Diversity in College, Academic Achievement and Students' Interactions: Evidence from Turnstile Data

Tatiana Velasco Rodriguez
Columbia University

January, 2022

Motivation

1. Segregation is a pervasive issue in education.
   - **Segregation**: separation of groups by race, Socio–Economic Status (SES) or ethnicity, voluntarily or enforced. The opposite is desegregation, integration, or diversity.

Motivation

1. Segregation is a pervasive issue in education.
   - **Segregation**: separation of groups by race, Socio–Economic Status (SES) or ethnicity, voluntarily or enforced. The opposite is desegregation, integration, or diversity.

2. To combat segregation **across schools**, policy makers have implemented desegregation policies (busing programs, financial aid programs and affirmative action).
   - These policies foster access of low–income and minority students to elite institutions

Motivation

1. Segregation is a pervasive issue in education.
   - **Segregation**: separation of groups by race, Socio–Economic Status (SES) or ethnicity, voluntarily or enforced. The opposite is desegregation, integration, or diversity.

2. To combat segregation **across schools**, policy makers have implemented desegregation policies (busing programs, financial aid programs and affirmative action).
   - These policies foster access of low–income and minority students to elite institutions

3. However, previous research suggests **segregation within schools** may persist (Armstrong and Hamilton, 2015; Carrel et al., 2013; Corredor et al., 2019).

Affirmative Action (AA) and Desegregation Policies

1. Attending selective colleges benefits low–income students
   (Chetty et al., 2020; Hoekstra, 2009; Londoño-Velez, et al., 2020; Bleemer, 2021; others)

Affirmative Action (AA) and Desegregation Policies

1. Attending selective colleges benefits low–income students
   (Chetty et al., 2020; Hoekstra, 2009; Londoño-Velez, et al., 2020; Bleemer, 2021; others)

2. Changes in the composition of students have consequences:

Affirmative Action (AA) and Desegregation Policies

1. Attending selective colleges benefits low–income students
   (Chetty et al., 2020; Hoekstra, 2009; Londoño-Velez, et al., 2020; Bleemer, 2021; others)

2. Changes in the composition of students have consequences:
   - (-) Potentially negative peer effects (Arcidiacono et al., 2015; Carrell et al., 2018)
   - (-) If desegregation induces achievement gaps, segregation within groups may persist (Carrell et al., 2013)

Affirmative Action (AA) and Desegregation Policies

1. Attending selective colleges benefits low–income students
   (Chetty et al., 2020; Hoekstra, 2009; Londoño-Velez, et al., 2020; Bleemer, 2021; others)

2. Changes in the composition of students have consequences:
   - (-) Potentially negative peer effects (Arcidiacono et al., 2015; Carrell et al., 2018)
   - (-) If desegregation induces achievement gaps, segregation within groups may persist (Carrell et al., 2013)
     - → Students tends to interact with others like themselves – *homophily*
       (Baker et al., 2011; Marmaros and Sacerdote, 2006; Mayer and Puller, 2008)

Affirmative Action (AA) and Desegregation Policies

1. Attending selective colleges benefits low–income students
   (Chetty et al., 2020; Hoekstra, 2009; Londoño-Velez, et al., 2020; Bleemer, 2021; others)

2. Changes in the composition of students have consequences:
   (-) Potentially negative peer effects (Arcidiacono et al., 2015; Carrell et al., 2018)

   (-) If desegregation induces achievement gaps, segregation within groups may persist (Carrell et al., 2013)
   - $\rightarrow$ Students tends to interact with others like themselves – *homophily*
     (Baker et al., 2011; Marmaros and Sacerdote, 2006; Mayer and Puller, 2008)

   (+) Social connections matter: cultural and social capital drives positive impacts on low–income students (social mobility)
   (Lleras-Muney et al., 2020; Michelman, et al., 2021; Zimmerman, 2019)

This Paper

Research questions:

1. Does the increased exposure to low–income peers impact wealthy students' academic achievement?

2. Does desegregation leads to more diverse social interactions?

This Paper

Research questions:

1. Does the increased exposure to low–income peers impact wealthy students' academic achievement?

2. Does desegregation leads to more diverse social interactions?

I use the shock induced by **Ser Pilo Paga (SPP)**, a financial aid program that dramatically increased the enrollment of students from low socio–economic status at elite universities in Colombia.

This Paper

Research questions:

1. Does the increased exposure to low–income peers impact wealthy students' academic achievement?
2. Does desegregation leads to more diverse social interactions?

I use the shock induced by **Ser Pilo Paga (SPP)**, a financial aid program that dramatically increased the enrollment of students from low socio–economic status at elite universities in Colombia.

I exploit the variation in **intensity of exposure to low–income peers** within cohorts and across majors in a **difference-in-difference** framework

This Paper

Research questions:

1. Does the increased exposure to low–income peers impact wealthy students' academic achievement?
2. Does desegregation leads to more diverse social interactions?

I use the shock induced by **Ser Pilo Paga (SPP)**, a financial aid program that dramatically increased the enrollment of students from low socio–economic status at elite universities in Colombia.

I exploit the variation in **intensity of exposure to low–income peers** within cohorts and across majors in a **difference-in-difference** framework

- To identify student social interactions I leverage **turnstiles records** to capture students co-movements across campus.

Findings and Contributions

1. No significant impacts of exposure to desegregation on the academic achievement of wealthy students

    Aligned with K–12 studies (Angrist and Lang, 2004; Dobbie and Fryer, 2014)

    → Consistent with Bleemer (2021) and in contrast to Arcidiacono and Vigdor (2010).

Findings and Contributions

1. No significant impacts of exposure to desegregation on the academic achievement of wealthy students

    Aligned with K–12 studies (Angrist and Lang, 2004; Dobbie and Fryer, 2014)

    → Consistent with Bleemer (2021) and in contrast to Arcidiacono and Vigdor (2010).

2. Lack of peer effects is not explained by lack of diversity in social interactions. On average the interactions between low–income and wealthy doubled

    → Contrast with Zimmerman and co–authors (2019, 2021)

    Measure of effects on interactions refines on previous literature (e.g. Marmaros and Sacerdote, 2006; Baker, Mayer and Puller, 2011).

Findings and Contributions

1. No significant impacts of exposure to desegregation on the academic achievement of wealthy students

   Aligned with K–12 studies (Angrist and Lang, 2004; Dobbie and Fryer, 2014)

   → Consistent with Bleemer (2021) and in contrast to Arcidiacono and Vigdor (2010).

2. Lack of peer effects is not explained by lack of diversity in social interactions. On average the interactions between low–income and wealthy doubled

   → Contrast with Zimmerman and co–authors (2019, 2021)

   Measure of effects on interactions refines on previous literature (e.g. Marmaros and Sacerdote, 2006; Baker, Mayer and Puller, 2011).

