Is Integration Equal to Inclusion? College Integration and Students' Social Interactions: Evidence from Turnstiles Data

Tatiana Velasco*

November 5th, 2020

[DO NOT CITE]

Abstract

The efforts to desegregate schools by helping low–income students to attend elite institutions have spread around the world. However, the evidence on how students' social networks respond to such policies remains scant. Importantly, there is a concern that school integration benefits may be undermined if social interactions within schools remain segregated. In this paper, I study a natural experiment at an elite university which experienced a sharp and unexpected increase in its enrollment of low–income students. To measure social interactions, I develop a measure based on students' co-movements across campus as recorded by turnstiles guarding all entrances. My findings indicate wealthy students are more likely to interact with low-income ones when exposed to larger shares of them in their majors and classes. However, the size of the effect is small. These findings suggest segregation may persist within colleges, even when large integration policies drastically change student composition.

Keywords: school segregation, social networks, difference-in-differences

*Teachers College, Columbia University. e-mail: tv2225@tc.columbia.edu.

This draft was presented at the 2020 Association of Public Policy Analysis and Managment APPAM - 2020. I am grateful Peter Bergman, Jordan Matsudaira, Tomás Rodriguez-Barraquer, and Román A. Zárate their feedback and comments throughout different stages of this project. In addition, I am grateful for the comments of Maria José Alvarez, Alex Bowers, Sarah Cohodes, Juan David Herreño, Elise Marifian, Viviana Rodriguez, Susha Roy, and the seminar participants at the 2020 Association for Education Finance & Policy (AEFP) Annual conference. I would also like to thank Tito Andrés Gutierrez for his research assistance. All errors are my own.

1 Introduction

The benefits of school integration may be undermined if social interactions within schools remain segregated. For decades, policy makers have seen integration of under–represented students at elite schools as key to foster upward mobility and more positive views towards others from different backgrounds. While this is an argument supported by extensive research (e.g. Boisjoly et al., 2006; Chetty et al., 2020; Fryer, 2011), the evidence also suggest that the benefits of school integration may depend on under–represented students forming networks with privileged ones (Carrell et al., 2013; Marmaros and Sacerdote, 2002; Rao, 2019). However, the extent to which this can be achieved through desegregation policies is unknown. In fact, researchers have documented how the culture at elite colleges prevents the social integration of low–income students, who often lack the social and financial resources to navigate these privileged environments (Armstrong and Hamilton, 2015; Corredor et al., 2019; Jack, 2019). Despite that, efforts to desegregate schools by helping low–income students attend elite institutions have spread, while the evidence on how students social networks respond to such policies remains scant.

In this project I ask: what happens to students' social interactions when a desegregation policy forces socio-economic integration? Are low–income and wealthy students more likely to interact with one another? To study this, I make use of a natural experiment at a large elite college in Colombia which experienced a sharp and unexpected increase in its enrollment of low–income students. To measure social interactions, I assemble novel data based of over a hundred million records of students' movements across campus as recorded by turnstiles guarding all campus 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 social network. I combine these data with student–level records on course enrollment and implement a difference–in–differences approach that exploits the plausibly random variation in the share of low–income peers within degree majors and across entry cohorts.

The question of how students form social networks has a long tradition in the sociology literature, and has recently became a topic of interest in economics. One consistent finding is that students prefer to interact with students like themselves (Baker et al., 2011; Marmaros and Sacerdote, 2006; Mayer and Puller, 2008) – a phenomenon known as *homophily*. While school integration creates opportunities for relations among students with different backgrounds, creating less homophilic social interactions within school settings relies mostly on students' choices (Christakis et al., 2010; Jackson, 2008; Rogers and Jackson, 2007) [1]. Therefore, it is important to understand how students' interactions change as a result to integration efforts.

This is particularly important given that interactions may be a key mechanism that influences academic achievement and other long–term outcomes. For example, Carrell et al. (2013) found that policies designed to foster achievement through changes in the share of advanced peers had harmful effects due to a relative increase in the tendency of advanced students to interact among themselves. While other evidence examining desegregation majors find no effects (Angrist and Lang, 2004; Echenique et al., 2006; Abdulkadiroglu et al., 2014), the research examining how student networks explain academic achievement remains sparse. In contrast, research that examines the role of social capital and attendance to elite college in employment and social mobility outcomes finds that students' social interactions largely explain success after college, even after accounting for academic achievement (Zimmerman, 2019; Michelman et al., 2020).

Importantly, recent evidence has documented the positive effects that socio-economic desegregation policies have on wealthy students' social preferences and behaviors, and has pointed to the importance of student interactions to support these results. Researchers have found that integration led to more pro-social and less discriminatory behaviors of wealthy students (Rao, 2019), and that exposure to low–income peers raised high–income students

[1]Even under perfectly integrated environments, students may perceive benefits from interacting with others with whom they share certain characteristics (Christakis et al., 2010; Currarini et al., 2009). Plus, students may have preferential attachment for people with whom they have friends in common (Rogers and Jackson, 2007).

concerns about fairness as well as their support for progressive taxation (Londoño-Vélez, 2020)[2]. Both results were associated with more diverse social interactions.

A key challenge to quantify the effect of school integration policies on students' social interactions is the lack of data on students' social networks. I overcome this challenge by leveraging on individual records of students' entries and exits to campus. The university I study is guarded by 18 entrances, each of which requires students and staff to tap their University ID in turnstiles to either enter or exit any of these buildings. I use individual–level turnstiles' records from 2016 to 2018 and develop a measure to identify pair-level students' interactions. Specifically, I define a pair of students as interacting when their IDs are tapped at a turnstile in a time window of three seconds or less, and in the same entrance and direction. Importantly, I must observe the same pair of students co-moving through the turnstiles at least twice in the semester. I validate this measure using survey data for a subgroup of students in the Fall 2017 entry cohort. I find robust evidence indicating my turnstile-based interactions definition does capture students' social interactions.

I use my turnstile-based measure of students' social interactions as outcome and I exploit the plausibly exogenous variation in the number of low-income peers introduced by a college desegregation policy in Colombia. In October of 2014, the Colombian government launched *Ser Pilo Paga*, a policy that targeted low-income students with outstanding academic achievement to promote their attendance to high-quality universities in the country. The program consisted of a loan that covered 100 percent of the tuition cost plus a small stipend. The loan was forgave upon completion of the degree. *Ser Pilo Paga* induced an influx in the number of low-income students enrolled at several high-quality private universities in the country (Londoño-Velez et al., 2020). At the university I study, the number of low-income students enrolled tripled in the Spring of 2015. This forced the exposure to

[2] Rao (2019) studied a school desegregation policy in India and examined the effect of socio-economic integration on wealthy students' behaviors, finding that integration led to more pro-social and less discriminatory behaviors by wealthy students. Similarly, Londoño-Vélez (2020) studied the effect of socio–economic diversity on high income students' perceptions of income distribution and redistributive preferences. Like in this proposal, she studied the case of SPP at one university and finds that exposure to low–income peers raised high–income students concerns about fairness as well as their support for progressive taxation.

low-income peers by relatively wealthy students that traditionally attended this university. Importantly, the timing of the policy gave students and university officers little time to change their major and courses compositions in ways driven by their preferences for low-income peers. Thus, I exploit the variation in the share of low-income students enrolled in each major and cohort in a difference-in-difference setting.

My analyses focus on relatively wealthy students and the effects exposure to larger shares of low-income peers had on their social interactions. I classify relatively wealthy students in two groups: high- and middle-income students, and examine their interactions among them and with other low-income peers. This classification is motivated by previous studies that highlight the importance of middle-income students in facilitating the college experience of new low-income ones (Alvarez-Rivadulla, 2019). I define the treatment as the share of low-income students each wealthy student had in their major and entry cohort. I consider 31 majors and 4 entry cohorts, between 2014 and 2015.

My findings indicate relatively wealthy students are more likely to interact with low-income peers when exposed to them either because they are enrolled in their major and entry cohort, or because their presence in first semester classes. Additionally, middle-income students maintain their interactions with other relatively wealthy peers unaltered, but high-income students are less likely to interact with other wealthy peers. These findings may be evidence of middle-income students serving as bridges for indirect interactions between low- and high-income students.

Despite being significant, the size of the effects is quite small. Under the assumption that all wealthy students are exposed to the same average increase in the share of low-income peers - an very optimistic scenario, the effect size is never above the 1 percent. This indicates relatively wealthy students' social networks changed little in the presence of more low-income peers. Thus, the findings suggest segregation may persist within colleges, even after large enrollment of low-income students.

These findings contribute to the literature studying the impacts of school integration.

My work is closely align to that of Londoño-Vélez (2020), who studied the effect of socio-economic diversity at an elite college in Colombia on students' redistribute preferences. In this work, Londoño-Vélez finds positive impacts of exposure to low-income peers on wealthy students preferences - a result that seems to be related to more interactions with low-income peers. My work validates the latter finding, while pointing out that the change in social interactions is rather small.

In similar lines, my work contributes to the research examining the role of peers on academic and non-academic achievement. My results relate to that of Angrist and Lang (2004) and Carrell et al. (2013) who document how social interactions explain the effects found on academic achievement. The evidence in this paper validates their finding of lack of social interactions across groups of peers as a potential mechanism explaining the effects on academic achievement. Similarly, my work is related to that of Zimmerman (2019) and Michelman et al. (2020) who identified the persistence of social interactions across privileged groups as a channel explaining long-term social mobility and success.

