0.07	0.37	6.95	0.07	0.03	3.36
Internal migrant	0.23	0.35	2.67	0.24	0.57	8.55
No. of HS peers in cohort	11.54	3.16	12.53	8.81	1.96	18.72
ID swipes in 6th and 7th terms	1340.19	1349.79	0.11	1311.73	1099.80	3.91
Link Characteristics						
Age difference	0.58	0.65	0.97	0.65	0.65	0.00
Share of same-gender friends	0.49	0.51	0.48	0.49	0.47	0.94
Courses taken together in first term	1.47	1.31	1.44	1.34	1.40	0.46
SB11 difference	0.70	0.77	1.28	0.79	0.67	3.53
Share of friends from same high school	0.04	0.01	4.80	0.03	0.01	6.67
Number of students	2,669	139		1,358	463	
Number of majors	31	31		31	31	

Note: This table shows the descriptive statistics for the sample of students described in Section 3. The 2015 entry cohort includes only the spring term (2015-1). High-SES students are those from household strata three to six, and low-SES students are from household strata one and two. The t-test corresponds to the hypothesis that the difference in means between high- and low-SES students is equal to zero. The t-tests are based on standard errors clustered at the program level.

Table 2: Correlation between numbers of high- and low-SES students in a program and entry cohort

	(1)	(2)	(3)	(4)
No. of low-SES peers in program–cohort	1.195***	1.654***	−0.232	−0.201
	(0.325)	(0.421)	(0.171)	(0.166)
Average of student characteristics		x	x	x
Major fixed effects			x	x
Entry cohort fixed effects				x
No. of program–cohort groups	93	93	93	93

Note: This table displays OLS estimates correlating the number of high- and low-SES students in a program and entry cohort between 2014 and 2015-1. The number of high-SES students is the dependent variable, and the number of low-SES students is the explanatory variable. Each observation in the data corresponds to one program and entry cohort. The average of student characteristics in a major–cohort group included are the share female, average age in years at entry, share of students whose mothers have no college education, average SB11 standardized test scores, share of students from SES 2 and 3, and share of SPP students. *** p <0.01, ** p <0.05, *, * p<0.1.

Table 3: Comparison of survey– and turnstile–elicited links

Time window	A. Two seconds					B. Three seconds					C. Five seconds				
Frequency	One	Two	Three	Four	Five	One	Two	Three	Four	Five	One	Two	Three	Four	Five
1. Turnstiles															
No. of dyads	850	366	234	179	148	1182	506	313	250	198	1858	887	549	399	315
No. of students	106	106	105	105	103	106	106	106	106	105	106	106	106	106	106
2. Are friends															
Dyads			*505*					*505*					*505*		
Survey & turnstiles															
Matched	342	256	201	165	140	389	305	248	215	179	433	368	337	295	263
False negatives (type II)	0.32	**0.49**	0.60	0.67	0.72	0.23	**0.40**	0.51	0.57	0.65	0.14	0.27	**0.33**	0.42	0.48
False positives (type I)	0.10	**0.02**	0.01	0.00	0.00	0.16	**0.04**	0.01	0.01	0.00	0.28	0.10	**0.04**	0.02	0.01
3. Acquaintances															
Dyads			*1033*					*1033*					*1033*		
Survey & turnstiles															
Matched	497	311	219	174	144	606	391	284	235	191	734	537	425	348	293
False negatives (type II)	0.52	**0.70**	0.79	0.83	0.86	0.41	**0.62**	0.73	0.77	0.82	0.29	0.48	**0.59**	0.66	0.72
False positives (type I)	0.08	**0.01**	0.00	0.00	0.00	0.13	**0.03**	0.01	0.00	0.00	0.25	0.08	**0.03**	0.01	0.00

Note: N students = 106. Number of dyads possible (N*(N-1))/2 = 5,565. Survey sample consists of economics undergraduates from the August 2017 cohort. The false negatives or type II error rate is the share of links in the survey that were not found in the turnstile-based links. The false positive or type I error is the rate of turnstile-elicited links unmatched to survey links over the unlinked survey dyads (5,565 survey dyads observed). I look to minimize the rate of false negatives when the rate of false positives is below five percent.

Table 4: ATT on measurement error proxies

		No. of courses in the semester interacting with peers in:		
	ID swipes	Two seconds	Three seconds	Five seconds
	(1)	(2)	(3)	(4)
Percentage of low-income peers	-123.791	0.057	-0.299	-0.180
	(284.499)	(0.457)	(0.521)	(0.470)
Pretreatment statistics for outcomes				
Mean	1340.192	1.091	1.118	1.135
Standard deviation	1017.184	1.367	1.399	1.414
No. of students	4,027	4,027	4,027	4,027
No. of major–cohort groups	93	93	93	93

Note: Results from estimating Equation 1 using measurement error proxies on the left-hand side. "ID swipes" is the total number of ID swipes of each student, either to enter or exit campus, in the sixth and seventh terms after first enrollment. "No. of courses with peers interacted" is the total number of courses that the student took with the peers whom I identify as turnstile-elicited links. All estimations include fixed effects by major and entry cohort and the covariates described for Equation 1. All standard errors are clustered at the major–cohort level.

Table 5: Placebo impact of exposure to desegregation on academic achievement and persistence – 2012 and 2013 cohorts

	(1) 1st-term credits	(2) 1st-term GPA	(3) 3rd-term cum. credits	(4) 3rd-term cum. GPA	(5) 6th-term cum. credits	(6) 6th-term cum. GPA	(7) Dropout by 5th term	(8) Graduation on time
A. Continuous Treatment								
Percentage of low-SES peers	0.011	0.000	0.052	0.003*	–0.073	0.001	0.002	0.001
	(0.013)	(0.002)	(0.038)	(0.002)	(0.086)	(0.002)	(0.001)	(0.001)
B. Outcomes Statistics (2014 Cohort)								
Mean	15.540	3.837	49.088	3.813	99.928	3.844	0.129	0.053
Standard deviation	3.079	0.476	8.580	0.386	16.818	0.342	0.336	0.225
No. of students	3,569	3,563	3,253	3,253	2,981	2,981	3,569	3,569
No. of major–cohort groups	93	93	93	93	93	93	93	93

Note: This table displays placebo estimates of the effect of exposure to different percentages of low-SES peers on high-SES students' academic outcomes. Results from estimating Equation 1 in the sample of students enrolling in the 2012 and 2013 entry cohorts. All regressions control for a female indicator, age in years at the time of entry, SB11 standardized test scores, mother without a college degree, socioeconomic stratum two or three (i.e., intermediate SES), receipt of an SPP loan, and number of high-school peers enrolled in the same cohort as the student. All standard errors are clustered at the program–cohort level. *** $p < 0.01$, ** $p < 0.05$, *, * $p < 0.1$.