3. High academic achievement is a driver of integration

Structure of the talk

1. Motivation, framework and contribution
2. Policy context
3. Data and identification
4. Findings
5. Assumptions and robustness checks
6. Conclusions

Policy Context

Ser Pilo Paga at Elite University

- SPP offered forgivable loans to low–income students scoring on the top 10 percent of the SABER 11 exam

- Only colleges certified as High Quality were eligible for enrollment

- Loan is fully forgiven upon completion of the degree

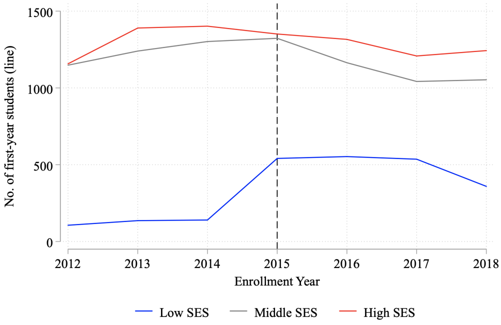


Figure: No. of students per entry year and SES

SPP Context

SPP Policy Shock Timeline

2014	Spring (-1) and Fall (-2) cohorts enroll (Pre-SPP)
Early October 2014	Students take National Exam SB11
October 2014	Ser Pilo Paga (SPP) is announced - media coverage is intense
Nov-Dec 2014	Students submit applications to college-majors
January 2015	First cohort of SPP recipients - New cohorts each Spring
February 16	Turnstile data starts to be stored by the University
Fall/16 to Fall/18	Turnstile-elicited interactions are measured

No relation with changes in the number of wealthy students

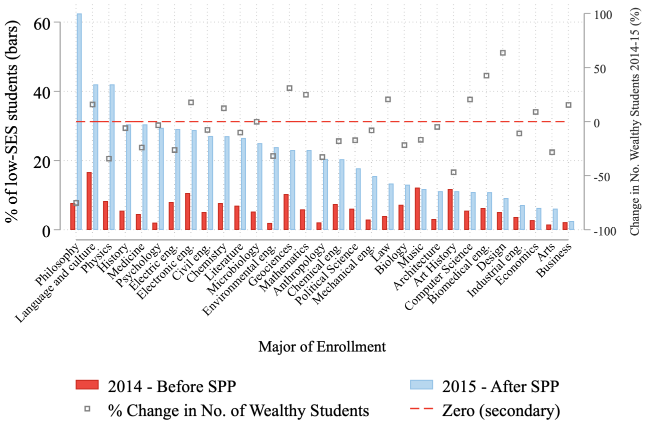


Figure: Changes in Major and Entry Cohort Student Composition After SPP

▸ Correlation ▸ Size of groups

Classrooms occupation increased, but not over capacity

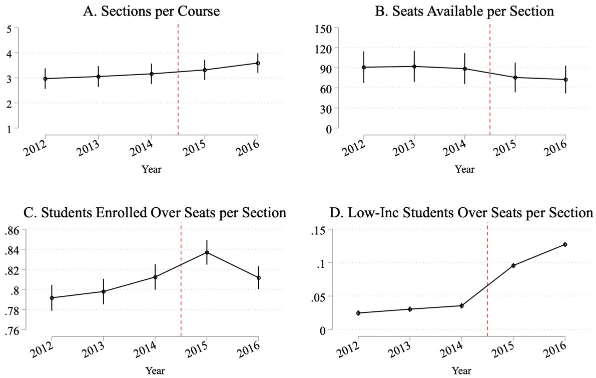


Figure: Average Composition of Courses taken by First–Term Students (95% C.I.)

Number of sections per course and seats available per section did not change with SPP.

Differences in Students' Characteristics

	2014 entry cohort				2015 entry cohort			
	Wealthy	Low-SES			Wealthy	Low-SES		
	Mean	Mean	Difference		Mean	Mean	Difference	
Peers composition								
Prop. of middle–SES	0.49				0.50			
Student characteristics								
Female	0.43	0.35	-0.09	**	0.46	0.41	-0.06	
SB11 standardized test score	0.00	-0.09	-0.10		0.04	-0.22	-0.26	***
SPP recipient	0.00	0.00	0.00		0.05	0.81	0.76	***
Other scholarship recipient	0.07	0.37	0.30	***	0.07	0.05	-0.02	
No. Of programs	31	31			31	31		
No. Of students	2669	139			2609	538		

Table: Student Characteristics

Low–income students have significantly lower test scores than their wealthy peers

	2014 entry cohort				2015 entry cohort			
	Wealthy	Low-SES			Wealthy	Low-SES		
	Mean	Mean	Difference		Mean	Mean	Difference	
Peers composition								
Prop. of middle–SES	0.49				0.50			
Student characteristics								
Female	0.43	0.35	-0.09	**	0.46	0.41	-0.06	
SB11 standardized test score	0.00	-0.09	-0.10		0.04	-0.22	-0.26	***
SPP recipient	0.00	0.00	0.00		0.05	0.81	0.76	***
Other scholarship recipient	0.07	0.37	0.30	***	0.07	0.05	-0.02	
No. Of programs	31	31			31	31		
No. Of students	2669	139			2609	538		

Table: Student Characteristics

SPP is the main financial aid source among 2015 low–income students

	2014 entry cohort				2015 entry cohort			
	Wealthy	Low-SES			Wealthy	Low-SES		
	Mean	Mean	Difference		Mean	Mean	Difference	
Peers composition								
Prop. of middle–SES	0.49				0.50			
Student characteristics								
Female	0.43	0.35	-0.09	**	0.46	0.41	-0.06	
SB11 standardized test score	0.00	-0.09	-0.10		0.04	-0.22	-0.26	***
SPP recipient	0.00	0.00	0.00		0.05	0.81	0.76	***
Other scholarship recipient	0.07	0.37	0.30	***	0.07	0.05	-0.02	
No. Of programs	31	31			31	31		
No. Of students	2669	139			2609	538		

Table: Student Characteristics

Among 2015 cohorts, low–income students perform significantly worse than wealthy students

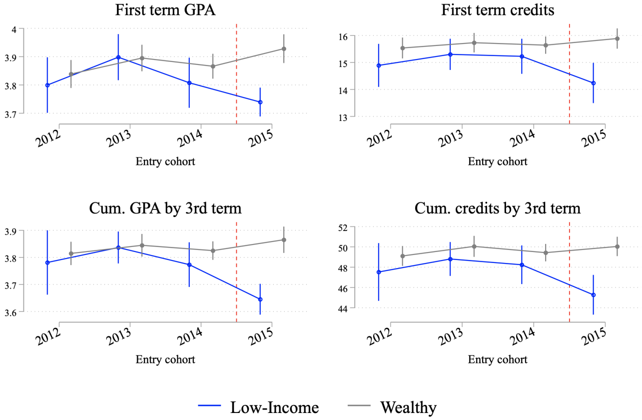


Figure: Ave. cumulative GPA and credits attempted by entry cohorts

Data and Identification

Turnstiles Data

The University campus is guarded by 18 entrances all with turnstiles.

I use records since fall of 2016 to the fall of 2019 matched with administrative student-term level data.

Turnstile-elicited link: *when a pair student IDs are swiped at a turnstile in the same entrance and direction, in a window of three seconds or less, and at least twice in a term.*

Figure: Turnstile Example

Measuring Social Interactions Using Turnstiles Data

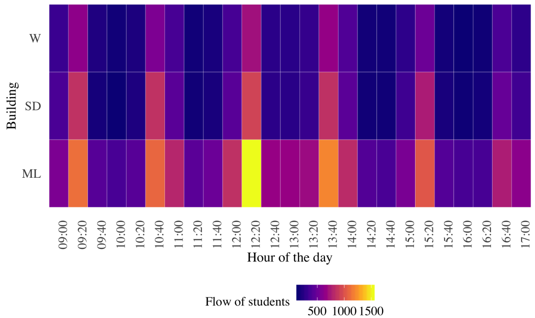


Figure: Average Student ID Taps During Week Days on Three of the Entrances

Students use the turnstiles throughout the day and at different intensities.