The remainder of this paper is organized as follows: In section 2. I present the study context; in section 3. I present the data and variables used for this analysis, along with the procedure used to identify social interactions, in section 4. I present the empirical strategy; in section 5. I discuss the results: in section 6. I discuss robustness checks under alternative definitions of the treatment; and in section 7. I discuss the results and conclude.

2 Policy Context

In this project, I examine the effect of a socio–economic desegregation policy on students' social interactions. Specifically, I study the case of a large private university located in Bogotá, Colombia (from now on *Elite University*[3]), which in 2015 experienced a large and unexpected increase in the number of low–income students enrolled, while keeping the

[3] This is a made-up name. I do not provide the real name of the university I study for confidentiality reasons.

enrollment of relatively wealthy students constant. The increase was driven by *Ser Pilo Paga* (SPP) – a forgivable loan program for high–achieving low–income students who wished to attend a high-quality university. Importantly, the increased enrollment of low–income students varied across the thirty-one degree majors offered at Elite University. In my research design, I focus on relatively wealthy students and compare students from the entry cohorts before and after SPP (2014 vs. 2015). I use the change in the share of low–SES students across degree majors as the treatment. In this section, I explain the context of SPP and Elite University where the natural experiment took place.

Higher education in Colombia is strongly segregated by socio–economic background. High–quality private universities have exceedingly expensive tuition rates relative to average salaries in the country; supply of public university seats is stagnant; and financial aid is scant (Marta Ferreyra et al. 2017). These factors led students to sort across colleges by socio-economic status (Camacho et al., 2017). SPP aimed to combat this segregation by providing low-income students a loan that covered tuition plus a small allowance for attending a high-quality accredited institution[4], which was forgiven upon completion of the degree. Eligibility to SPP required that students were classified as poor under the governments' index of household wealth[5], and scored in the top ten percentile of the national high school exit exam SABER 11[6]. SPP awarded loans for new cohorts of students between 2015 and 2018 benefiting about 40,000 students nationwide. Previous research has found SPP increased diversity at top private universities by making the selection mechanism based more on ability than on income (Londoño-Velez et al., 2020).

The timing of SPP and the admission rules at Elite University set the conditions of the

[4]The high-quality accreditation is granted to higher education institutions by the National Council of Accreditation. It is granted after a detailed review from a panel formed by the Institution, the academic community, and the Council. By 2014, the year of the first round of SPP, 32 universities in Colombia.

[5]This index is known as SISBEN and it is based on a census-based survey targeted to household previously screened as potentially poor. Londoño-Velez et al. (2020) provide more details about how SISBEN was used to screen SPP eligible students.

[6]SABER 11 is a requirement for all students in the country who are about to complete their high school education. The exam is applied twice a year, following the two academic calendar of schools in the country: January – November and August – June.

natural experiment I exploit in my research design. First, admissions to Elite University are open for the Spring and Fall term of each year and are determined by the applicant's score in the SABER 11 standardized test. Students must apply to a degree major[7] and entry cohort for which admission officers had pre–determined a specific SABER 11 weighting formula[8] and cutoff score. Second, SPP was widely unexpected by students and higher education institutions. SPP was launched in October of 2014 and only students who had taken that October's test were eligible. Candidates had to apply for enrollment in the following Spring of 2015, for which 10,000 forgivable loans were offered. Thus, students who traditionally applied to Elite University had very little time to change their application portfolio and university officers could not adjust the admission criteria to limit the influx of admitted and eventually enrolled students. As a result, the number of middle– and high–income students enrolled in 2015 remained similar to that from 2014, but the number of low-income students increased significantly.

Figure 1 depicts the first–term enrollment trends by socio–economic status (SES) at Elite University. Between 2012 and 2014, less than 150 first–term students came from low–SES backgrounds. Once the first cohort of SPP beneficiaries enrolled, the number of low–SES students tripled to 541, while the number of students from other socio–economic backgrounds remained almost the same. Figure 2 compares the shares of low–SES students across majors, before and after SPP. The variation in the change of the share of low–SES students is important. majors such as Business and Music experienced virtually no change in the share of low–SES students, while others like Philosophy or Psychology experienced a notable increase. In Figure 5, I document that observed characteristics like demographics or dropout rates that predict future interactions of relatively wealthy students do not change significantly across entry cohorts, which provides support for the comparison of different cohorts of first–term

[7]As opposed to the U.S., applicants to higher education must apply to a major for degree as well to as a college.

[8]The SABER 11 is made of five modules which are given different weights depending on the major of application. For example, for admission to engineering majors, quantitative reasoning is a assigned a higher weight than the social sciences module

students.

3 Data

The data for this project comes from two sources: administrative records from Elite University, and detailed records from turnstiles located in each of the 18 access points to Elite University campus.

Elite University administrative records. I use records from all students enrolled at Elite University between 2012 and 2018 which contained student-course level data on student characteristics (i.e. sex, age, mother's education, High School ID), SABER 11 standardized test score, SPP recipient status, selected degree major, entry cohort and term of enrollment. For each semester, I observe each of the courses in which the student is enrolled and their course GPA. More importantly, I observe the student's household social strata indicator. This indicator has six categories which are used to provide homes with subsidies in utility bills. Plus, it is also widely known in the country as a proxy of social status. I use the household social strata to classify students in three socio-economic status (SES): low-, middle-, and high-SES. Low-SES are students from strata one and two, middle-SES are students from strata three and four, and high-SES are students from strata five and six. Students benefiting from SPP mostly fall in the low-SES category. As depicted in Figure 1, the majority of students at Elite University are classified as high- and middle-SES.

Turnstile records. I use records on student access and exits to Elite University campus to identify students' social interactions. Elite University campus is guarded by turnstiles located at the 18 entrances to main buildings and campus areas. In order to enter or exit through any of these entrances students and university staff must tap their University ID. Security officers at Elite University provided me individual-level records of University ID taps on the turnstiles from February 1st, 2016 to November 1st, 2019. These records include student ID number, entrance, action (IN or OUT of campus), and the date, hour, minute and second of the tap. Figure 7 in the Appendix displays a heat map of the average frequency of student ID taps at each of the 18 entrances to campus by 20 minutes blocks and for each academic term since the Spring of 2016. Yellow cells 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 change as well as during lunch hours.

Students' interactions: I define a pair of students as interacting –a link– when their IDs are tapped at a turnstile in a time window of three seconds or less, in the same entrance and direction (either entering or exiting campus), and when I observed the same pair of IDs co-moving at least twice in a semester.

Validation of student interactions definition. I define a time window and frequency thresholds by comparing turnstile-elicited with survey-elicited links among first-term undergraduate students of Economics from the fall of 2017 cohort[9]. The survey I analyzed inquired about two types of links: friendships and acquaintances. Table 1 shows the results of the comparison. The time windows tested in Table 1 were selected based on in-person observations to different entrances[10]. I select a time window and a frequency criterion by minimizing the sum of the type I and type II measurement errors (that is, the number of unmatched survey-elicited links over the total number of survey links, and the number of unmatched turnstile-elicited links over the total number of survey links).

To illustrate how to interpret the results in Table 1, I ask the reader to focus on the time window of three seconds and the acquaintances survey links. The numbers in bold indicate the combinations of time-windows and frequencies that minimize the sum of type I and II errors, for each type of link. Thus, the frequency with which I should observe two

[9]The survey was conducted online between December 7, 2017, and January 5, 2018, and elicited the network among 110 economics students from the 2017 fall cohort. The survey was conducted using Qualtrics. Students who completed the survey received a free lunch voucher for a recognized chain restaurant of the campus area. Cárdenas et al. (2019) provide a detail description of the survey. I am very grateful to Professor Tomás Rodríguez-Barraquer for providing me access to these data.

[10]The observations of entrances to campus were conducted between August 26th and 30th of 2019. Because there are multiple turnstiles at each entrance, students walking together can essentially tap their IDs simultaneously using different scanners.

student IDs tapped on a turnstile entrance so that it resembles an acquaintances link should be minimum twice in the semester. Under that rule, the likelihood of Type II error or false positives - i.e. the likelihood of defining a pair of students as linked when according to the survey they are not, is 11 percent. Conversely, the likelihood of a Type I error or false negative - i.e., the likelihood of not identifying a pair of students as acquaintances when according to the survey they are, is 62 percent. While a five-seconds and three times in the term criteria would yield a lower sum of errors, it would do so by leaving one student from the 110 in the sample without turnstile-based links information –an omission I want to avoid. Notice that the acquaintances criteria has a lower threshold in terms of the frequency of the co-movements in the semester than the friendship criteria. I chose to use the acquaintances instead of the friendship criteria because it allows me to identify social interactions that students did not identify as friendships in the first term of college, but that may eventually evolve as such.

The results in Table 1 indicate that under the baseline definition, it is highly likely that the turnstiles-elicited links capture survey-like links. However, an important share of survey links may not be captured by the turnstiles. This is an issue to the extent that those I do capture are not representative of the survey-elicited links. To assess this, I compare whether turnstile-elicited links plausibly reflect survey-elicited network characteristics. Results are displayed in Figure 3. The goal of this exercise is to estimate how far from random are the turnstile-elicited links' characteristics, and how close the average characteristics of the links are to those of the survey-elicited links. The computation proceeds as follows: I use the acquaintances minimizing criteria from Table 1 for each of the time windows and randomly assign the number of turnstile-elicited links under that criteria to the 110 students in the sample. Then, I compute the average of the following network individual attributes: age difference, number of courses students are taking together, GPA difference, degree or number of links, and local clustering. I conduct this procedure 1000 times and plot the distribution of the characteristics. I include the average value I observe for the turnstile- and survey-elicited links with its 95 percent confidence interval. I find statistically significant support indicating turnstile-elicited network characteristics resemble closely those of the friendship and acquaintances networks elicited by the survey, and are not the result of random links formation

Sample. My analytic sample consist of all the first–term students in the entry cohorts before and after SPP (i.e. Fall and Spring of 2014 and 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 degree major. For example, I match students in the entry cohort of Spring of 2014 with their interactions as captured by the turnstiles during the Fall of 2016 and the Spring of 2017. I merge administrative records and pairwise–level students' interactions data using the student ID number which is available in both data sources. My final sample consist of 5,955 students across 31 degree majors and 4 entry cohorts. This sample captures the universe of students enrolled in these majors and cohorts except for two majors (Government, and the Directed Studies major) which started after SPP.