Table 6: Estimates of relationship between share of low-SES peers and high-SES student characteristics

	(1) Female	(2) Age	(3) Mother w/o college	(4) Test scores	(5) Mid-SES	(6) HS peers
A. Continuous Treatment						
Percentage of low-SES peers	–0.001	–0.001	0.000	0.001	0.001	0.016
	(0.001)	(0.003)	(0.001)	(0.003)	(0.001)	(0.029)
B. Pretreatment Statistics for Outcomes						
Mean	0.434	17.591	0.083	–0.052	0.487	11.544
Standard deviation	0.496	0.912	0.276	0.955	0.500	11.711
No. of students	4,027	4,027	4,027	4,027	4,027	4,027
No. of major–cohort groups	93	93	93	93	93	93

Note: This table displays estimates of the effect of exposure to different shares of low-SES peers on high-SES students' observed characteristics from Equation 1 and on students enrolling between 2014 and 2015-1. All standard errors are clustered at the program–cohort level. *** p <0.01, ** p <0.05, *, * p<0.1.

Table 7: Effects of increased exposure to low-SES peers on academic outcomes

	(1) 1st-term credits	(2) 1st-term GPA	(3) 3rd-term cum. credits	(4) 3rd-term cum. GPA	(5) 6th-term cum. credits	(6) 6th term cum. GPA	(7) Dropout by 5th term	(8) Graduation on time
A. Continuous Treatment								
Percentage of low-SES peers	0.015***	0.001	0.029*	0.001	0.020	0.001	0.001*	0.000
	(0.005)	(0.001)	(0.017)	(0.001)	(0.035)	(0.001)	(0.001)	(0.001)
B. 50th Percentile								
$\mathbb{I}$[% of low-SES peers $>$ 24%]	0.063	−0.010	−0.030	0.015	0.226	0.015	0.022	−0.021
	(0.141)	(0.030)	(0.448)	(0.022)	(0.747)	(0.020)	(0.018)	(0.028)
C. 75th Percentile								
$\mathbb{I}$[% of low-SES peers $>$ 36%]	0.071	0.011	0.505	0.017	0.541	0.025	0.037*	−0.035
	(0.177)	(0.030)	(0.610)	(0.027)	(0.828)	(0.025)	(0.019)	(0.038)
D. Percentage of SPP								
Percentage of SPP	0.006	0.000	0.021	0.001	0.024	0.000	0.002***	−0.001
	(0.006)	(0.001)	(0.017)	(0.001)	(0.033)	(0.001)	(0.001)	(0.001)
Pretreatment Statistics for Outcomes								
Mean	15.637	3.863	49.386	3.822	100.865	3.859	0.122	0.069
Standard deviation	2.949	0.449	8.496	0.378	16.223	0.344	0.327	0.254
No. of students	4,027	4,024	3,730	3,730	3,407	3,407	4,027	4,027
No. of major–cohort groups	93	93	93	93	93	93	93	93

Note: Panel A displays the results of estimating Equation 1. Panels B and C display the results based on dummy variables for treatment assignment designating programs with low-SES student shares above the median and 75th percentile, respectively. Panel D displays the results based on the percentage of SPP recipients in the program. All regressions control for a female indicator, age in years at the time of entry, SB11 standardized test scores, mother without a college degree, socioeconomic stratum two or three (i.e., intermediate SES), receipt of an SPP loan, and number of high-school peers enrolled in the same cohort as the student. All standard errors are clustered at the program–cohort level. *** p <0.01, ** p <0.05, *, * p<0.1.

Table 8: Effects of increased exposure to low-SES peers on students' social interactions

	I. Probability of a Link with a:		II. Number of Links with:			III. % of Links with:
	(1) Low SES	(2) High SES	(3) Low SES	(4) High SES	(5) Any Student	(6) Low SES
A. Continuous Treatment						
Percentage of low-SES peers	0.008***	–0.002	0.039***	–0.036**	0.003	0.752***
	(0.001)	(0.001)	(0.004)	(0.018)	(0.019)	(0.056)
Mean Increase (18.0 points)	0.144	–0.036	0.702	–0.648	0.054	13.536
B. 50th Percentile						
$\mathbb{I}$[% of low-SES peers $>$ 24%]	0.110***	–0.034	0.717***	–0.856*	–0.139	15.086***
	(0.040)	(0.028)	(0.167)	(0.477)	(0.498)	(2.318)
C. 75th Percentile						
$\mathbb{I}$[% of low-SES peers $>$ 36%]	0.141***	–0.069**	1.042***	–1.287**	–0.245	18.707***
	(0.049)	(0.033)	(0.163)	(0.583)	(0.655)	(3.057)
Pretreatment Statistics for Outcomes						
Mean	0.180	0.752	0.231	4.771	5.002	4.386
Standard deviation	0.384	0.432	0.548	4.816	5.040	11.145
No. of students	4,027	4,027	4,027	4,027	4,027	3,079
No. of major–cohort groups	93	93	93	93	93	91

Note: A socially interacting pair of students is defined when their IDs are swiped at a turnstile at the same entrance and in the same direction within five seconds or less and at least three times during a semester. The outcomes in Panel I are indicators equal to one when the student has interacted with at least one low- or high-SES peer. Panel II uses the number of peers whom the student has interacted with, and Panel III uses the percentage of low-SES links. Panel A displays the results of estimating Equation 1. Panels B and C display results based on dummy variables for treatment assignment designating programs with low-SES student shares above the median and 75th percentile, respectively. All regressions control for a female indicator, age in years at the time of entry, SB11 standardized test scores, mother without a college degree, socioeconomic stratum two or three (i.e., intermediate SES), receipt of an SPP loan, and number of high-school peers enrolled in the same cohort as the student. All standard errors are clustered at the program–cohort level. *** p <0.01, ** p <0.05, *, * p<0.1.