Descriptive statistics of students' links

	2014 entry cohort				2015 entry cohort			
	Wealthy	Low-SES			Wealthy	Low-SES		
	Mean	Mean	Difference		Mean	Mean	Difference	
No. Of links	5.21	4.94	0.27		5.53	4.60	0.93	**
No. of low SES links	0.24	0.35	0.11	*	0.59	1.73	1.14	***
Homophilic characteristics								
Age Difference	0.60	0.65	0.06		0.63	0.63	0.00	
Same Gender	0.50	0.51	0.00		0.51	0.45	0.05	**
Ave. No. Of courses w/links	1.49	1.37	0.13		1.51	1.31	0.20	
SB11 difference	0.73	0.76	0.03		0.79	0.67	0.12	**
Share of friends from same high school	0.04	0.01	0.03	***	0.04	0.00	0.04	***
No. Of students	2669	139			2609	538		
Individuals without links in their program-cohort	600	40			541	151		

Table: Student Characteristics

Identification Strategy Framework

Sample: relatively wealthy students.

$$Outcome_i^{mc} = \beta_l R_{mc}^l + \mathbf{X_i'} B + \beta_m + \beta_c + \varepsilon_{imc}$$

- **Outcomes:** academic achievement (GPA and attempted credits, persistance measures), number of links with other students.
- $R_{mc}^l = \frac{N_{mc}^{low-inc}}{N_{mc}} * 100$ with $N_{mc}^l =$ number of low–income students, and N_{mc} is the number of students in major m and entry cohort c.
  - $\rightarrow$ β_l is the estimate of interest. Huber–White Clustered S.E. by major–cohort
- X_i: Student i characteristics (Female, Age, Mother' highest education level is high school, SB11 score, middle–SES, SPP)

Results are Robust to Identification Assumptions

1. *Parallel trends assumption:* in the absence of SPP, the trends in the outcome should change the same for both treated and control groups. ▸ test

2. *Measurement Error in the Outcomes:*
   - Measurement Error in turnstile–elicited interactions. ▸ Turnstiles Validation
   - GPA is a noisy measure of academic performance $\rightarrow$ use other indicators like credits and persistence measures

3. *Unobserved Exposure Effects:* exposure through courses differs from exposure through majors–cohorts ▸ threats

   The following results are robust to these forms of biases

Findings

Effects on Academic Achievement and Persistence

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	1st term credits	1st term GPA	3rd term cum. Credits	3rd term cum. GPA	6th term cum. Credits	6th term cum. GPA	Dropout by 5th term	Graduated on time
A. OLS								
%tage of low–income peers	0.011*	0.001	0.034**	0.001	0.031	0.001	0.002**	-0.000
	(0.006)	(0.001)	(0.017)	(0.001)	(0.031)	(0.001)	(0.001)	(0.001)
B. Non-linear Effects								
I[%tage of low–income peers $>$ 30%]	0.063	-0.021	0.277	0.002	-0.065	-0.004	0.028	-0.034
	(0.174)	(0.033)	(0.571)	(0.026)	(0.815)	(0.022)	(0.020)	(0.029)
mu(Outcome)	15.64	3.863	49.39	3.822	100.9	3.859	0.122	0.0693
sd(Outcome)	2.949	0.449	8.496	0.378	16.22	0.344	0.327	0.254
No. Students	5,278	5,274	4,895	4,895	4,507	4,507	5,278	5,278

- No impacts on GPA or graduation rates, some small and positive impacts on number of credits attempted and dropout

Effects on Academic Achievement and Persistence

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	1st term credits	1st term GPA	3rd term cum. Credits	3rd term cum. GPA	6th term cum. Credits	6th term cum. GPA	Dropout by 5th term	Graduated on time
A. OLS								
%tage of low–income peers	0.011*	0.001	0.034**	0.001	0.031	0.001	0.002**	-0.000
	(0.006)	(0.001)	(0.017)	(0.001)	(0.031)	(0.001)	(0.001)	(0.001)
B. Non-linear Effects								
I[%tage of low–income peers $>$ 30%]	0.063	-0.021	0.277	0.002	-0.065	-0.004	0.028	-0.034
	(0.174)	(0.033)	(0.571)	(0.026)	(0.815)	(0.022)	(0.020)	(0.029)
mu(Outcome)	15.64	3.863	49.39	3.822	100.9	3.859	0.122	0.0693
sd(Outcome)	2.949	0.449	8.496	0.378	16.22	0.344	0.327	0.254
No. Students	5,278	5,274	4,895	4,895	4,507	4,507	5,278	5,278

- Size of the effects on credits attempted is small ($\sim$ 0.05 s.d.)

Effects on Academic Achievement and Persistence

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	1st term credits	1st term GPA	3rd term cum. Credits	3rd term cum. GPA	6th term cum. Credits	6th term cum. GPA	Dropout by 5th term	Graduated on time
A. OLS								
%tage of low–income peers	0.011*	0.001	0.034**	0.001	0.031	0.001	0.002**	-0.000
	(0.006)	(0.001)	(0.017)	(0.001)	(0.031)	(0.001)	(0.001)	(0.001)
B. Non-linear Effects								
I[%tage of low–income peers $>$ 30%]	0.063	-0.021	0.277	0.002	-0.065	-0.004	0.028	-0.034
	(0.174)	(0.033)	(0.571)	(0.026)	(0.815)	(0.022)	(0.020)	(0.029)
mu(Outcome)	15.64	3.863	49.39	3.822	100.9	3.859	0.122	0.0693
sd(Outcome)	2.949	0.449	8.496	0.378	16.22	0.344	0.327	0.254
No. Students	5,278	5,274	4,895	4,895	4,507	4,507	5,278	5,278

- No evidence suggesting non–linear effects on students at the top 25th percentile of the distribution of exposure to low–incs.

Interpreting the Lack of Effects on Achievement

Peer effects literature would suggest increased exposure to relatively low achievers should affect the performance of high achievers (Arcidiacono et al., 2015)

My results suggest otherwise. Why?

H: Segregation within groups is persistent Lack of effects on achievement due to lack of interactions between wealthy and low–income students.

Interpreting the Lack of Effects on Achievement

Peer effects literature would suggest increased exposure to relatively low achievers should affect the performance of high achievers (Arcidiacono et al., 2015)

My results suggest otherwise. Why?

H: Segregation within groups is persistent Lack of effects on achievement due to lack of interactions between wealthy and low–income students.

Estimate the effect of increased exposure to low–income peers on wealthy students':

A. Probability of a link with a low–income (extensive margin)

B. Number of links with the low–income (intensive margin)

C. Friendships composition as measured by the percentage of links with low–income peers

Effects on Social Interactions

	A. Probability of a Link with		B. Number of Links with			C. %tage of Links with
	(1) Wealthy	(2) Low Income	(3) Any	(4) Wealthy	(5) Low Income	(6) Low Income
A. OLS						
%tage of low–income peers	-0.002**	0.008***	-0.011	-0.043***	0.031***	0.714***
	(0.001)	(0.001)	(0.016)	(0.016)	(0.004)	(0.060)
At the mean increase (9.51 p.p.)	**-0.019**	**0.076**		**-0.409**	**0.295**	**6.790**
B. Non-linear Effects						
1[%tage of low–income peers> 30%]	-0.043	0.117***	0.017	-0.667	0.684***	14.566***
	(0.028)	(0.038)	(0.495)	(0.507)	(0.131)	(2.915)
Pre-treatment statistics						
mu(No. of friends)	0.770	0.188	5.212	4.973	0.239	4.404
sd(No. of friends)	0.421	0.391	5.154	4.922	0.558	11.34
No. Students	5,278	5,278	5,278	5,278	5,278	4,137

On average, the probability of an interaction between a wealthy and a low–income increased in 42% relative to pre–SPP

the number of interactions with low–income peers increased in 120% relative to pre–SPP