Table 2 provides descriptive statistics of the sample. I divide the sample between non-low-SES students - that is, middle-SES and high-SES students, and low-SES students, before and after the implementation of SPP (i.e., 2014 vs. 2015). About half of non-low-SES students are middle income students according to the household strata indicator. In 2015, 81 percent of low-SES students at Elite University were SPP recipients. In both 2014 and 2015 cohorts, non-low-SES students are more likely to be females, are slightly older and with mothers more educated than low-SES students. Also, non-low-SES students have higher SB11 test scores than low-SES ones, and the gap increases for the 2015 cohort. Among students from the 2014 cohort, non-low-SES students had an average standardized test score of -0.05 while low-SES had a -0.13. In 2015, the scores where 0.03 and -0.49 respectively. Importantly, non-low-SES students have on average more peers from high school enrolling at Elite University in their same cohort than low-SES students (11.54 and 3.17 high school peers respectively). This difference increases in 2015. For this entry cohort, non-low-SES students enrolled with an average o 11.73 high school peers, while low-SES students enrolled with 1.98 peers. High school social networks are important predictors of college and employment outcomes (Zimmerman, 2019; Michelman et al., 2020 . Thus, I control for indicators of high school peers throughout my analyses.

Importantly, Table 2 presents descriptive statistics of students' networks characteristics as captured by the turnstiles-elicited links. In 2014, wealthy students have 5.21 links, while low-SES students have 4.94 links. Both groups have similar local clustering (i.e., share of friends who are friends) (0.29 and 0.26, respectively), and similar values of their Bonacich power centrality (0.47 and 0.48) -a measure of popularity of each student in the overall network. In 2015, the gap between wealthy and low-SES students increases, indicating wealthy students are overall better connected than low-income ones.

4 Methods

In this paper I ask, are relatively wealthy students more likely to interact with low–income ones? As a research design, I use a difference–in–differences approach that compares entry cohorts of students before and after SPP and exploits the change in the share of low–SES students across majors and during the first semester in college. I examine social interactions after six semesters in college. I constrain my sample to relatively wealthy students which I define as those from a middle– and high–SES backgrounds and examine the interactions among them, and with low–SES students. I distinguish between high- and middle-SES students, because I care to understand the role that middle-income students have in facilitating social interactions for low-income ones.

$$Pr(H_{ij} = 1|\boldsymbol{X}) = \alpha_R R_i^{pc} + \alpha_X X_i + \alpha_{|X|}|X_i - X_j| + \delta_p + \delta_c + \epsilon_{ij} \tag{1}$$

I use Equation 1 to estimate the change in the probability that a student interacts with a peer from a different SES background. Define student i and j as two students who started college in the same major and entry cohort. H_{ij} is an indicator equal to 1 when i and j are socially interacting according to my turnstiles-based definition and when they have different SES background (i.e. a SES-heterophilic link), and to zero otherwise. R_i^{pc} represents the share of low-SES students i had among their cohort and major peers in the first term of college and constitutes my treatment variable. X_i captures the student characteristics that affect the interaction probability (gender, age, mother's education, SABER 11 score, GPA, number of credits taken in the first term), $|X_i - X_j|$ measures the degree of homophily of the link and it is defined as the distance between student i and student j characteristics[11]. δ_p is a major indicator which captures time-invariant major characteristics, and δ_c is an entry cohort indicator capturing cohort–level unobserved characteristics which are common across all majors. My estimation of equation 1 follows the approach suggested by Angrist and Pischke (2009) and uses a Linear Probability Model (LPM) estimation with clustered standard errors at the cohort-major level. My estimator of interest is α_R, which under the LPM estimation I interpret as the change in probability points of a SES-heterophilic interaction driven by a unit change in the share of low–SES peers in student i major and cohort.

As explained in Section 2, identification of α_R is ensured by the timing of SPP, which gave little time to traditional students at Elite University to adjust their application portfolio in ways driven by preferences for low-SES students and unobserved by the researcher. Figures 1 and 4 provide evidence in this regard. Changes in enrollment preferences of wealthy students would reflect as changes in the overall numbers of students enrolled in 2015 compared to prior cohorts. In Figure 1, I document how the trends in enrollment numbers of middle- and high-SES students at Elite University remained stable with the implementation of SPP in 2015. In Figure 4, I do the same but with enrollment at each of the 31 majors in my sample.

[11] As homophilic characteristics I use: Age difference, same gender, both immigrants, SABER 11 score difference, both from the same high school.

In both cases, I do not observe a dramatic change in the enrollment of relatively wealthy students in 2015, compared to prior years.

Importantly, since I compare different students across cohorts and majors, student characteristics related to their socialization preferences could change due to SPP, even if the number of students enrolled does not. I provide evidence that this is not the case in Figure 5. In this figure, I present point estimates of the entry-cohort indicator in an ordinary least square regression that controls for major dummies. I display the results by SES group. I do not find evidence suggesting the characteristics of middle- and high-SES students changed significantly in 2015, once SPP is implemented.

One concern about this identification strategy is that the estimates of α_R are also the result of the change in group composition induced by SPP. Intuitively, define the likelihood as the rate between potential interactions and actual ones. If the number of interactions remains the same, but the number of possible interactions increases, then the estimated probability would show a reduction. To discern between these two responses I develop a benchmark measure of how interactions would respond to changes in the share of low-SES students if social interactions were formed completely at random among students from the same major and cohort. I use a Monte Carlo simulation procedure to compute this benchmark. For each major-cohort pool of students, I take the number of interactions observed and re-assign them randomly, drawing from a binomial distribution. Then, I derive H_{ij}, estimate equation 1 and obtain an estimate of α_R and its t-statistic. I conduct this procedure 500 times. Then, I use the Monte Carlo estimates of α_R to compute 95 percent confidence intervals. This interval indicates what the value of α_R would be if the number of social interactions I observe were formed completely at random, among the students at each major-cohort group. Thus, deviations from this benchmark would be explained by students' social interaction preferences, and not by changes in sample size. The average t-statistic indicates whether that relation would be statistically significant in expectation.

In the following sections, I present the results of this estimation procedure, and subsequent robustness checks and analyses.

5 Results

Table 3 displays the results of estimating Equation 1 on the sample of High-SES students. The first two columns show interactions with other relatively wealthy students, while the last one shows interactions with low-SES ones. That is, in column 1, I display the results using as outcome an indicator equal to one when the interaction is with another high-SES student; in column 2, I use an indicator equal to one when the interaction is with a middle-SES student; and in column 3, I use an indicator equal to one when the interaction is with a low-SES student. I present the results of the Monte Carlo simulation in the next row. I display 95 percent confidence intervals of what the point estimate for the effect of R_i^{pc} would be if the effect was fully explained by changes in sample size and composition and therefore random. Importantly, I interpret the point estimates as the change in the likelihood of interaction and how far from random that change is.

The increase in the number of low-SES students made high-SES students less likely to interact with other relatively wealthy students and more likely to interact with low-SES ones. The point estimates for the effect of R_i^{pc} indicate that a unit increase in the share of low-SES peers in the student's major and cohort would significantly decrease the likelihood of interaction with other high-SES student in 7.7 percentage points, and with middle-SES peers in 2.4 percentage points. Conversely, a unit increase in the share of low-SES peers would increase the likelihood of interaction with them in 2.4 percentage points.

Importantly, the Monte Carlo simulations suggest these findings are not likely to be driven by pure changes in the composition of groups. With the exception of the estimate for the interactions among high-SES students - which fall only 0.2 percentage points below the benchmark, these estimates are far from those that would be obtained if interactions among students were being formed completely at random.

Table 4 displays the results of estimating Equation 1 for middle-SES students. As I did for the high-SES group, in column 1, I display the results for interactions among middle-SES students; in column 2, I display the results for interactions with high-SES students; and in column 3, I present the results for interactions with low-SES ones. Likewise, I present the results of the benchmark Monte Carlo simulation in the next row.

The results in Table 4 indicate more low-SES students did not affect the likelihood of interaction of middle-SES students with other wealthy peers. However, the increase did have a positive impact on the likelihood of interaction with low-SES ones. For the latter effect, the estimate of R_i^{pc} suggest a unit increase in the share of low-SES peers increased the likelihood of interaction between middle- and low-SES students in 4.5 percentage points. In this case, the point estimates are far from what they would be if the effects were driven by a composition effect as indicated by the Monte Carlo benchmark.

Overall, these results provide suggestive evidence indicating relatively wealthy students did form social interactions with the incoming low-SES students. Additionally, high-SES students reduced their likelihood of interaction with other wealthy students while middle-SES students kept this likelihood unchanged. These results do not seem to be driven by changes in the size of the cohorts induced by SPP. Importantly, these findings are robust to the alternative definitions of the outcome variable discussed in Section 3. Estimations under the alternative definition of the outcomes are presented in Appendix 7.