Table 9: Effects of increased exposure to low-SES peers on students' interactions with high-achieving low-SES students

	I. Probability of Link w/ Low-Income High Achiever by:			II. Number of Links with Low-Income High Achiever by:		
	(1) SB11	(2) GPA	(3) Credits Attempted	(4) SB11	(5) GPA	(6) Credits Attempted
A. Continuous Treatment						
Percentage of low-SES peers	0.007***	0.006***	0.009***	0.017***	0.021***	0.029***
	(0.002)	(0.001)	(0.002)	(0.003)	(0.003)	(0.004)
Mean Increase (18.0 p.p.)	0.126	0.108	0.162	0.306	0.378	0.522
B. 50th Percentile						
$\mathbb{I}$[% of low-SES peers > 24%]	0.119**	0.050	0.180***	0.323***	0.277**	0.614***
	(0.048)	(0.047)	(0.054)	(0.092)	(0.122)	(0.137)
C. 75th Percentile						
$\mathbb{I}$[% of low-SES peers > 36%]	0.175***	0.168***	0.215***	0.497***	0.609***	0.787***
	(0.057)	(0.040)	(0.062)	(0.112)	(0.085)	(0.173)
Pretreatment Statistics for Outcomes						
Mean	0.090	0.121	0.113	0.100	0.146	0.131
Standard deviation	0.286	0.327	0.316	0.334	0.427	0.393
No. of students	4,027	4,027	4,027	4,027	4,027	4,027
No. of major–cohort groups	93	93	93	93	93	93

Note: A socially interacting pair of students is defined when their IDs are swiped at a turnstile at the same entrance and in the same direction within five seconds or less and at least three times during a semester. The outcomes are indicators equal to one when the student has interacted with at least one low-SES student with performance above her high-SES peers' mean in terms of SB11 (columns (1) and (4)), first-term GPA (columns (2) and (5)) or first-term credits attempted (columns (3) and (6)). Panel A displays the results of estimating Equation 1. Panels B and C display the results based on dummy variables for treatment assignment designating programs with low-SES student shares above the median and 75th percentile, respectively. All regressions control for a female indicator, age in years at the time of entry, SB11 standardized test scores, mother without a college degree, socioeconomic stratum two or three (i.e., intermediate SES), receipt of an SPP loan, and number of high-school peers enrolled in the same cohort as the student. All standard errors are clustered at the program–cohort level. *** p <0.01, ** p <0.05, *, * p<0.1.

Table 10: Effects of increased exposure to low-SES peers on links between high- and low-SES students during off-peak hours

	I. Before 7 am		II. 5:30 pm to 9 pm	
	(1)	(2)	(3)	(4)
Number of Links				
Percentage of low-SES peers	0.001*	0.001*	0.011***	0.009***
	(0.000)	(0.001)	(0.001)	(0.001)
Number of courses together		0.963***		0.257***
		(0.012)		(0.027)
Pretreatment Outcomes				
Mean	0.012		0.173	
Standard deviation	0.107		0.378	
No. of students	4,027	4,027	4,027	4,027
No. of major–cohort groups	93	93	93	93

Note: Regressions in this table use as outcome the number of unique links high-SES students have with low-SES students, based on their having at least one comovement within a five-second window during the off-peak hours. Regressions in columns (1), (3), (5) and (7) control for a female indicator, age in years at the time of entry, SB11 standardized test scores, mother without a college degree, socioeconomic stratum two or three (i.e., intermediate SES), receipt of an SPP loan, and number of high-school peers enrolled in the same cohort as the student. Regressions in columns (2), (4), (6), and (8) also control for the number of courses the student took with the low-SES link that either start or end during the off-peak block. All standard errors are clustered at the program–cohort level. *** p <0.01, ** p <0.05, *, * p<0.1.

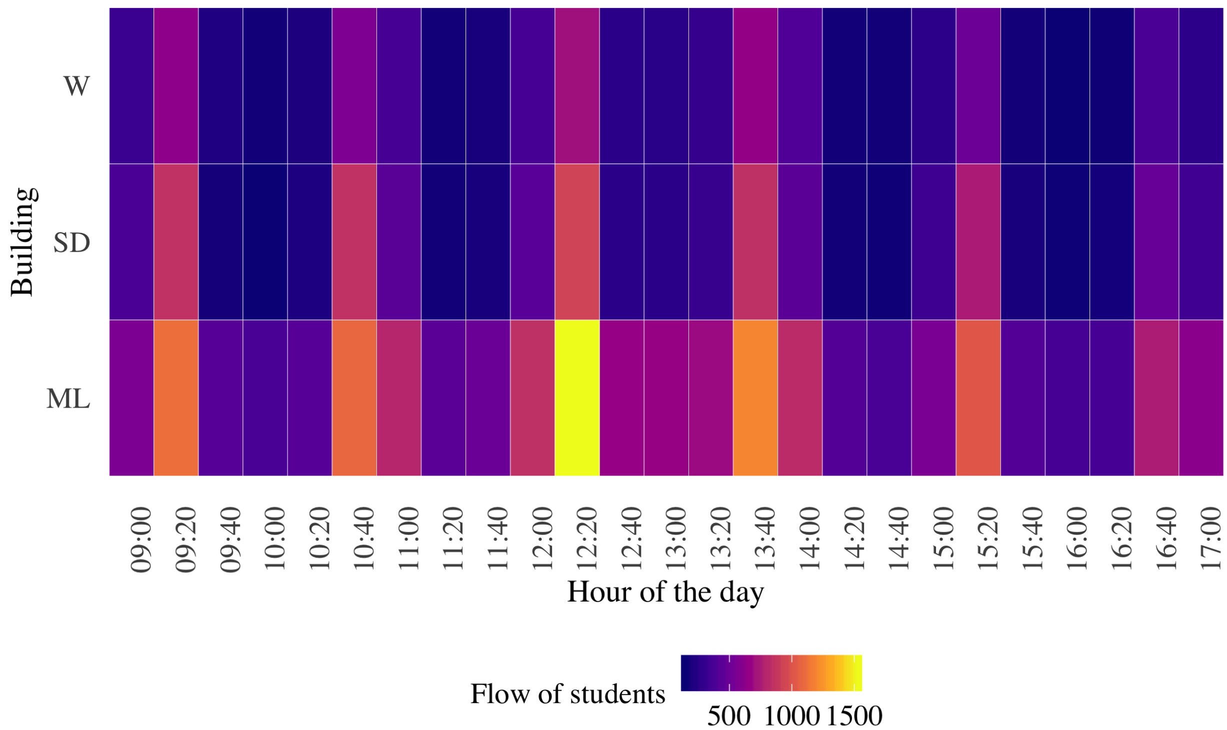
Online Appendix: Additional Descriptives on Turnstile-Elicited Interaction and Robustness Results