Effects on Social Interactions

	A. Probability of a Link with		B. Number of Links with			C. %tage of Links with
	(1) Wealthy	(2) Low Income	(3) Any	(4) Wealthy	(5) Low Income	(6) Low Income
A. OLS						
%tage of low–income peers	-0.002**	0.008***	-0.011	-0.043***	0.031***	0.714***
	(0.001)	(0.001)	(0.016)	(0.016)	(0.004)	(0.060)
At the mean increase (9.51 p.p.)	**-0.019**	**0.076**		**-0.409**	**0.295**	**6.790**
B. Non-linear Effects						
1[%tage of low–income peers$>$ 30%]	-0.043	0.117***	0.017	-0.667	0.684***	14.566***
	(0.028)	(0.038)	(0.495)	(0.507)	(0.131)	(2.915)
Pre-treatment statistics						
mu(No. of friends)	0.770	0.188	5.212	4.973	0.239	4.404
sd(No. of friends)	0.421	0.391	5.154	4.922	0.558	11.34
No. Students	5,278	5,278	5,278	5,278	5,278	4,137

Interactions with other wealthy peers decreased, but size of the effect is small relative to pre–SPP ($<$ 1% decrease at the intensive and extensive margins)

Effects on Social Interactions

	A. Probability of a Link with		B. Number of Links with			C. %tage of Links with
	(1) Wealthy	(2) Low Income	(3) Any	(4) Wealthy	(5) Low Income	(6) Low Income
A. OLS						
%tage of low–income peers	-0.002**	0.008***	-0.011	-0.043***	0.031***	0.714***
	(0.001)	(0.001)	(0.016)	(0.016)	(0.004)	(0.060)
At the mean increase (9.51 p.p.)	**-0.019**	**0.076**		**-0.409**	**0.295**	**6.790**
B. Non-linear Effects						
1[%tage of low–income peers> 30%]	-0.043	0.117***	0.017	-0.667	0.684***	14.566***
	(0.028)	(0.038)	(0.495)	(0.507)	(0.131)	(2.915)
Pre-treatment statistics						
mu(No. of friends)	0.770	0.188	5.212	4.973	0.239	4.404
sd(No. of friends)	0.421	0.391	5.154	4.922	0.558	11.34
No. Students	5,278	5,278	5,278	5,278	5,278	4,137

On average, wealthy students increase their share of links with low–income peers in 6.8 p.p. Change is not one-to-one.

Effects maintain at the top quarter of the distribution

Is Academic Achievement Explaining Integration?

Segregation within groups IS NOT persistent

Carrel et al., (2013): high achievers segregate from low achieving peers, explaining the lack of peer effects.

→ Could academic performance drive integration?

27% of the low–income students enrolling during 2015 had a SB11 performance equal or above the mean of that of their wealthy peers in the major–cohort

Identify the links that were with high–achieving low–income students

SB11 test score distribution between low–income and wealthy 2015-1 students

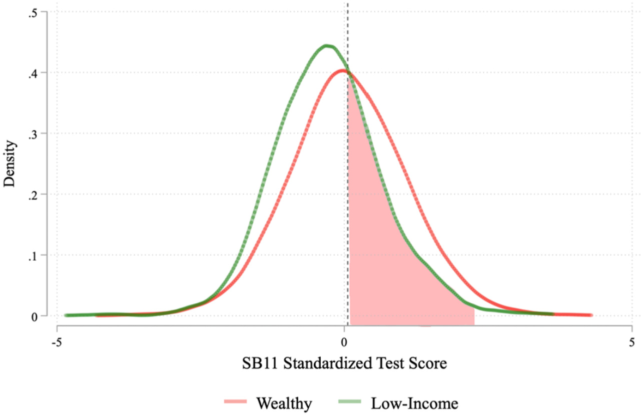


Figure: Low–income high–achieving are in the shaded area

Effects on Social Interactions by Achievement of the Low–Income

	A. Number of Links with low–Income High Achievers By:			B. Probability of a Link with High Achievers By:		
	(1) SB11	(2) First Term GPA	(3) First Term Credits Attempted	(4) SB11	(5) First Term GPA	(6) First Term Credits Attempted
A. Low – Income						
%tage of low–income peers	0.013***	0.016***	0.021***	0.006***	0.006***	0.008***
	(0.003)	(0.002)	(0.003)	(0.001)	(0.001)	(0.001)
At the mean increase (9.51 p.p.)	**0.124**	**0.152**	**0.200**	**0.057**	**0.057**	**0.076**
Pre-treatment statistics of the outcomes						
mean	0.0892	0.135	0.118	0.0798	0.111	0.102
standard Deviation	0.317	0.419	0.375	0.271	0.315	0.303
No. Students	5,278	5,278	5,278	5,278	5,278	5,278

Effects on Social Interactions by Achievement of the Low–Income

	A. Number of Links with low–Income High Achievers By:			B. Probability of a Link with High Achievers By:		
	(1) SB11	(2) First Term GPA	(3) First Term Credits Attempted	(4) SB11	(5) First Term GPA	(6) First Term Credits Attempted
A. Low – Income						
%tage of low–income peers	0.013***	0.016***	0.021***	0.006***	0.006***	0.008***
	(0.003)	(0.002)	(0.003)	(0.001)	(0.001)	(0.001)
At the mean increase (9.51 p.p.)	**0.124**	**0.152**	**0.200**	**0.057**	**0.057**	**0.076**
Pre-treatment statistics of the outcomes						
mean	0.0892	0.135	0.118	0.0798	0.111	0.102
standard Deviation	0.317	0.419	0.375	0.271	0.315	0.303
No. Students	5,278	5,278	5,278	5,278	5,278	5,278

About half of the increase in the number of interactions with low–income is due to interactions with high–achievers.

Effects on Social Interactions by Achievement of the Low–Income

	A. Number of Links with low–Income High Achievers By:			B. Probability of a Link with High Achievers By:		
	(1) SB11	(2) First Term GPA	(3) First Term Credits Attempted	(4) SB11	(5) First Term GPA	(6) First Term Credits Attempted
A. Low – Income						
%tage of low–income peers	0.013***	0.016***	0.021***	0.006***	0.006***	0.008***
	(0.003)	(0.002)	(0.003)	(0.001)	(0.001)	(0.001)
At the mean increase (9.51 p.p.)	**0.124**	**0.152**	**0.200**	**0.057**	**0.057**	**0.076**
Pre-treatment statistics of the outcomes						
mean	0.0892	0.135	0.118	0.0798	0.111	0.102
standard Deviation	0.317	0.419	0.375	0.271	0.315	0.303
No. Students	5,278	5,278	5,278	5,278	5,278	5,278

Estimates on links with low–income high–achievers by first–term performance are even larger.

Is Academic Achievement Explaining Integration?

Wealthy students form links with low–income students of high academic performance.