However, the size of the effects is quite small. In terms of pre-SPP standard deviations of the outcomes and treatment variables, a one standard deviation increase in the share of low-SES students in the major and entry cohort would increase the likelihood of interaction with low-SES students in 0.012 and 0.018 standard deviations for high- and middle-SES students, respectively. Since the number of low-SES students before SPP was so small, a more realistic estimate would be to use the increase in the share of low-SES students due to SPP. Specifically, in the Spring of 2015, high-SES and middle-SES students were on average exposed to 15 and 17 percentage points more low-SES students in their program and cohorts, respectively. Thus, the estimated average increase in the probability of interaction with low-SES peers for these groups are 0.4 and 0.8 percentage points. The latter estimates are particularly small, specially taking into account that the number of low-SES students in the Spring of 2015 tripled with respect to prior periods.

6 Robustness Checks

One important concern with the results this far is that exposure to low-SES students may be different even across students within the same major and cohort. Such a situation is likely to happen at Elite University as certain majors may require the same core courses in the first semester of courses in the curriculum. For example, Economics, Business, and Engineering all share Differential Calculus as a core first-semester course. Therefore, Differential Calculus has multiple sections to accommodate all students. Thus, a student enrolled in Business - a major with little changes in the number of low-SES students enrolled, may still be exposed to low-SES students through the courses' sections they enrolled in. To capture this, I construct an alternative definition of the treatment. Specifically, I define the share of low-SES peers of student i –R_i– as the number of low-SES classmates student i had throughout all their first-term course sections over the total number of students enrolled across those course sections. I use R_i in the same difference-in-differences framework depicted in Equation 1. Hence, the identification assumption is that student selection of their first-term courses portfolio is not correlated with their preferences for social interactions in ways unobserved by the researcher, and unaccounted after controlling for common shocks within majors and within cohorts (i.e. fixed effects by major and by cohort).

Using R_i as treatment may carry identification issues if wealthy students i) manipulate their first-term courses portfolio, and ii) self-select in specific sections within courses, in ways driven by their preferences for low-SES students. I do not find evidence suggesting either is the case. To test these mechanisms, I examine the number of high- and middle-SES students enrolled by course and by course-section. Specifically, I compute the number of first-term students enrolled in each course and in each section by SES group, and I estimate a regression using these numbers as outcome and conditioning by entry cohort dummies and fixed effects by course name[12]. Figure 6 displays the entry cohort dummies estimates. Panel A. documents changes in enrollment by course and SES group, and Panel B. documents changes within courses' sections. If wealthy students were segregating in different courses or sections to supposedly avoid low-SES students, I should observe changes in these variations from 2014 to 2015. However, the variation in students enrollment by SES group is not statistically different between these periods. This is consistent with the timeline of the policy, which limited the ability of students and university officers to alter behaviors and policies in ways induced by peers preferences.

The results using the share of low-SES peers as treatment are presented in Table 5 and 6. Table 5 displays the estimates for high-SES students, and Table 6 displays the estimates for middle-SES students. As in the latter section, column 1 present the estimated effects for interactions with students in the same SES group, column 2 the effects for interactions with other relatively wealthy students, and column 3 the results for interactions with low-SES students. Each table also displays the results of the Monte Carlo simulation, which provides a benchmark value for what the coefficient of the effect would be if social interactions were formed fully at random, and therefore were driven purely by composition effects.

In the case of high-SES students, exposure to larger shares of low-SES peers in their classes made them less likely to interact among themselves, but did not affect the likelihood of interactions with other less wealthy middle- or low-SES peers. Specifically, a unit increase in the share of low-SES peers reduced the likelihood of interaction among high-SES students in 5.3 percentage points (significant at the 90 percent of confidence). These findings are not likely to be driven by changes in the sample size, as they fall well outside the confidence interval of the Monte Carlo simulated estimate.

[12] I use information on 769 courses taught at least once between Fall of 2013 and Fall of 2015.

Second, I analyze the results for middle-SES students. In this case, an increase in the share of low-SES peers in middle-SES students' first-term courses leads to no changes in the likelihood of interaction with other relatively wealthy students, but does increase the likelihood of interaction with low-SES ones in 6.6 percentage points. Importantly, the estimated point estimate of the effect on interactions among middle-SES students is close to the estimated benchmark. The latter suggest interactions among this group are closed to be as–random. For the remaining two effects, the estimates do fall far off the confidence interval benchmark, which indicates these effects are not driven by changes in cohort sizes.

In terms of pre-SPP standard deviations, the point estimates under this treatment definition are also small. Specifically, a one standard deviation increase in the share of low-SES peers on first semester classes leads to a reduction in the likelihood of interaction among high-SES students of 0.004 standard deviations, and an increase in the probability of interaction between middle- and low-SES students of 0.015 standard deviations. Again, a more realistic assessment would be to compare the estimates with the size of the policy – that is, the average increase in the share of low-SES peers in the cohort of Spring 2015. That increase was 11 percentage points among middle-SES students. Hence, the average increase in the likelihood of interaction between middle and low-SES students due to more low-SES peers in the classroom is on average 0.7 percentage points.

7 Discussion and conclusions

Governments and policy makers have invested large efforts and resources in diversifying schools. These policies are motivated by the expectation that school diversity will foster academic achievement and upward mobility of disadvantaged students, while promoting more inclusive behaviors and positive views towards others among the students traditionally attending elitist schools. Notoriously, researchers have found suggestive evidence indicating these effects can be achieved through socio-economic integration policies (Rao, 2019; Chetty et al., 2020). However, there is little evidence documenting the role that students' social interactions have on explaining these outcomes. Understanding how social interactions respond to integration policies is important because socialization preferences are a likely mechanism driving the positive effects in long-term outcomes. In fact, prior research has found students may respond to policies aiming to optimize peer effects by segregating their interactions to other students similar to them, leading to unintended detrimental effects (Carrell et al., 2013). Despite the importance of students' social responses to integration policies, there is little evidence documenting this relation.

This paper addresses this gap in by studying the effect that a college integration policy had on wealthy students' social interactions. Specifically, I study the case of a large private university in Colombia (*Elite University*) which experienced a sharp increase in the number of low-income students enrolled in the entry cohort of Spring 2015. The influx in the number of low-income students was driven by SPP, a government program that granted low–income students forgivable loans to attend high quality universities in the country. At Elite University, SPP tripled the number of low-income students enrolled, while keeping the numbers of relatively wealthy students fixed. Importantly, the policy did not alter the characteristics of the students who traditionally enrolled at Elite University. Thus, I exploit the increase in the number of students from a low socio-economic background in a difference-in-differences approach that uses the plausible random variation in the share of low-income students within majors and across entry cohorts.

A key limitation to study students' social interactions within schools is the lack of data documenting individuals' social networks. I address this limitation by leveraging on a novel measure of social interactions based on students' co-movements across Elite University campus. I validate this measure using a survey of students' social networks. I document how my turnstiles-based measure of social interactions likely represents students' friendships. To conduct my analysis, I divide the sample of relatively wealthy students in two groups: high- and middle- socio-economic background (SES) and examine each of these groups responses to the increase the number of low-SES peers. I consider two types of exposition to low-SES peers: first, peers in the same major and entry cohort and second, peers in first semester courses. While the first measure documents the effects of peers in the same major and cohort, the second documents the effects of exposition through courses shared with students at another majors.

My findings indicate relatively wealthy students had a positive response to the increase in the number of low-income students among their major and cohort peers. Both income groups were more likely to interact with low-SES students when exposed to larger numbers of them, either in their major or through their first semester courses. However, the size of the effect is small. If students were exposed to the average increase in the share of low-SES peers driven by SPP -which ranges between 8 to 17 percent, the estimated effect is still below the 1 percentage point. The evidence is less conclusive in terms of the effect more low-income peers had on the likelihood of interactions among wealthy students. My results suggest high-SES students are more likely to reduce their likelihood of interaction with other relatively wealthy peers when exposed to low-SES students. Middle-SES students - a group that prior research suggest could be pivotal in facilitating integration, do not change their interactions with other relatively wealthy peers.

A remaining question is about the extent to which social interactions and its responses to integration policies affect other academic and long-term outcomes. Arguably, it is not necessary to impact students' interactions, as long as the lack of change is not detrimental to other academic and long-term outcomes. Next steps in this research involve addressing the extent to which changes in social networks matter in explaining academic and labor market outcomes.

References

Abdulkadiroglu, A., Angrist, J. D., and Pathak, P. (2014). The Elite Illusion: Achievement Effects at Boston and New York Exam Schools. *Econometrica*, 82(1):137–196.

Alvarez-Rivadulla, M. J. (2019). ¿"Los becados con los becados y los ricos con los ricos"? Interacciones entre clases sociales distintas en una universidad de elite. *Destacados*, 59:50–67.

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

Angrist, J. D. and Pischke, J. S. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press, Princeton, NJ., first edit edition.

Armstrong, E. A. and Hamilton, L. T. (2015). *Paying for the Party: How College Mantains Inequality*. Harvard University Press.

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

Boisjoly, J., Duncan, G. J., Kremer, M., Levy, D. M., and Eccles, J. (2006). Empathy or antipathy? The impact of diversity. *American Economic Review*, 96(5):1890–1905.

Camacho, A., Messina, J., and Uribe, J. P. (2017). The Expansion of Higher Education in Colombia: Bad Students or Bad Programs?

Cárdenas, J. C., Okulicz, D., Pietrobon, D., and Rodríguez, T. (2019). Propensity to Trust and Network Formation.

Carrell, S. E., Sacerdote, B., and West, J. E. (2013). From Natural Variation to Optimal Policy? The Importance of Endogenous Peer Group Formation. *Econometrica*, 81(3):855–882.