In this Appendix, I provide further evidence validating my definition of turnstile-elicited interactions, present additional descritive estatistics, and results under alternative interactions definitions.

The results in Table 3 indicate that, under the baseline definition, it is highly likely that the turnstile-elicited links capture links similar to those reported in the survey. However, an important share of the survey-reported links may not be captured by the turnstiles (i.e., they may be false negatives). This could be an issue to the extent that the links that I do capture are not representative of the survey-elicited links. To address this concern, I compare whether the turnstile-elicited links plausibly reflect the survey-elicited network characteristics. The results are displayed in Figure 9. The goal of this exercise is to estimate how far from random the characteristics of the turnstile-elicited links are and how close the average characteristics of the links are to those of the survey-elicited links. The computation proceeds as follows: I take the results from Table 3 and randomly assign the turnstile-elicited links that minimize the measurement error to the 106 students in the sample. Then, I compute the average of the following network individual attributes: age difference, number of courses that the students take together, GPA difference, degree or number of links, and local clustering. I conduct this procedure 1,000 times and plot the distribution of the characteristics. I include the average value that I observe for the turnstile- and survey-elicited links with the 95 percent confidence interval. I find statistically significant evidence that the turnstile-elicited network characteristics closely resemble those of the friendship and acquaintance networks elicited by the survey and are not the result of random link formation.

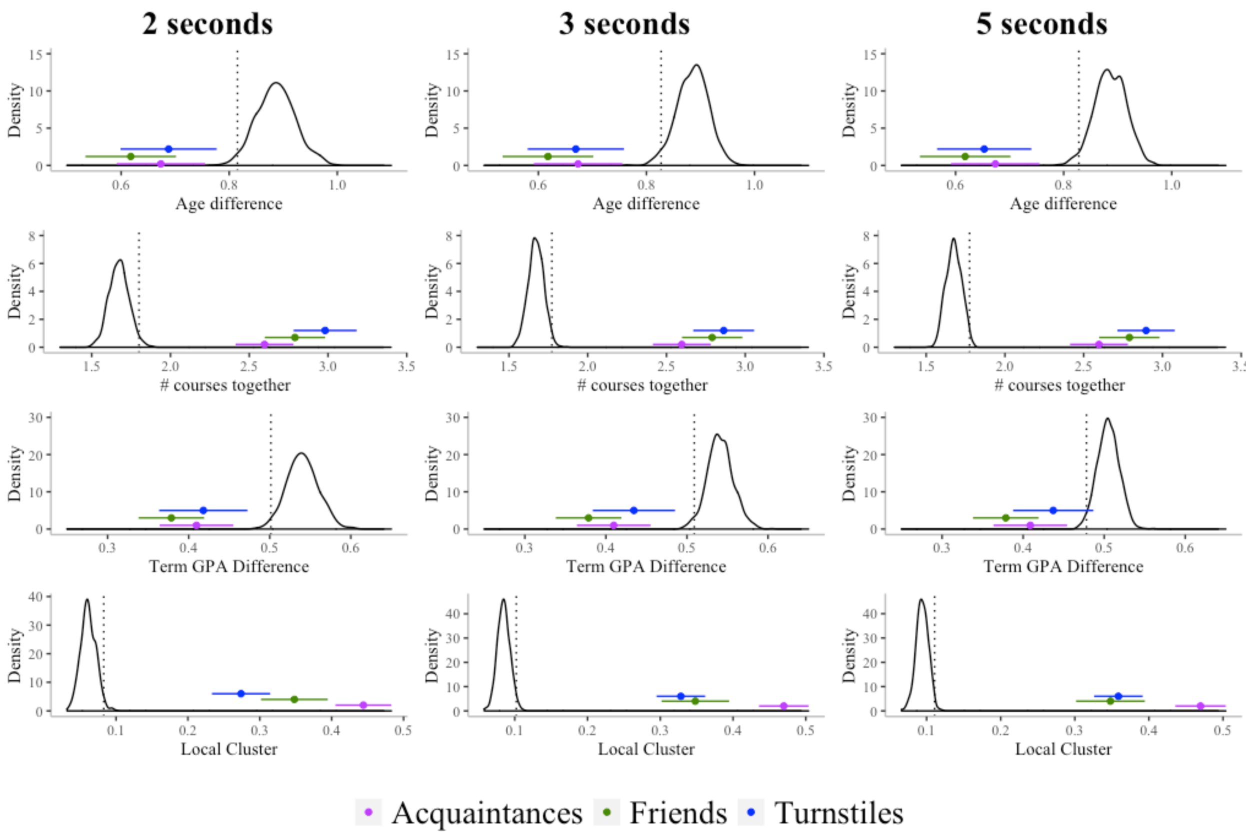
The validity of the turnstile-elicited interaction data could be sensitive to the hours of the day during which comovements are captured. Comovements captured around lunch hours could be more susceptible to yielding false positives or could exclude interactions not related to students' usual lunchtime activities, thereby increasing the chances of false negatives affecting my dataset. I test the extent to which this is an issue by replicating the comparison with the survey-elicited interactions from Table 3 but for comovements happening around lunchtime hours (from 11:40 am to 2:20 pm) with comovements at other times. The results are displayed in Table 11. For simplicity, I focus on friends' links and comovements within a three-second and a five-second window. Comovements captured during lunchtime are more susceptible to suffering from false negatives than comovements captured outside lunch hours. These results suggest that, to minimize measurement error, searching for comovements at any time of day is more reliable than focusing on comovements happening at specific times of day.