Hypothesis: socio-economic diversity is driven by links among the very–high–achieving

$\rightarrow$ Consistent with homophily. Matching with others of similar achievement out-weights socio–economic differences

$$L_{imc}^{w=0,HighAchiv} = \alpha_d (R^l_{mc} * HighAchiever_i) + \alpha_l R^l_{mc} + \alpha_p HighAchiever_i + \mathbf{X_i'} A + \alpha_m + \alpha_c + \varepsilon_{imc}$$

SB11 test score distribution between low–income and wealthy 2015-1 students

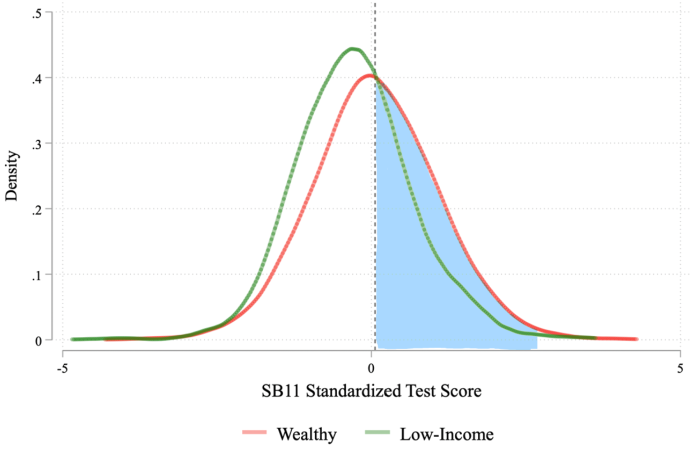


Figure: Low–income and wealthy high–achieving are in the shaded area

Interactions with the low–income by SB11 performance

	High Achievers according to SB11 scores	
	(1)	(2)
	A. Probability of a Link	B. Number of Links
*%tage of low–income peers * high achiever*	-0.000	-0.001
	(0.001)	(0.002)
%tage of low–income peers	0.007***	0.015***
	(0.002)	(0.003)
High Achiever by SB11	0.006	0.002
	(0.019)	(0.023)
No. Students	5,278	5,278

Wealthy students of above average SB11 test scores are no more likely to link with high achieving low–income peers

Interactions with the low–income by first–term performance

	A. Probability of a Link with a Low-Income High Achievers By:		B. Number of Links with Low-Income High Achievers By:	
	(1)	(2)	(3)	(4)
	First Term GPA	First Term Credits Attempted	First Term GPA	First Term Credits Attempted
*%tage of low–income peers * high achiever*	0.005***	0.002**	0.014***	0.007***
	(0.002)	(0.001)	(0.004)	(0.003)
%tage of low–income peers	0.004***	0.008***	0.011***	0.020***
	(0.001)	(0.001)	(0.003)	(0.003)
High Achiever by first term performance	0.024	0.044***	0.009	0.030
	(0.015)	(0.014)	(0.028)	(0.020)
No. Students	5,278	5,278	5,278	5,278

Wealthy students who are high achievers according to their *first–term results* are more likely to link with other low–income peers who are also high–achievers under the same criteria. Results are potentially endogenous

Assumptions and robustness checks

Discussion on Identification Assumptions

1. *Parallel trends assumption:* in the absence of SPP, the trends in the outcome should change the same for both treated and control groups.

2. *Measurement Error in the Outcomes:*
   - Measurement Error in turnstile–elicited interactions.
   - GPA is a noisy measure of academic performance $\rightarrow$ use other indicators like credits and persistence measures

3. *Unobserved Exposure Effects:* exposure through courses differs from exposure through majors–cohorts

1. Parallel Trends Assumption

Wealthy students change their entry major and cohort due to unobserved preferences for low–income peers

$\rightarrow$ If such is the case, observed characteristics should show changes.

$$Y_{imc} = \sum_{c=2012}^{c=C} \mu_{IC} R^{I}_{m,c=C} + \mathbf{X_i'} M + \mu_m + \mu_c + \varepsilon_{imc}$$

The relation between the dependent variable Y_{imc} and the share of low–income students in each entry cohort captured by $\hat{\mu}_{IC}$ should not differ in the pre– and after–SPP.

(If a line is drawn across estimates it should be flat) ▸ Identification Assumptions

No Changes in the Characteristics of Wealthy Students

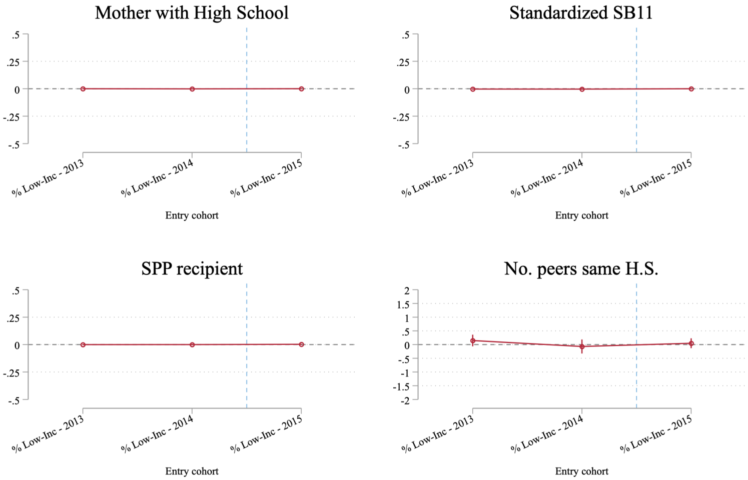


Figure: Variation in Students' Chars.

2. Measurement Error in the Outcomes: Turnstile-Elicited Interactions

Suppose we were able to observe who is friends with whom (i.e., links) in *real life*

A. **False Negatives:** links that exist in real life but are not captured by turnstile-elicited links,

B. **False Positives:** turnstile-elicited links that do not correspond to real life ones

2. Measurement Error in the Outcomes: Turnstile-Elicited Interactions

Suppose we were able to observe who is friends with whom (i.e., links) in *real life*

A. **False Negatives:** links that exist in real life but are not captured by turnstile-elicited links,

B. **False Positives:** turnstile-elicited links that do not correspond to real life ones

$\Rightarrow$ To approximate *real life* social links, I use data from a survey asking students about their networks

$\rightarrow$ Assumes survey-elicited links capture *real-life* links.

The survey was conducted among the Economics undergrads of Elite University from the 2017 entry cohort, and during their first term of college (Thank you to Cardenas et al. (2019) who provided access to the data!).

2. Measurement Error in the Outcomes: Turnstile-Elicited Interactions

1. Turnstile–elicited interactions are the result of random co–movements
2. If false–negatives are non–random, there will be bias in the turnstile–elicited links
3. Rate of false–positives and negatives across majors and entry cohorts could be impacted by the exposure to low–income peers.

2. Measurement Error in the Outcomes: Turnstile-Elicited Interactions

1. Turnstile–elicited interactions are the result of random co–movements

2. If false–negatives are non–random, there will be bias in the turnstile–elicited links

3. Rate of false–positives and negatives across majors and entry cohorts could be impacted by the exposure to low–income peers.

2. Measurement Error in the Outcomes: Turnstile-Elicited Interactions

1. Turnstile–elicited interactions are the result of random co–movements
2. If false–negatives are non–random, there will be bias in the turnstile–elicited links
3. Rate of false–positives and negatives across majors and entry cohorts could be impacted by the exposure to low–income peers.

2. Measurement Error in the Outcomes: Turnstile-Elicited Interactions

1. Turnstile–elicited interactions are the result of random co–movements
   - → I compare with the survey data. The rate of false–positives is below 10 percent, but the rate of false–negatives is over 60 percent. ▸ See comparison with survey
2. If false–negatives are non–random, there will be bias in the turnstile–elicited links
   - → Using simulations, I show turnstile–elicited interactions capture well links' characteristics, albeit the number of interactions missing. ▸ Links' simulations
3. Rate of false–positives and negatives across majors and entry cohorts could be impacted by the exposure to low–income peers.
   - → Use a 2x2 Dif-in-Dif framework and measurement error proxies to show exposure to low–income peers does not alter measurement error ▸ Dif-in-Dif Framework ▸ assumptions

3. Unobserved Exposure Effects

$$Y_i^{mc} = \rho_l IN_{imcs}^l + \rho_N N_{imcs} + \rho_s S_{imc} + \mathbf{X_i'} P + \rho_{mc} + \upsilon_{imc}$$

- IN_{imcs}^l Individual exposure to low–income peers. Number of low–income students enrolled in all the first–term courses s taken by the student i

- N_{imcs} Number of peers in all first term courses. Computed as IN_{imcs}^l but without conditioning on SES.