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

Christakis, N., Fowler, J., Imbens, G., and Kalyanaraman, K. (2010). An Empirical Model for Strategic Network Formation. *National Bureau of Economic Research*.

Corredor, J., Álvarez-Rivadulla, M. J., and Maldonado-Carreño, C. (2019). Good will hunting: social integration of students receiving forgivable loans for college education in contexts of high inequality. *Studies in Higher Education*.

Currarini, B. S., Jackson, M. O., and Pin, P. (2009). An Economic Model of Friendship: Homophily, Minorities, and Segregation. *Econometrica*, 77(4):1003–1045.

Echenique, F., Fryer, R. G., and Kaufman, A. (2006). Is school segregation good or bad? *American Economic Review*, 96(2):265–269.

Fryer, R. G. (2011). The Importance of Segregation, Discrimination, Peer Dynamics, and Identity in Explaining Trends in the Racial Achievement Gap. In *Handbook of Social Economics*, chapter 21. Elsevier B.V.

Jack, A. A. (2019). *The Priviledge Poor: How Elite College Are Failing Disadvantaged Students*. Harvard University Press.

Jackson, M. O. (2008). *Social and Economic Networks*. Princeton University Press, Princeton, NJ.

Londoño-Vélez, J. (2020). The Impact of Diversity on Distributive Perceptions and Preferences for Redistribution.

Londoño-Velez, J., Rodriguez, C., and Sanchez, F. (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.

Marmaros, D. and Sacerdote, B. (2002). Peer and social networks in job search. *European Economic Review*, 46:870–879.

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

Marta Ferreyra, M., Avitabile, C., Botero Álvarez, J., Haimovich Paz, F., and Urzúa, S. (2017). *At a Crossroads Higher Education in Latin America and the Caribbean Human Development*. The Worl Bank, Washington D.C.

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

Michelman, V., Price, J., and Zimmerman, S. D. (2020). The Distribution of and Returns to Social Success at Elite Universities.

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

Rogers, B. W. and Jackson, M. O. (2007). Meeting Strangers and Friends of Friends: How Random Are Social Networks? *American Economic Review*, 97(3):890–915.

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

Tables

Table 1: Survey– and Turnstile–elicited links comparison

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 links	868	368	235	180	148	1209	509	314	251	198	1906	898	552	401	315
2. Are friends															
Links			*505*					*505*					*505*		
Survey & Turnstiles															
Matched	342	256	201	165	140	389	305	248	215	179	433	368	337	295	263
False Negatives (Type I)	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 II)	1.04	0.22	0.07	0.03	0.02	1.62	0.40	0.13	0.07	0.04	2.92	1.05	0.43	0.21	0.10
Sum	1.36	0.71	**0.67**	0.70	0.74	1.85	0.80	**0.64**	0.65	0.68	3.06	1.32	0.76	0.63	**0.58**
3. Acquaintances															
Links			*1033*					*1033*					*1033*		
Survey & Turnstiles															
Matched	497	311	219	174	144	606	391	284	235	191	734	537	425	348	293
False Negatives (Type I)	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 II)	0.36	0.06	0.02	0.01	0.00	0.58	0.11	0.03	0.02	0.01	1.13	0.35	0.12	0.05	0.02
Sum	*0.88*	***0.75***	*0.80*	*0.84*	*0.86*	*1.00*	***0.74***	*0.75*	*0.79*	*0.82*	*1.42*	*0.83*	***0.71***	*0.71*	*0.74*

Notes: N students = 110. Number of links possible (N*(N-1))/2 = 5995. Survey sample consist of economics undergrads from the August 2017 cohort. 113 students surveyed. One student did not report information and two do not show enrolled as of 2017-2. The survey asked each student who among the 113 students were an Acquaintance, and among those, who was considered a friend. Type I error rate is the share of links in survey that were not found in turnstiles-based links. Type II error is the links in turnstiles that were not matched with the links in survey, over the total links in survey

Table 2: Descriptive statistics of the sample

	Before		After	
	Non low-SES	Low-SES	Non low-SES	Low-SES
	Mean	Mean	Mean	Mean
R_i^{pc}	0.04	0.06	0.14	0.29
R_i	0.04	0.07	0.11	0.24
Prop. Of middle–SES	0.49	0.00	0.50	0.00
Student characteristics				
Female	0.43	0.35	0.46	0.41
Age	17.59	17.25	17.59	17.14
Mother w/high school	0.08	0.25	0.09	0.40
SB11 score	-0.05	-0.13	0.03	-0.49
Tot. Credits in first term	15.64	15.21	15.92	14.25
SPP recipient	0.00	0.00	0.05	0.81
No. Of peers from H.S. in same cohort	11.54	3.17	11.73	1.98
Dropout	0.15	0.22	0.12	0.17
Cum GPA in 7th term	3.86	3.82	3.91	3.71
Cum. Credits in 7th term	115.80	113.50	117.88	110.69
Turnstiles characteristics				
Tot. ID taps in 6th and 7th terms	1507.36	1617.42	1314.20	1310.68
No. Of Links	5.21	4.94	5.53	4.60
Ego-network characteristics				
Local Clustering	0.29	0.28	0.29	0.25
Bonacich power centrality	0.47	0.48	0.54	0.41
Age Difference	0.77	0.92	0.79	0.88
Same Gender	0.65	0.71	0.64	0.63
Ave. No. courses w/links 6th and 7th	2.66	2.35	2.69	2.93
SB11 difference	0.94	1.07	0.99	0.93
No. Of majors	31	31	31	31
No. Of students	2669	139	2609	538
Individuals without links	600	40	541	151

Notes: This table displays the estimates from Equation 1. All regressions include the controls depicted in equation 1 and described in the empirical analysis section. I address missing values by including an indicator eual to 1 if any of the covariates contains a missing value for the dyad ij. Each observation represents a dyad which results from all possible combinations of pairs of students from the same major and cohort, when one of the pair is a high–SES student.

Table 3: Effect of major-cohort low-income peers on High–SES students' interactions

	Interactions of High SES students with:		
	(1) High SES	(2) Middle SES	(3) Low SES
R_i pc	-0.077***	-0.024**	0.024***
	(0.023)	(0.012)	(0.004)
Monte Carlo Benchmark			
95% C.I.	[-0.0748, -0.0746]	[-0.0582, -0.0581]	[0.0679, 0.0680]
mean(t-test)	8.53	7.89	21.22
Pre-treatment statistics			
sd(H_ij)	0.19	0.15	0.05
sd(R_i pc)	0.026	0.026	0.026
No. Students	2,674	2,674	2,674
Num. Obs.	224,690	224,690	224,690

Notes: This table displays the estimates from Equation 1. R_i^{pc} is the proportion of low-SES students in the major and entry cohort as treatment. All regressions include the controls depicted in Equation 1 and described in the empirical analysis section. I address missing values by including an indicator equal to 1 if any of the covariates contains a missing value for the link ij. Each observation represents a link which results from all possible combinations of pairs of students from the same major and cohort, when one of the pair is a middle–SES student. Monte Carlo Confidence Intervals are estimated as described in Section 4. All standard errors are clustered at the major-cohort level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 4: Effect of major-cohort low-income peers on Middle–SES students' interactions

	Interactions of Middle SES students with:		
	(1) Middle SES	(2) High SES	(3) Low SES
R_i pc	0.008	-0.020	0.045***
	(0.017)	(0.015)	(0.016)
Monte Carlo Benchmark			
95% C.I.	[-0.0514, -0.0513]	[-0.0476, -0.0475]	[0.0702, 0.0703]
	7.06	7.04	19.29
Pre-treatment statistics			
sd(H_ij)	0.20	0.16	0.07
sd(R_i pc)	0.028	0.028	0.028
No. Students	2,604	2,604	2,604
Num. Obs.	199,315	199,315	199,315

Notes: This table displays the estimates from Equation 1. R_i^{pc} is the proportion of low-SES students in the major and entry cohort as treatment. All regressions include the controls depicted in Equation 1 and described in the empirical analysis section. I address missing values by including an indicator equal to 1 if any of the covariates contains a missing value for the link ij. Each observation represents a link which results from all possible combinations of pairs of students from the same major and cohort, when one of the pair is a middle–SES student. Monte Carlo Confidence Intervals are estimated as described in Section 4. All standard errors are clustered at the major-cohort level. * p < 0.10, ** p < 0.05, *** p < 0.01.