Figure 8: Flow of students at selected entrances – Term and hour according to turnstiles



Note: Average number of swipes by day, entrance and 20-minute block. Swipes include building entries and exits. Only observations from weekdays during the official academic calendar are included in the data.

Figure 9: Comparison with randomly generated distribution



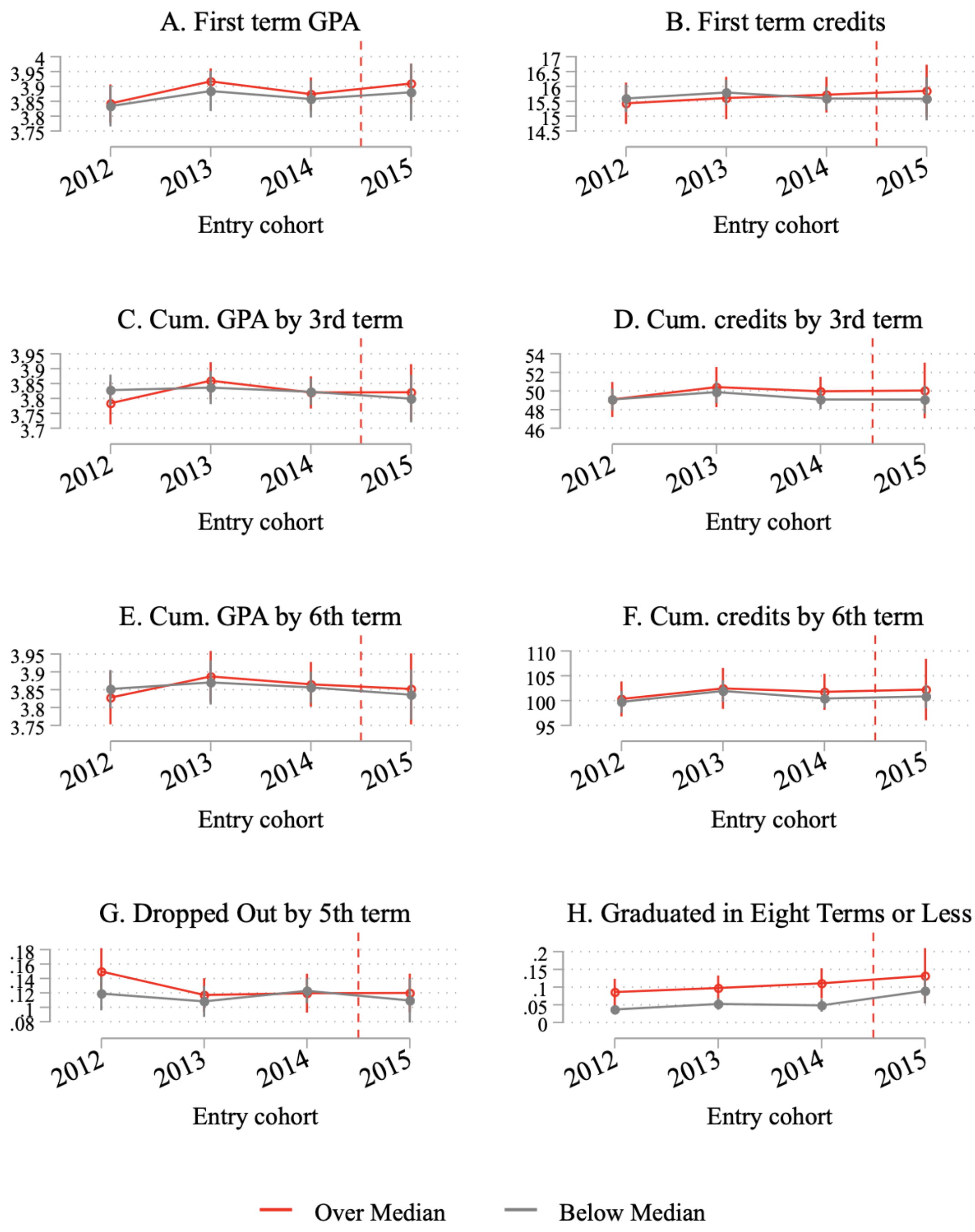
Note: Turnstile-elicited links matched with the survey links are randomly assigned in 1,000 draws among 106 students, forming all possible 5,565 dyads. The 95 percent confidence intervals are presented. Matches for a 2-second time window with 2 comovements: 366 links. Matches for 3-second time window with 2 comovements: 506 links. Matches for a 5-second time window with 3 comovements: 549 links. The dotted vertical lines indicate the 95 percent confidence points in the distribution.

Table 11: Comparison of survey- and turnstile-elicited links during and outside lunch hours

Time window	A. Three seconds						B. Five seconds					
Type	11:40 am to 2:20 pm			Other times			11:40 am to 2:20 pm			Other times		
Frequency	One	Two	Three	One	Two	Three	One	Two	Three	One	Two	Three
1. Turnstiles												
No. of dyads	551	213	135	869	373	232	860	355	233	1444	636	393
No. of students	106	106	104	106	106	104	110	109	108	110	110	108
2. Are friends												
Dyads	*505*						*505*					
Survey & turnstiles												
Matched	231	155	119	338	252	192	277	203	174	392	316	274
False negatives (type II)	0.78	**0.85**	0.88	0.67	**0.76**	0.81	0.45	**0.60**	0.66	0.22	0.37	**0.46**
False positives (type I)	0.06	**0.01**	0.00	0.10	**0.02**	0.01	0.12	**0.03**	0.01	0.21	0.06	**0.02**