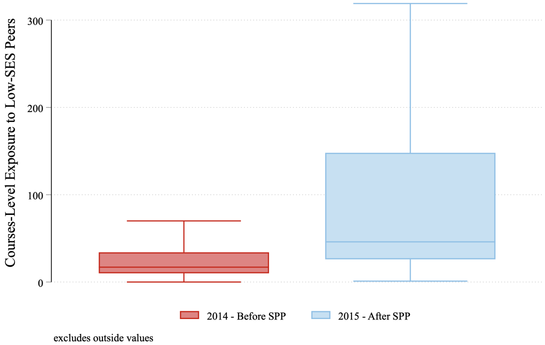
- S_{imc} Number of courses taken by the student.

- ρ_{mc} fixed effect by major and entry cohort. It absorbs shocks common to students in each major–cohort group.

▸ Assumptions

Individual–Level Exposure

Figure: Course-Level Exposure Index to Low–SES Peers in the First Term of Enrollment



Notes: This figure plots the distribution of the course–level exposure to low–SES peers in the first term. The plot follows the standard display of 75th percentile, median and 25th percentile references.

Endogeneity in Exposure within Major–Cohort

Identification of $\hat{\rho}_l$ is challenged by students selecting in courses in ways associated with achievement or socializing preferences but unobserved.

- Use data from 2012 and 2013 to predict the number of low–income students had the distribution not changed with SPP: $PIN_{ims}^{l,c<2014}$

First Stage:

$$IN_{imcs}^{l} = \mu_l PIN_{ims}^{l,c<2014} + \mu_N N_{imcs} + \mu S_{imc} + \mathbf{X_i'} M + \mu_{mc} + \upsilon_{imc}$$

Effects of course–level exposure on academic achievement

Table: Effect of Courses-Level Exposure to Low–SES Peers on Wealthy Students' Achievement

	(1)	(2)	(3)	(4)	(5)	(6)
	6th term cum. Credits	6th term cum. GPA	7th term cum. Credits	7th term cum. GPA	Graduated in 8 term	Graduated in 9 terms
2SLS						
IN_{imc}	0.023	-0.001	-0.003	-0.001	-0.002	0.001
	(0.071)	(0.001)	(0.079)	(0.001)	(0.002)	(0.002)
First Stage						
Predicted IN_{imc}	0.511***	0.511***	0.533***	0.533***	0.460***	0.574***
	(0.110)	(0.110)	(0.112)	(0.112)	(0.102)	(0.102)
F-test excluded instruments	21.74	21.74	22.52	22.52	20.28	31.92
Reduced Form						
Predicted IN_{imc}	0.012	-0.001	-0.002	-0.001	-0.001	0.000
	(0.036)	(0.001)	(0.043)	(0.001)	(0.001)	(0.001)
OLS						
IN_{imc}	-0.033***	-0.001**	-0.043***	-0.001***	-0.000	-0.000
	(0.012)	(0.000)	(0.013)	(0.000)	(0.000)	(0.000)
No. Students	4,507	4,507	4,447	4,447	5,278	4,027

Effects of course–level exposure on academic achievement

Table: Effect of Courses–level Exposure to Low–SES Peers on Wealthy Students' Friendships

	2 - seconds window			3 - seconds window			5 - seconds window		
	(1) Any	(2) Wealthy	(3) Low SES	(4) Any	(5) Wealthy	(6) Low SES	(7) Any	(8) Wealthy	(9) Low SES
2SLS									
IN_{imc}	0.001	-0.002	0.003	0.007	-0.001	0.008	0.014	0.003	0.011
	(0.018)	(0.015)	(0.006)	(0.024)	(0.019)	(0.008)	(0.025)	(0.020)	(0.008)
First Stage									
Predicted IN_{imc}	0.460***			0.460***			0.460***		
	(0.102)			(0.102)			(0.102)		
F-test excluded instruments	20.28	20.28	20.28	20.28	20.28	20.28	20.28	20.28	20.28
Reduced Form									
Predicted IN_{imc}	0.001	-0.001	0.002	0.003	-0.000	0.004	0.006	0.001	0.005
	(0.009)	(0.007)	(0.003)	(0.011)	(0.009)	(0.004)	(0.012)	(0.009)	(0.004)
No. Students	5,278	5,278	5,278	5,278	5,278	5,278	5,278	5,278	5,278

Summary of Results

1. Increased exposure to low–income peers did not affect wealthy students performance, but it led to more diverse social interactions. How?

1. Relative to the average increase in the share of low–income peers in a major and cohort (9.5 p.p.):
   a. Wealthy students increase their links with the low–income in their major and cohort in 0.3 links (120%relative to pre–SPP), and the probability in 8 p.p. (42% to pre–SPP)

   b. Marginal impact on links with other wealthy suggest expansion rather than substitution of wealthy’s friendships.

Summary of Results (Cont.)

3. Similarities on academic achievement may offset differences in socio–economic status
   c. Half of the effect on interactions with the low–income is explained by interactions with low–income who are high achievers

   d. Moreover, wealthy students with high–achievement in their first term are also more likely to match low–income high achievers

Conclusions

Conclusions

Exposure to more low–income peers did not hurt achievement and led to more interactions between wealthy and low–income peers

Why?

- Similarities among very high–achievers seem to partially offset differences in socio–economic status
  (Carrell et al., 2013; Baker et al., 2011; Marmaros and Sacerdote, 2006; others)

AA and financial aid policies can be effective at generating social integration when students find similarities in terms of achievement.

Potentially positive consequences for post–college outcomes if the effects of 'Elite Boys Clubs' are lessened (Zimmerman and co–authors, 2019, 2021)

Thank you!

tv2225@tc.columbia.edu

Thank you!

tv2225@tc.columbia.edu

SCAN ME

Appendix

Policy Variation: SPP at Elite University

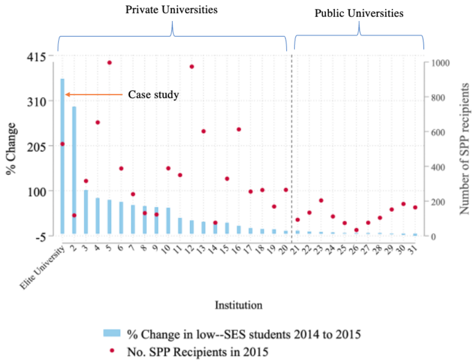


Figure: Low–SES students at SPP Eligible Colleges 2014-15

Note: Figures based on data from the impact evaluation report of the SPP program (Alvarez et al., 2017) and public records on college–level enrollment (SPADIES, 2020)

SPP at Elite University

Testing Crowding–Out of Wealthy Students [NEW!]