Table 5: Effect of low-income classroom peers on High–SES students interactions

	Interactions of High SES students with:		
	(1) High SES	(2) Middle SES	(3) Low SES
R_i	-0.053*	0.018	0.013
	(0.028)	(0.018)	(0.008)
Monte Carlo Benchmark			
95% C.I.	[-0.0442, -0.0440]	[-0.0302, -0.0300]	[0.0447, 0.0448]
mean(t-test)	3.55	2.88	9.85
Pre-treatment statistics			
sd(H_ij)	0.19	0.15	0.05
sd(R_i pc)	0.014	0.014	0.014
No. Students	2,674	2,674	2,674
Num. Obs.	224,690	224,690	224,690

Notes: This table displays the estimates from Equation 1. R_i is the proportion of low-SES students in the major and entry cohort as treatment. All regressions include the controls depicted in Equation 1 and described in the empirical analysis section. I address missing values by including an indicator equal to 1 if any of the covariates contains a missing value for the link ij. Each observation represents a link which results from all possible combinations of pairs of students from the same major and cohort, when one of the pair is a middle–SES student. Monte Carlo Confidence Intervals are estimated as described in Section 4. All standard errors are clustered at the major-cohort level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

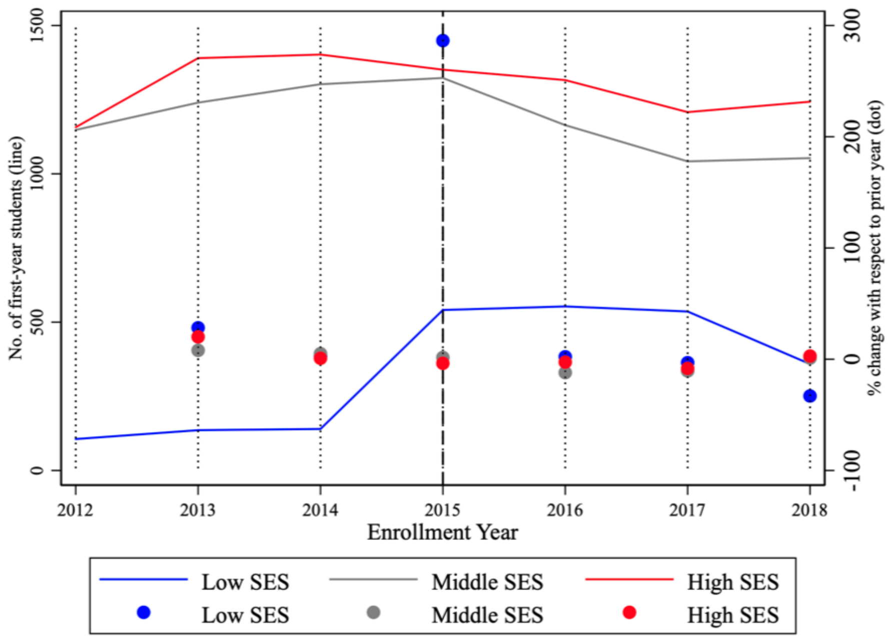
Table 6: Effect of low-income classroom peers in Middle–SES students' interactions

	Interactions of Middle SES students with:		
	(1) Middle SES	(2) High SES	(3) Low SES
R_i	-0.030	-0.029	0.066***
	(0.020)	(0.018)	(0.012)
Monte Carlo Benchmark			
95% C.I.	[-0.0099. -0.0098]	[-0.0210, -0.0208]	[0.0324, 0.0325]
mean(t-test)	1.23	2.4	6.88
Pre-treatment statistics			
sd(H_ij)	0.20	0.16	0.07
sd(R_i pc)	0.016	0.016	0.016
No. Students	2,604	2,604	2,604
Num. Obs.	199,315	199,315	199,315

Notes: This table displays the estimates from Equation 1. R_i is the proportion of low-SES students in the major and entry cohort as treatment. All regressions include the controls depicted in Equation 1 and described in the empirical analysis section. I address missing values by including an indicator equal to 1 if any of the covariates contains a missing value for the link ij. Each observation represents a link which results from all possible combinations of pairs of students from the same major and cohort, when one of the pair is a middle–SES student. Monte Carlo Confidence Intervals are estimated as described in Section 4. All standard errors are clustered at the major-cohort level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

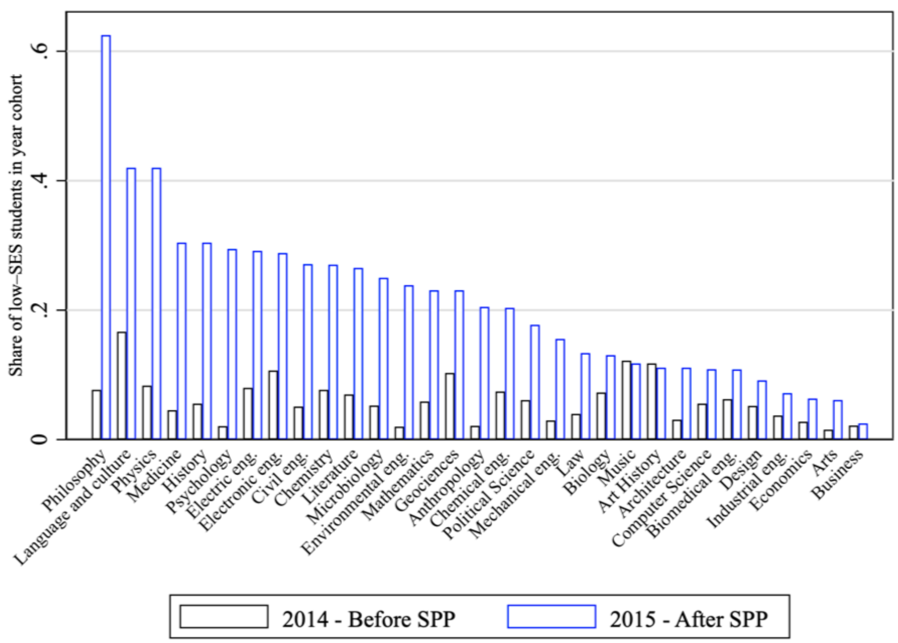
Figures

Figure 1: Number of first term students by socio–economic status (SES)



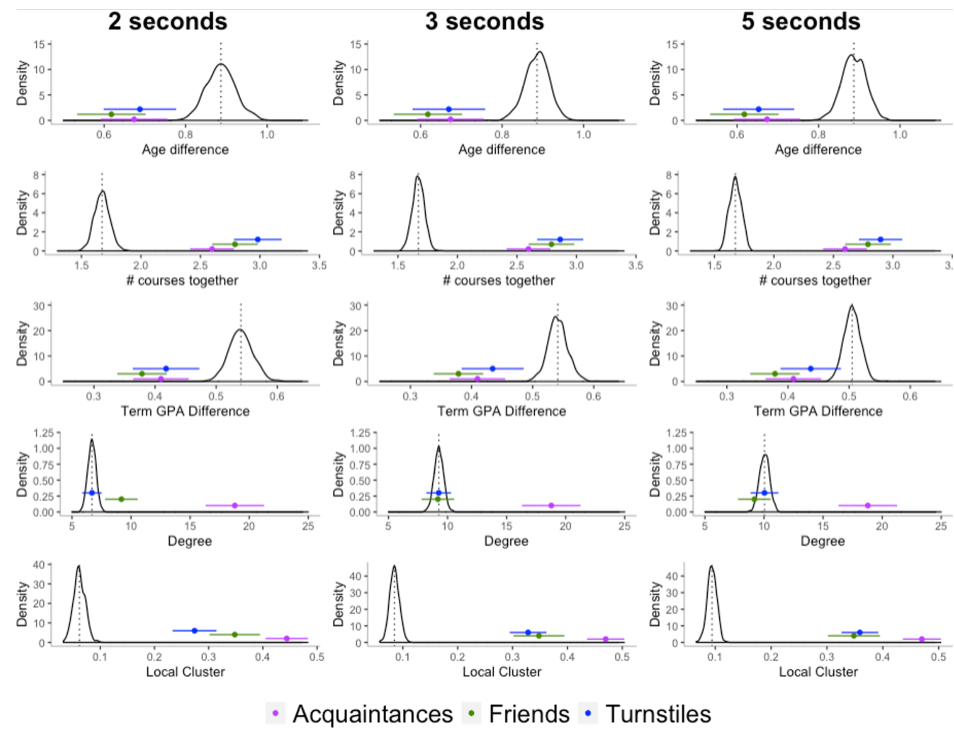
Notes: This figure displays the total number of first–term students by socio–economic (SES) background. I add both spring and fall enrollment per year

Figure 2: Share of low–SES students by major and before and after SPP



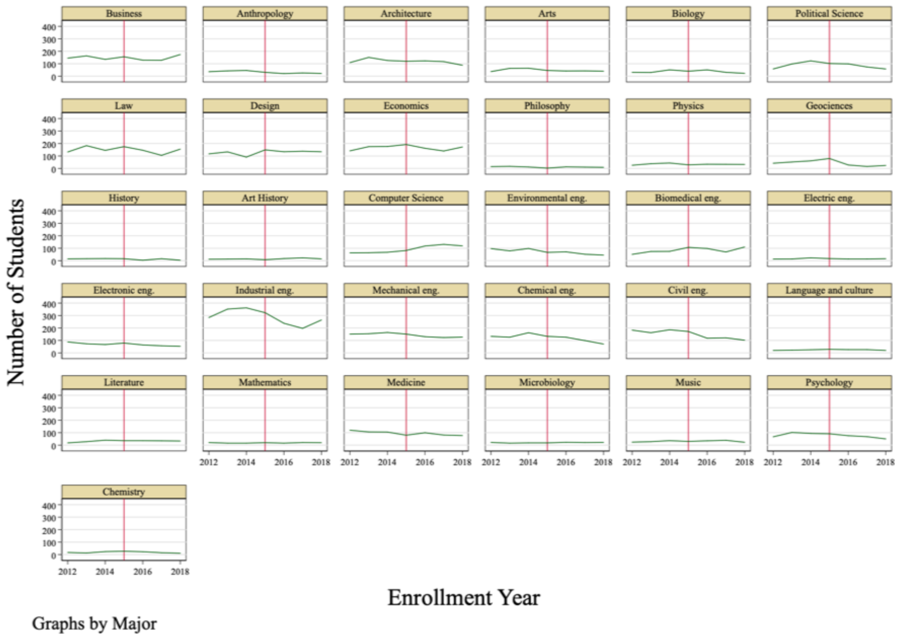
Notes: This figure displays the share of low–SES students per major and in the first term of enrollment. The shares are calculated per year, by adding the total number of new students enrolled in each major and semester (fall and spring).