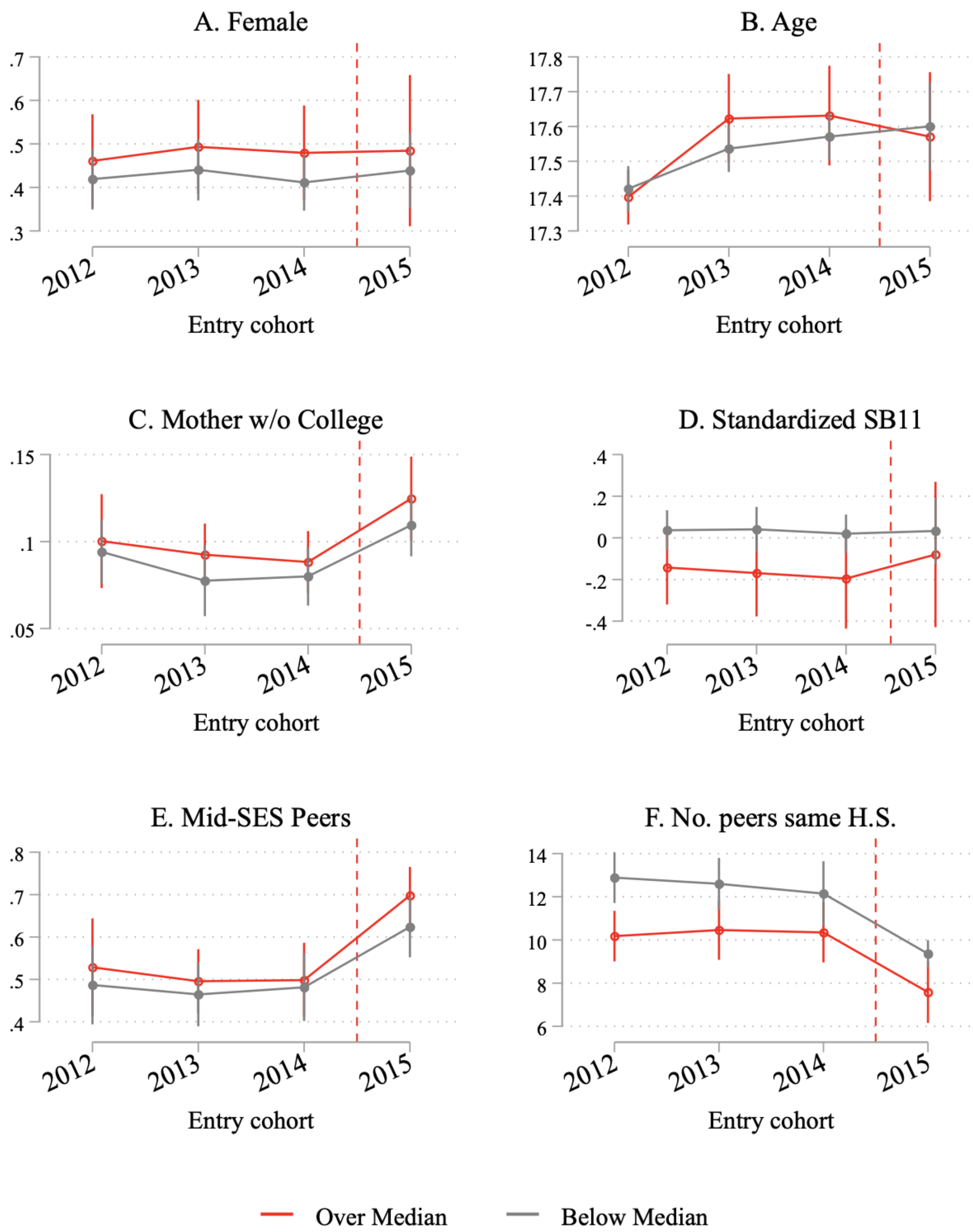
Note: N students = 106. Number of links possible (N*(N-1))/2 = 5,565. Survey sample consist of economics undergraduates from the August 2017 cohort. The false negatives or type II error rate is the share of links in the survey that were not found in the turnstile-based links. The false positive or type I error is the rate of turnstile-elicited links unmatched to survey links over the unlinked survey dyads (5,565 survey dyads observed). I look to minimize the rate of false negatives when the rate of false positives is below five percent.

Figure 10: Pretreatment trends in observed outcomes – Programs with above- and below-median shares of low-SES peers in 2015



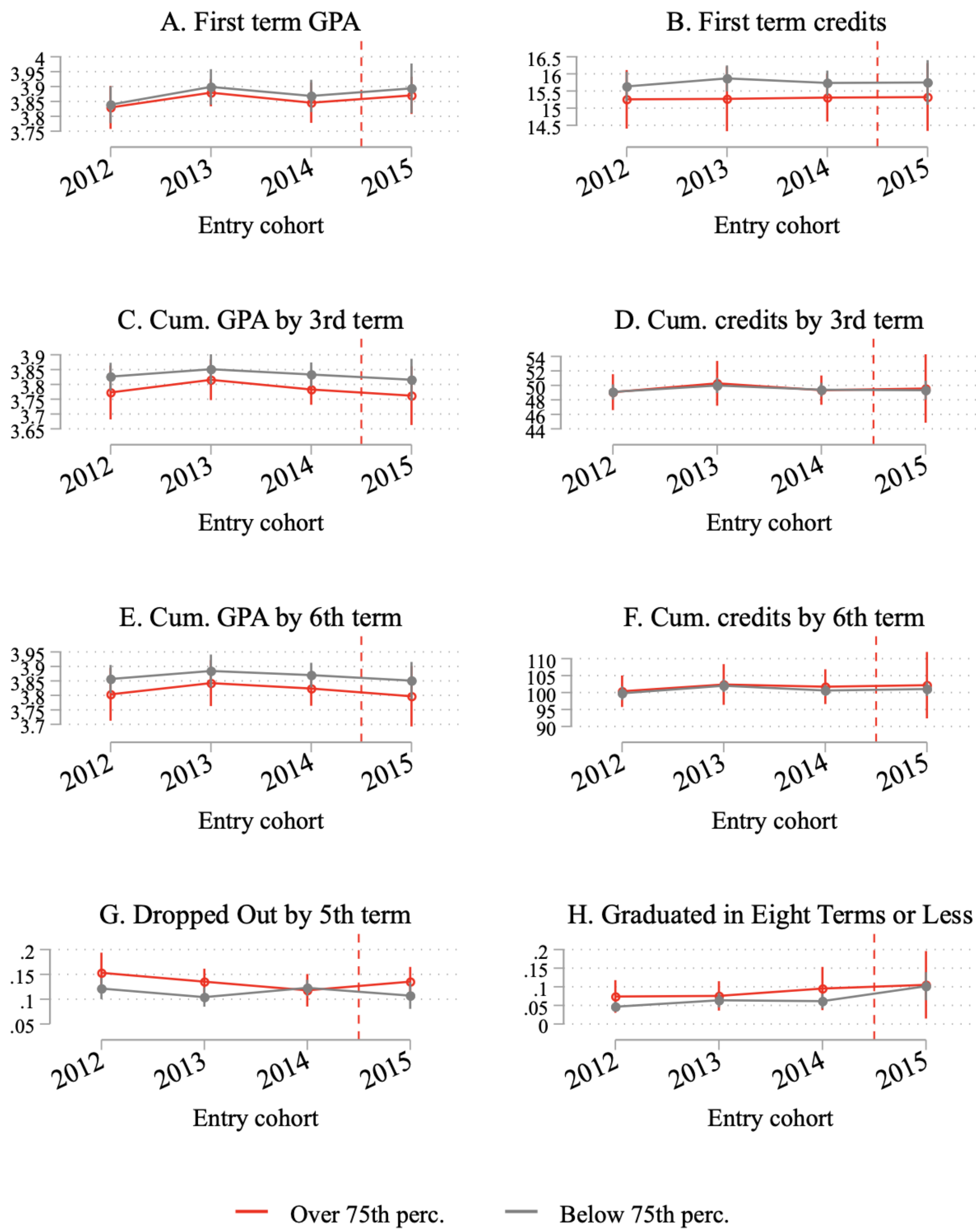
Note: Graphs display point estimates of a cohort dummy variable from an OLS regression with no intercept using student outcomes as the dependent variable. Programs with a low-SES student share above the median value during the SPP period are classified as "Over Median" and others as "Below Median." The 95 percent confidence intervals use clustered standard errors at the program–cohort level. Yearly entry cohorts include both spring and fall, with a red line dividing pre-SPP and post-SPP (2015-1) cohorts.