Table: Relation between number of wealthy students and number of low–income students in each major–cohort

	(1)	(2)	(3)	(4)
No. of low-income peers	1.117***	1.747***	-0.179	-0.222
	(0.319)	(0.408)	(0.196)	(0.212)
Average of student characteristics		x	x	x
Major Fixed Effects			x	x
Entry Cohort Fixed Effects				x
No. of major-cohort groups	124	124	124	124

Initial correlation is positive. Likely capturing size of major (i.e., larger majors tend to receive more low-income students than smaller majors)

Once major F.E. are included, correlation disappears. ▸ Policy Shock at Elite University

Proportion of low–SES students in each major increases

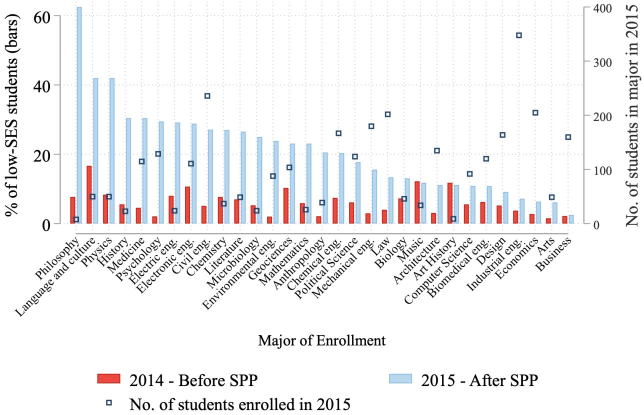


Figure: Changes in Major and Entry Cohort Student Composition After SPP

▸ Variation

Validation of Turnstile–Elicited Links Definition

Table: Rate of false positives and negatives of turnstile-elicited links relative to survey-elicited links

Time window	A. two seconds			B. three seconds			C. Five seconds		
Frequency per term	One	Two	Three	One	Two	Three	One	Two	Three
1. Turnstiles									
No. Of links found	868	368	235	1209	509	314	1906	898	552
No. of students in links	110	110	108	110	110	109	110	110	109
2. Acquaintances (Survey)									
Links		*1033*			*1033*			*1033*	
Survey & Turnstiles									
Matched	497	311	219	606	391	284	734	537	425
False Negatives	0.52	0.70	0.79	0.41	0.62	0.73	0.29	0.48	0.59
False Positives	0.36	0.06	0.02	0.58	0.11	0.03	1.13	0.35	0.12
Sum	*0.88*	***0.75***	*0.80*	*1.00*	***0.74***	*0.75*	*1.42*	*0.83*	***0.71***

Survey was conducted among 113 Economics students from the 2017 entry cohort. Among other questions, it asked who were the student's friends and acquaintances in their majors and cohort.

Validation of Turnstile–Elicited Links Definition

Table: Rate of false positives and negatives of turnstile-elicited links relative to survey-elicited links

Time window	A. two seconds			B. three seconds			C. Five seconds		
Frequency per term	One	Two	Three	One	Two	Three	One	Two	Three
1. Turnstiles									
No. Of links found	868	368	235	1209	509	314	1906	898	552
No. of students in links	110	110	108	110	110	109	110	110	109
2. Acquaintances (Survey)									
Links		*1033*			*1033*			*1033*	
Survey & Turnstiles									
Matched	497	311	219	606	391	284	734	537	425
False Negatives	0.52	0.70	0.79	0.41	0.62	0.73	0.29	0.48	0.59
False Positives	0.36	0.06	0.02	0.58	0.11	0.03	1.13	0.35	0.12
Sum	*0.88*	***0.75***	*0.80*	*1.00*	***0.74***	*0.75*	*1.42*	*0.83*	***0.71***

I compute turnstile–elicited links under different co–movements time–windows and different frequencies in a term

Validation of Turnstile–Elicited Links Definition

Table: Rate of false positives and negatives of turnstile-elicited links relative to survey-elicited links

Time window	A. two seconds			B. three seconds			C. Five seconds		
Frequency per term	One	Two	Three	One	Two	Three	One	Two	Three
1. Turnstiles									
No. Of links found	868	368	235	1209	509	314	1906	898	552
No. of students in links	110	110	108	110	110	109	110	110	109
2. Acquaintances (Survey)									
Links		*1033*			*1033*			*1033*	
Survey & Turnstiles									
Matched	497	311	219	606	391	284	734	537	425
False Negatives	0.52	0.70	0.79	0.41	0.62	0.73	0.29	0.48	0.59
False Positives	0.36	0.06	0.02	0.58	0.11	0.03	1.13	0.35	0.12
Sum	*0.88*	***0.75***	*0.80*	*1.00*	***0.74***	*0.75*	*1.42*	*0.83*	***0.71***

I keep the turnstile-elicited links for the students in the survey

Validation of Turnstile–Elicited Links Definition

Table: Rate of false positives and negatives of turnstile-elicited links relative to survey-elicited links

Time window	A. two seconds			B. three seconds			C. Five seconds		
Frequency per term	One	Two	Three	One	Two	Three	One	Two	Three
1. Turnstiles									
No. Of links found	868	368	235	1209	509	314	1906	898	552
No. of students in links	110	110	108	110	110	109	110	110	109
2. Acquaintances (Survey)									
Links		*1033*			*1033*			*1033*	
Survey & Turnstiles									
Matched	497	311	219	606	391	284	734	537	425
False Negatives	0.52	0.70	0.79	0.41	0.62	0.73	0.29	0.48	0.59
False Positives	0.36	0.06	0.02	0.58	0.11	0.03	1.13	0.35	0.12
Sum	*0.88*	***0.75***	*0.80*	*1.00*	***0.74***	*0.75*	*1.42*	*0.83*	***0.71***

I compare both types of elicited links. Turnstile-elicited interactions are unlikely to be a false–negative ($\sim$ 10% chance), but are likely to miss survey–elicited interactions ($\sim$ 60 %). Measurement Error

Assessing Size of Measurement Error

Rate of False–Positives is low: links captured using turnstiles are very likely to be as good as survey–elicited links.

Assessing Size of Measurement Error

Rate of False–Positives is low: links captured using turnstiles are very likely to be as good as survey–elicited links.

BUT!, rate of False–Negatives is large: many links that would be captured by a survey are not captured as a turnstile–elicited link.

- This is an issue to the extent that those I do capture are not representative of the survey-elicited links.

Assessing Size of Measurement Error

Rate of False–Positives is low: links captured using turnstiles are very likely to be as good as survey–elicited links.

BUT!, rate of False–Negatives is large: many links that would be captured by a survey are not captured as a turnstile–elicited link.

- This is an issue to the extent that those I do capture are not representative of the survey-elicited links.
- Turnstile–elicited links that form at random can introduce bias

Assessing Size of Measurement Error

Rate of False–Positives is low: links captured using turnstiles are very likely to be as good as survey–elicited links.

BUT!, rate of False–Negatives is large: many links that would be captured by a survey are not captured as a turnstile–elicited link.

- This is an issue to the extent that those I do capture are not representative of the survey-elicited links.
- Turnstile–elicited links that form at random can introduce bias

To assess this, I compare whether turnstile-elicited links plausibly reflect survey-elicited network characteristics using the survey sample.

Assessing Size of Measurement Error

Rate of False–Positives is low: links captured using turnstiles are very likely to be as good as survey–elicited links.

BUT!, rate of False–Negatives is large: many links that would be captured by a survey are not captured as a turnstile–elicited link.

- This is an issue to the extent that those I do capture are not representative of the survey-elicited links.
- Turnstile–elicited links that form at random can introduce bias

To assess this, I compare whether turnstile-elicited links plausibly reflect survey-elicited network characteristics using the survey sample.