Figure 3: Comparison against randomly generated distribution



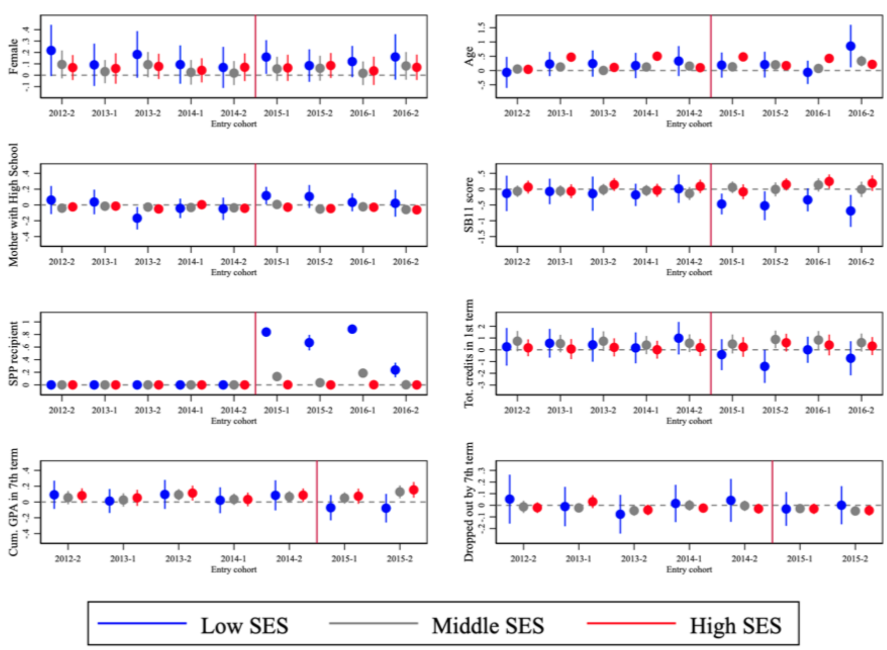
Notes: Turnstile-elicited links matched with the survey are randomly assigned in 1000 draws among 110 students forming all possible 5595 dyads. Confidence intervals of 95% confidence are presented. Matches for 2 seconds - 2 times window: 368 links. Matches for 3 seconds - 2 times window: 368 links. Matches for 5 seconds - 3 times window: 552 links.

Figure 4: Number of non-low-SES first-term students enrolled by major



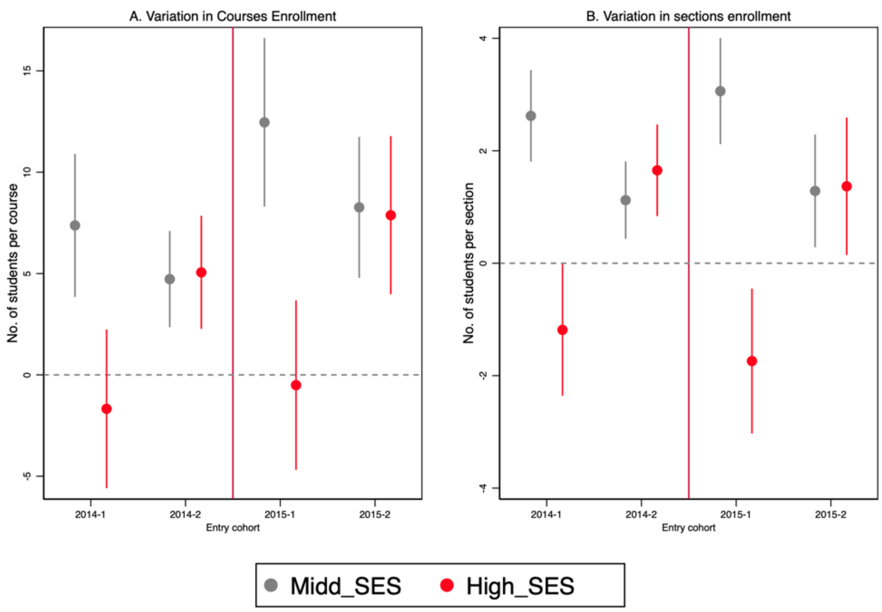
Notes: This figure displays the number of first-term non-low–SES students per major. The number of students enrolled are aggregated by year. Thus, I add the total number of new students enrolled in each major and semester (fall and spring).

Figure 5: Student characteristics trends, before and after SPP



Notes: -1 refers to the Spring and -2 refers to the Fall cohort. Point estimates of a cohort dummy from a regression where the dependent variable is the characteristic of the students. I control for major of enrollment dummies. Cluster standard errors at the major-cohort level. 95% confidence intervals.

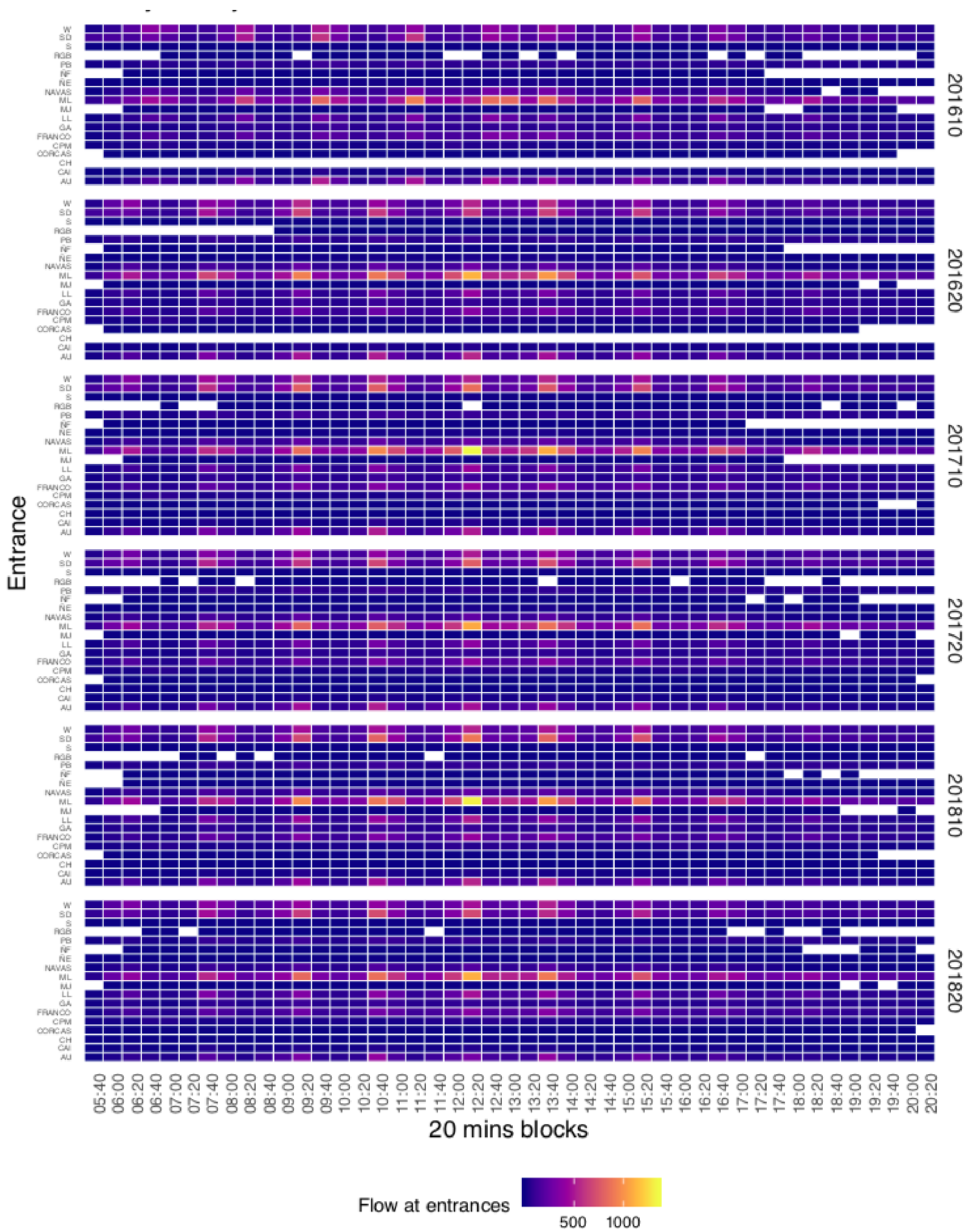
Figure 6: Variation in course enrollment by SES group



Notes: -1 refers to the Spring and -2 refers to the Fall cohort. Point estimates of a cohort dummy from a regression where the dependent variable is the number of students enrolled. Fall of 2013 is the omitted term. I include all courses taken by at least one first-term student in each period. Both panels include fixed effects by course. Cluster standard errors at the course level. 95% confidence intervals.