Figure 11: Pretreatment trends in observed student sociodemographics – Programs with above- and below-median shares of low-SES peers in 2015



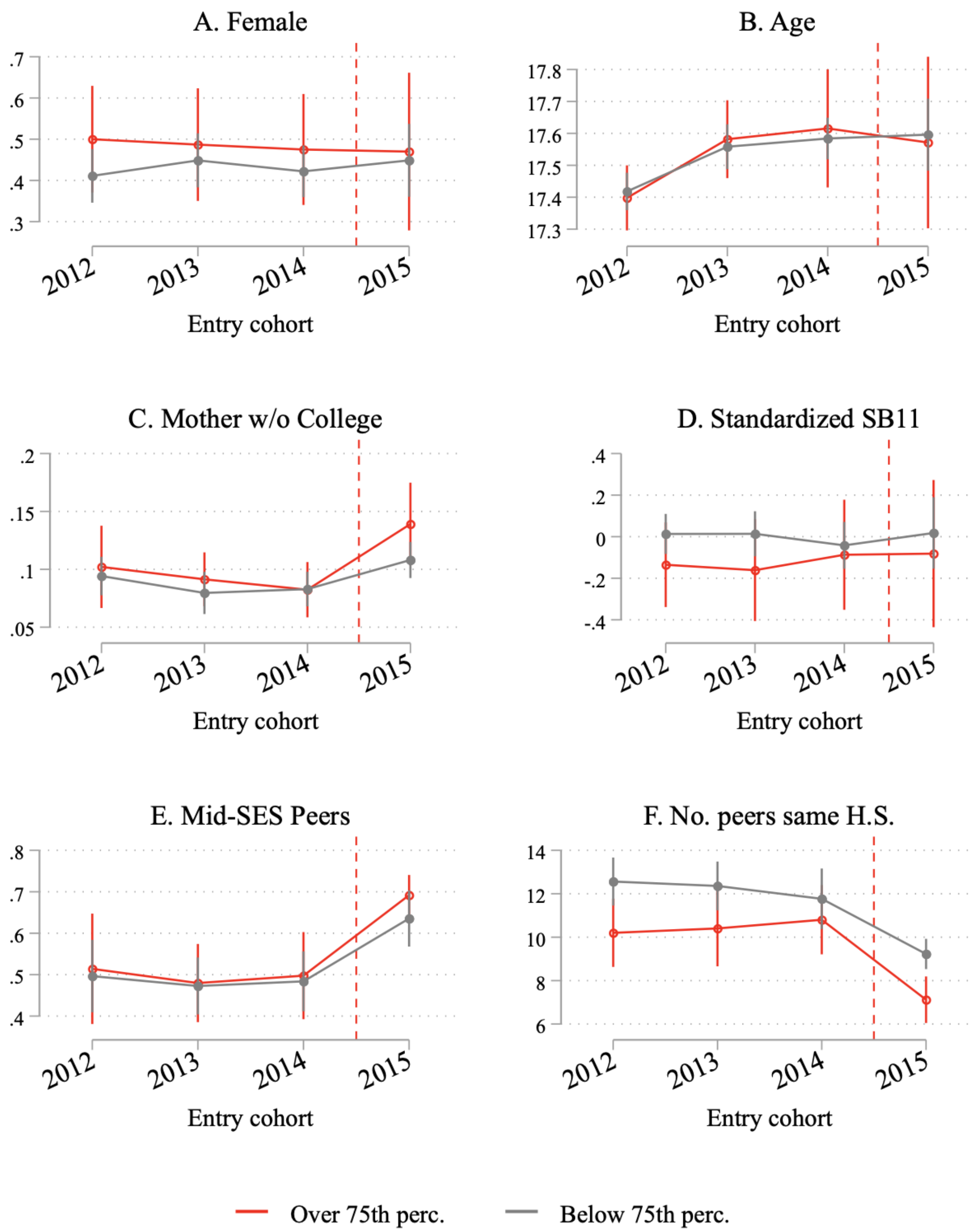
Note: Graphs display point estimates of a cohort dummy variable from an OLS regression with no intercept using student characteristics as the dependent variable. Programs with a low-SES student share above the median value during the SPP period are classified as "Over Median" and others as "Below Median." The 95 percent confidence intervals use clustered standard errors at the program–cohort level. Yearly entry cohorts include both spring and fall, with a red line dividing pre-SPP and post-SPP (2015-1) cohorts.