- I simulate a fully random assignment of turnstile–elicited links 500 times, plot the distribution of the characteristics of said links, and compared them with the observed turnstile– and survey–elicited links

Turnstile-elicited links capture network characteristics

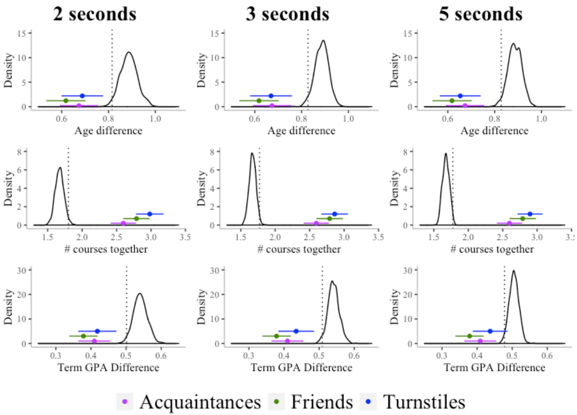


Figure: Average survey- and turnstile-elicited characteristics against distribution of characteristics in randomly-elicited network

Measurement error

Measurement Error in the Diff–in–Diff Context

Define L^{true} as the number of links of each student

$$L^{true} = L^{obs} - L^{F(+)} + L^{F(-)} \tag{1}$$

Measurement Error in the Diff–in–Diff Context

Define L^{true} as the number of links of each student

$$L^{true} = L^{obs} - L^{F(+)} + L^{F(-)} \tag{1}$$

Use 2x2 Dif–in–Dif potential outcomes framework (Goodman–Bacon, 2019; Cunninghan, 2021)

Measurement Error in the Diff–in–Diff Context

Define L^{true} as the number of links of each student

$$L^{true} = L^{obs} - L^{F(+)} + L^{F(-)} \tag{1}$$

Use 2x2 Dif–in–Dif potential outcomes framework (Goodman–Bacon, 2019; Cunninghan, 2021)

$$\begin{aligned} ATT^{estimated} &= E[L_t^{1,True}|Post] - E[L_t^{0,True}|Post] + \text{non-parallel trends bias (=0)} \\ &= E[L_t^{1,obs} - L_t^{1,F(+)} + L_t^{1,F(-)}|Post] - E[L_t^{0,obs} - L_t^{0,F(+)} + L_t^{0,F(-)}| \\ &= \underbrace{E[L_k^{1,obs}|Post] - E[L_k^{0,obs}|Post]}_{\text{Observed ATT}} + \\ &\underbrace{E[L_t^{1,F(-)} - L_t^{1,F(+)}|Post] - E[L_t^{0,F(-)} - L_t^{0,F(+)}|Post]}_{\text{Measurement Error Bias}} \end{aligned}$$

Measurement Error in Diff–in–Diff Context

After re-arraigning the terms, the estimated Average Treatment on the Treated in the 2x2 Difference-in-Difference framework is:

$$ATT^{estimated} = ATT^{obs} + \underbrace{E[L_t^{1,F(-)} - L_t^{0,F(-)}|Post]}_{\text{ATT on F(-)}} - \underbrace{E[L_t^{1,F(+)} - L_t^{0,F(+)}|Post]}_{\text{ATT on F(+)}} \tag{2}$$

Implications:

1. If treatment does not impact $L^{F(-)}$ or $L^{F(+)}$ among the treated, then $ATT = ATT^{obs}$,
2. If $ATT^{F(-)} = ATT^{F(+)}$, then measurement error cancels out.

Test if treatment impacts $L^{F(-)}$ or $L^{F(+)}$

Estimate ATT on proxies of $L^{F(-)}$ and $L^{F(+)}$.

Test if treatment impacts $L^{F(-)}$ or $L^{F(+)}$

Estimate ATT on proxies of $L^{F(-)}$ and $L^{F(+)}$.

$\rightarrow$ **total ID taps:** more taps could yield a higher chance of $L^{F(+)}$. Similarly, fewer taps could yield a higher chance of $L^{F(-)}$

Test if treatment impacts $L^{F(-)}$ or $L^{F(+)}$

Estimate ATT on proxies of $L^{F(-)}$ and $L^{F(+)}$.

→ **total ID taps:** more taps could yield a higher chance of $L^{F(+)}$. Similarly, fewer taps could yield a higher chance of $L^{F(-)}$

→ **No. of courses with turnstile-elicited links:** More courses with links may indicate a higher chance of $L^{F(+)}$. turnstile-elicited link is the product of a coincidence.

ATT on Measurement Error on Proxies is marginal and small

Table: ATT on Measurement Error

		No. of courses with peers who are links		
	ID taps	2 seconds	3 seconds	5 seconds
	(1)	(2)	(3)	(4)
percentage of low–income peers	-4.162 (2.691)	0.003 (0.007)	-0.001 (0.007)	0.000 (0.007)
Covariates	x	x	x	x
Cohort FE	x	x	x	x
Major FE	x	x	x	x
No. of Students	5,278	5,278	5,278	5,278

I do not find evidence of a change in the number of courses with turnstile-elicited links. ▸ Measurement error

Variation in turnstile-elicited interactions measurement criteria

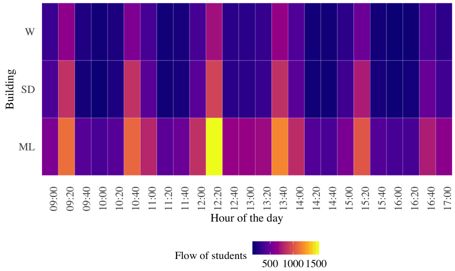


Figure: Average Student ID Taps During Week Days on Three of the Entrances

How susceptible are turnstile-elicited interactions to the time of the day when they are measured?

Comparing survey-elicited interaction against turnstile-elicited interactions during lunch time

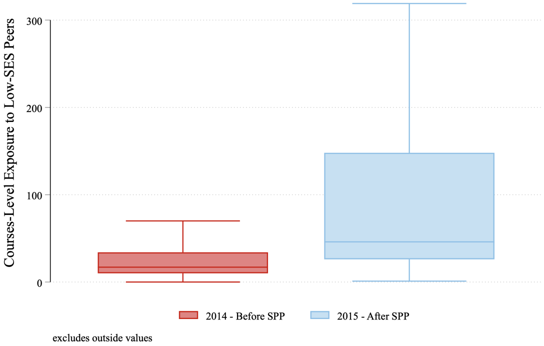
Table: False Positives and Negatives Rates by Lunch-Times

Time window	A. Two seconds						B. Three 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	397	159	100	654	272	172	554	213	135	893	376	233
No. of students	110	109	103	110	110	105	110	109	106	110	110	106
2. Acquaintances												
Dyads	*1033*						*1033*					
Survey & Turnstiles												
Matched	255	143	93	411	236	162	321	180	123	494	308	214
False Negatives	0.75	0.86	0.91	0.60	0.77	0.84	0.69	0.83	0.88	0.52	0.70	0.79
False Positives	0.14	0.02	0.01	0.24	0.03	0.01	0.23	0.03	0.01	0.39	0.07	0.02
Sum	*0.89*	***0.88***	*0.92*	*0.84*	***0.81***	*0.85*	*0.91*	***0.86***	*0.89*	*0.91*	***0.77***	*0.81*

False–positives would decrease, but rate of false–negative would increase to over 80 percent!

Individual–Level Exposure

Figure: Course-Level Exposure Index to Low–SES Peers in the First Term of Enrollment



Notes: This figure plots the distribution of the course–level exposure to low–SES peers in the first term. The plot follows the standard display of 75th percentile, median and 25th percentile references.

Instrumenting with the Predicted Exposure to low–SES

I use University data on course–level enrollment from 2012 and 2013 to estimate a distribution of low–SES students across courses and to predict the number of low–SES students in each course of the 2014 and 2015 cohort, had the distribution of low–SES students not changed due to the outset of SPP

Figure: Observed vs. Predicted Index of Course–Level Exposure to Low–SES Peers

threats