Appendix

Figure 7: Flow of students by entrance - term and hour according to turnstiles



Notes: Average number of taps per day, entrance and 20 minutes blocks. Taps include INs and OUT of the building. Each term has approximately 75 weekdays

Table 7: Effect of major-cohort low-income peers on High–SES students' interactions - outcome under 2 seconds time window

	Interactions of High SES students with:		
	(2) High SES	(3) Middle SES	(4) Low SES
R_i pc	-0.063***	-0.018*	0.018***
	(0.021)	(0.011)	(0.004)
Monte Carlo Benchmark			
95% C.I.	[-0.0584, -0.0583]	[-0.0502, -0.0501]	[0.0522, 0.0523]
mean(t-test)	7.51	7.65	18.46
Pre-treatment statistics			
sd(H_ij)	0.17	0.13	0.04
sd(R_i pc)	0.026	0.026	0.026
No. Students	2,674	2,674	2,674
Num. Obs.	224,690	224,690	224,690

Notes: This table displays the estimates from Equation 1. R_i is the proportion of low-SES students in the major and entry cohort as treatment. All regressions include the controls depicted in Equation 1 and described in the empirical analysis section. I address missing values by including an indicator equal to 1 if any of the covariates contains a missing value for the link ij. Each observation represents a link which results from all possible combinations of pairs of students from the same major and cohort, when one of the pair is a middle–SES student. Monte Carlo Confidence Intervals are estimated as described in Section 4. All standard errors are clustered at the major-cohort level. * p < 0.10, ** p < 0.05, *** p < 0.01.

Table 8: Effect of major-cohort low-income peers on Middle–SES students' interactions - outcome under 2 seconds time window

	Interactions of Middle SES students with:		
	(2) Middle SES	(3) High SES	(4) Low SES
R_i pc	-0.028*	-0.018	0.049***
	(0.015)	(0.015)	(0.010)
Monte Carlo Benchmark			
95% C.I.	[-0.0432, -0.0430]	[-0.0363, -0.0362]	[0.0543, 0.0544]
mean(t-test)	6.64	6.03	16.87
Pre-treatment statistics			
sd(H_ij)	0.18	0.14	0.06
sd(R_i pc)	0.028	0.028	0.028
No. Students	2,604	2,604	2,604
Num. Obs.	199,315	199,315	199,315

Notes: This table displays the estimates from Equation 1. R_i is the proportion of low-SES students in the major and entry cohort as treatment. All regressions include the controls depicted in Equation 1 and described in the empirical analysis section. I address missing values by including an indicator equal to 1 if any of the covariates contains a missing value for the link ij. Each observation represents a link which results from all possible combinations of pairs of students from the same major and cohort, when one of the pair is a middle–SES student. Monte Carlo Confidence Intervals are estimated as described in Section 4. All standard errors are clustered at the major-cohort level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 9: Effect of low-income classroom peers on High–SES students interactions - 2 seconds window

	Interactions of High SES students with:		
	(1) High SES	(2) Middle SES	(3) Low SES
R_i	-0.048**	0.015	0.009
	(0.024)	(0.013)	(0.007)
Monte Carlo Benchmark			
95% C.I.	[-0.0379, -0.0377]	[-0.0266, -0.0265]	[0.0348, 0.0350]
mean(t-test)	3.43	2.87	8.7
Pre-treatment statistics			
sd(H_ij)	0.17	0.13	0.04
sd(R_i pc)	0.014	0.014	0.014
No. Students	2,674	2,674	2,674
Num. Obs.	224,690	224,690	224,690

Notes: This table displays the estimates from Equation 1. R_i is the proportion of low-SES students in the major and entry cohort as treatment. All regressions include the controls depicted in Equation 1 and described in the empirical analysis section. I address missing values by including an indicator equal to 1 if any of the covariates contains a missing value for the link ij. Each observation represents a link which results from all possible combinations of pairs of students from the same major and cohort, when one of the pair is a middle–SES student. Monte Carlo Confidence Intervals are estimated as described in Section 4. All standard errors are clustered at the major-cohort level. * p < 0.10, ** p < 0.05, *** p < 0.01.

Table 10: Effect of low-income classroom peers on Middle–SES students interactions - 2 seconds window

	Interactions of Middle SES students with:		
	(1) Middle SES	(2) High SES	(3) Low SES
R_i	0.009	-0.011	0.030**
	(0.012)	(0.011)	(0.012)
Monte Carlo Benchmark			
95% C.I.	[-0.0085, -0.0084]	[-0.0144, -0.0143]	[0.0230, 0.0231]
mean(t-test)	1.2	1.86	5.54
Pre-treatment statistics			
sd(H_ij)	0.18	0.14	0.06
sd(R_i pc)	0.016	0.016	0.016
No. Students	2,604	2,604	2,604
Num. Obs.	199,315	199,315	199,315

Notes: This table displays the estimates from Equation 1. R_i is the proportion of low-SES students in the major and entry cohort as treatment. All regressions include the controls depicted in Equation 1 and described in the empirical analysis section. I address missing values by including an indicator equal to 1 if any of the covariates contains a missing value for the link ij. Each observation represents a link which results from all possible combinations of pairs of students from the same major and cohort, when one of the pair is a middle–SES student. Monte Carlo Confidence Intervals are estimated as described in Section 4. All standard errors are clustered at the major-cohort level. * p < 0.10, ** p < 0.05, *** p < 0.01.

Table 11: Effect of major-cohort low-income peers on High–SES students' interactions - outcome under 5 seconds time window

	Interactions of High SES students with:		
	(1) High SES	(2) Middle SES	(3) Low SES
R_i pc	-0.083***	-0.025**	0.030***
	(0.024)	(0.012)	(0.005)
Monte Carlo Benchmark			
95% C.I.	[-0.0765, -0.0763]	[-0.0559, -0.0558]	[0.0702, 0.0703]
mean(t-test)	8.86	7.71	22.162
Pre-treatment statistics			
sd(H_ij)	0.19	0.15	0.04
sd(R_i pc)	0.026	0.026	0.026
No. Students	2,674	2,674	2,674
Num. Obs.	224,690	224,690	224,690

Notes: This table displays the estimates from Equation 1. R_i is the proportion of low-SES students in the major and entry cohort as treatment. All regressions include the controls depicted in Equation 1 and described in the empirical analysis section. I address missing values by including an indicator equal to 1 if any of the covariates contains a missing value for the link ij. Each observation represents a link which results from all possible combinations of pairs of students from the same major and cohort, when one of the pair is a middle–SES student. Monte Carlo Confidence Intervals are estimated as described in Section 4. All standard errors are clustered at the major-cohort level. * p < 0.10, ** p < 0.05, *** p < 0.01.

Table 12: Effect of major-cohort low-income peers on Middle–SES students' interactions - outcome under 5 seconds time window

	Interactions of Middle SES students with:		
	(1) Middle SES	(2) High SES	(3) Low SES
R_i pc	-0.028	-0.032*	0.066***
	(0.019)	(0.018)	(0.013)
Monte Carlo Benchmark			
95% C.I.	[-0.0507, -0.0505]	[-0.0500, -0.0499]	[0.0717, 0.0718]
mean(t-test)	7.07	7.52	19.84
Pre-treatment statistics			
sd(H_ij)	0.194	0.16	0.07
sd(R_i pc)	0.028	0.028	0.028
No. Students	2,604	2,604	2,604
Num. Obs.	199,315	199,315	199,315

Notes: This table displays the estimates from Equation 1. R_i is the proportion of low-SES students in the major and entry cohort as treatment. All regressions include the controls depicted in Equation 1 and described in the empirical analysis section. I address missing values by including an indicator equal to 1 if any of the covariates contains a missing value for the link ij. Each observation represents a link which results from all possible combinations of pairs of students from the same major and cohort, when one of the pair is a middle–SES student. Monte Carlo Confidence Intervals are estimated as described in Section 4. All standard errors are clustered at the major-cohort level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 13: Effect of low-income classroom peers on High–SES students interactions - 5 seconds window

	Interactions of High SES students with:		
	(1) High SES	(2) Middle SES	(3) Low SES
R_i	-0.057**	0.015	0.017*
	(0.028)	(0.016)	(0.009)
Monte Carlo Benchmark			
95% C.I.	[-0.0445, -0.0443]	[-0.0279, -0.0278]	[0.0463, 0.0465]
mean(t-test)	3.63	2.71	10.32
Pre-treatment statistics			
sd(H_ij)	0.19	0.15	0.04
sd(R_i pc)	0.014	0.014	0.014
No. Students	2,674	2,674	2,674
Num. Obs.	224,690	224,690	224,690

Notes: This table displays the estimates from Equation 1. R_i is the proportion of low-SES students in the major and entry cohort as treatment. All regressions include the controls depicted in Equation 1 and described in the empirical analysis section. I address missing values by including an indicator equal to 1 if any of the covariates contains a missing value for the link ij. Each observation represents a link which results from all possible combinations of pairs of students from the same major and cohort, when one of the pair is a middle–SES student. Monte Carlo Confidence Intervals are estimated as described in Section 4. All standard errors are clustered at the major-cohort level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 14: Effect of low-income classroom peers on Middle–SES students interactions - 5 seconds window

	Interactions of Middle SES students with:		
	(1) Middle SES	(2) High SES	(3) Low SES
R_i	0.016	-0.020	0.049***
	(0.017)	(0.014)	(0.016)
Monte Carlo Benchmark			
95% C.I.	[-0.0094, -0.0092]	[-0.0225, -0.0224]	[0.0343, 0.0344]
mean(t-test)	1.21	2.61	7.34
Pre-treatment statistics			
sd(H_ij)	0.194	0.16	0.07
sd(R_i pc)	0.016	0.016	0.016
No. Students	2,604	2,604	2,604
Num. Obs.	199,315	199,315	199,315

Notes: This table displays the estimates from Equation 1. R_i is the proportion of low-SES students in the major and entry cohort as treatment. All regressions include the controls depicted in Equation 1 and described in the empirical analysis section. I address missing values by including an indicator equal to 1 if any of the covariates contains a missing value for the link ij. Each observation represents a link which results from all possible combinations of pairs of students from the same major and cohort, when one of the pair is a middle–SES student. Monte Carlo Confidence Intervals are estimated as described in Section 4. All standard errors are clustered at the major-cohort level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.