Figure 12: Pretreatment trends in observed outcomes – Programs with above– and below–75th percentile shares of low-SES peers in 2015



Note: Graphs display point estimates of a cohort dummy variable from an OLS regression with no intercept using student outcomes as the dependent variable. Programs with a low-SES student share above the 75th percentile during the SPP period are classified as "Over 75th perc." and others as "Below 75th perc." The 95 percent confidence intervals use clustered standard errors at the program–cohort level. Yearly entry cohorts include both spring and fall, with a red line dividing pre-SPP and post-SPP (2015-1) cohorts.

Figure 13: Pretreatment trends in observed student sociodemographics – Programs with above– and below–75th percentile shares of low-SES peers in 2015



Note: Graphs display point estimates of a cohort dummy variable from an OLS regression with no intercept using student characteristics as the dependent variable. Programs with a low-SES student share above the 75th percentile during the SPP period are classified as "Over 75th perc." and others as "Below 75th perc." The 95 percent confidence intervals use clustered standard errors at the program–cohort level. Yearly entry cohorts include both spring and fall, with a red line dividing pre-SPP and post-SPP (2015-1) cohorts.

Table 12: Effects of increased exposure to low-SES peers on students' social interactions

(Turnstile–elicited interactions based on a two-second window)

	I. Probability of a Link with a:		II. Number of Links with:			III. % of Links with:
	(1) Low SES	(2) High SES	(3) Low SES	(4) High SES	(5) Any Student	(6) Low SES
A. Continuous Treatment						
Percentage of low-SES peers	0.007***	–0.001	0.026***	–0.029*	–0.003	0.719***
	(0.001)	(0.001)	(0.003)	(0.016)	(0.015)	(0.080)
Mean Increase (18.0 points)	0.126	–0.018	0.468	–0.522	–0.054	12.942
B. 50th Percentile						
$\mathbb{I}$[% of Low–SES Peers > 24%]	0.097***	–0.028	0.420***	–0.592	–0.172	13.841***
	(0.035)	(0.035)	(0.106)	(0.437)	(0.415)	(2.632)
C. 75th Percentile						
$\mathbb{I}$[% of low-SES peers > 36%]	0.111**	–0.037	0.609***	–1.005*	–0.396	17.179***
	(0.044)	(0.048)	(0.111)	(0.555)	(0.540)	(3.886)
Pretreatment Statistics for Outcomes						
Mean	0.157	0.736	0.193	3.892	4.085	4.330
Standard deviation	0.364	0.441	0.494	4.019	4.213	11.511
No. of students	4,027	4,027	4,027	4,027	4,027	3,011
No. of major–cohort groups	93	93	93	93	93	90

Note: A socially interacting pair of students is defined when their IDs are swiped at a turnstile at the same entrance and in the same direction within two seconds or less and at least twice during a semester. The outcomes in Panel I are indicators equal to one when the student has interacted with at least one low- or high-SES peer. Panel II uses the number of peers whom the student has interacted with, and Panel III uses the percentage of low-SES links. Panel A displays the results of estimating Equation 1. Panels B and C display the results based on dummy variables for treatment assignment designating programs with low-SES student shares above the median and 75th percentile, respectively. All regressions control for a female indicator, age in years at the time of entry, SB11 standardized test scores, mother without a college degree, socioeconomic stratum two or three (i.e., intermediate SES), receipt of an SPP loan, and number of high-school peers enrolled in the same cohort as the student. All standard errors are clustered at the program–cohort level. *** p <0.01, ** p <0.05, *, * p<0.1.

Table 13: Effects of increased exposure to low-SES peers on students' social interactions

(Turnstile-elicited interactions based on a three-second window)

	I. Probability of a Link with a:		II. Number of Links with:			III. % of Links with:
	(1) Low SES	(2) High SES	(3) Low SES	(4) High SES	(5) Any Student	(6) Low SES
A. Continuous Treatment						
Percentage of low-SES peers	0.008***	−0.002	0.037***	−0.035*	0.002	0.750***
	(0.001)	(0.001)	(0.003)	(0.018)	(0.018)	(0.066)
Mean Increase (18.0 points)	0.144	−0.036	0.666	−0.630	0.036	13.500
B. 50th Percentile						
$\mathbb{I}$[% of low-SES peers > 24%]	0.109***	−0.034	0.630***	−0.690	−0.060	13.783***
	(0.041)	(0.027)	(0.149)	(0.494)	(0.483)	(2.544)
C. 75th Percentile						
$\mathbb{I}$[% of low-SES peers > 36%]	0.143***	−0.058*	0.940***	−1.239**	−0.298	17.715***
	(0.048)	(0.033)	(0.128)	(0.611)	(0.636)	(3.428)
Pretreatment Statistics for Outcomes						
Mean	0.188	0.770	0.239	4.973	5.212	4.404
Standard deviation	0.391	0.421	0.558	4.922	5.154	11.337
No. of students	4,027	4,027	4,027	4,027	4,027	3,141
No. of major–cohort groups	93	93	93	93	93	91

Note: A socially interacting pair of students is defined when their IDs are swiped at a turnstile at the same entrance and in the same direction within three seconds or less and at least twice during a semester. The outcomes in Panel I are indicators equal to one when the student has interacted with at least one low- or high-SES peer. Panel II uses the number of peers whom the student has interacted with, and Panel III uses the percentage of low-SES links. Panel A displays the results of estimating Equation 1. Panels B and C display the results based on dummy variables for treatment assignment designating programs with low-SES student shares above the median and 75th percentile, respectively. All regressions control for a female indicator, age in years at the time of entry, SB11 standardized test scores, mother without a college degree, socioeconomic stratum two or three (i.e., intermediate SES), receipt of an SPP loan, and number of high-school peers enrolled in the same cohort as the student. All standard errors are clustered at the program–cohort level. *** p <0.01, ** p <0.05, *, * p<0.1.