80
Primer semestre 2018
PP. 53-88, ISSN 0120-3584
E-ISSN 1900-7760

It's not what you know but who you know: Heterogeneous peer effects at a Colombian University

No es lo que sabes sino a quién conoces: efectos de pares heterogéneos en una universidad colombiana

Ana María Díaz[1],
Ignacio Penagos[2]

DOI: 10.29263/DYS.80.2

Abstract

This paper uses matched survey and administrative data on first-year Economics students who were studying at the Javeriana University in Bogotá, Colombia, in 2015 in order to estimate peer effects on student grades. We employ the strategy proposed by De Giorgi, Pellizzari & Redaelli (2010) to identify and estimate these peer effects. Our results show that peer effects are economically significant in their context, that they result from the sharing of specific rather than general skills among peers, and that they flow mainly from peers with whom students interact frequently and who are considered to be leaders.

1 Profesora asociada, Departamento Economía, Pontificia Universidad Javeriana.
E-mail: a.diaze@javeriana.edu.co.

2 Profesor instructor, Departamento de Economía, Pontificia Universidad Javeriana.
E-mail: ipenagos@ javeriana.edu.co.

Este artículo fue recibido el 28 de octubre del 2016, evaluado el 16 de junio del 2017 y finalmente aceptado el 23 de octubre del 2017.

Author's key words: Peer effects, social network formation, academic achievement, homophily

JEL Classification: D85, I21, I23, I26, J24

Resumen

Este artículo combina encuestas y datos administrativos sobre estudiantes de primer año de Economía en 2015 de la Universidad Javeriana en Colombia para estimar los efectos de pares en las calificaciones de los estudiantes. Empleamos la estrategia propuesta por De Giorgi *et al.* (2010) para identificar y estimar los efectos de pares. Los resultados muestran que los efectos de pares son económicamente significativos en su contexto, que resultan del intercambio de habilidades específicas en lugar de intercambio de habilidades generales entre pares, y que provienen principalmente de pares con quienes los estudiantes interactúan con frecuencia y que son vistos como líderes.

Palabras clave del autor: efectos entre iguales, formación de redes sociales, logros académicos, homofilia.

Clasificación JEL: D85, I21, I23, I26, J24.

Introduction

The influence of peers on higher educational outcomes has been extensively studied. Recent research has argued that peer interactions play an important role in the academic performance of students. However, two important questions remain unanswered. First, the mechanism through which peers affect education is unclear: There is no obvious economic incentive to help peers. Second, few studies on heterogeneous effects take the types of links between individuals into account. This paper addresses both of these issues by exploring the influence on the grades of first-year Economics students at a Colombian university of peer effects by exploiting their network structure.

Limited by the theory, methodology and related variables chosen for analysis, the empirical literature often finds evidence of different peer effects. For example, an empirical study of a quasi-random assignment of students to room and dormitory groups found that peer effects are relevant and significant (Sacerdote, 2001). However, empirical support for significant positive effects is not universal. Zimmerman (2003) finds, for example, that student performance declines if the assigned roommate comes from the bottom of the ability distribution. Finally, neither Foster (2006); McEwan, Soderberg & Kristen (2006) nor Carrell, Sacerdote & West (2011) find evidence that the background and current achievements of their peers affect the academic achievement of students.

Recent studies have shown that peer effects cannot be generalised, as peers might affect some subgroups more than others. Studies of heterogeneous peer effects have examined whether different types of individuals affect individual outcomes (e.g. male vs. female, highly skilled vs. less skilled). For example, using data from a Chinese college, Han & Li (2009) found evidence that females respond to peer influences, while males do not. Hoxby & Weingarth (2005) concluded that if students are initially very high achieving and their classroom has high median ability, then they benefit most from peers who are also very highly achieving. Carrell et al. (2011) find a negative and statistically significant peer effect for the lowest ability students, and a positive and significant peer effect for their middle ability counterparts. Finally, Brady, Insler & Rahman (2015) found that in large social settings, more favorable average peer attributes can, perversely, lower individual performance, while in settings where individuals are engaged in common work tasks, positive peer effects arise.

These varied findings may be due either to the presence of different types of endogeneity or the choice of empirical methodology employed in each analysis. For example, the linear-in-means model, which is frequently employed in the peer effects literature, augments regressions that include individual-level covariates using means of group-level characteristics (Sacerdote, 2011, provides a survey of this literature). Significant coefficients of group-level variables are assumed to indicate the presence of peer effects. The problem with this approach, as Manski (1993) points out, is that such models might be biased by the presence of reflection, correlated, and multiplier effects. Recent studies have addressed these problems by using randomised controlled trials on both group and treatment assignments (Sacerdote, 2001; Duflo & Saez, 2003; Babcock, Bedard, Charness, Hartman & Royer, 2015), instrumental variables (IV) techniques (Bramoullé, Djebbari & Fortin, 2009; De Giorgi et al., 2010), and other quasi-experimental methods.

We explore the presence of peer effects among first-year Economics students at the Javeriana University in Bogotá, Colombia, using a unique dataset that combines network data with university administrative records. We carried out an online survey to collect network data from students in four "Introduction to Economics" classes. Our dataset thus represents a directed network of contacts between students and describes the types of relationships between them. We also collected individual socio-economic information for each student.

Methodologically, we adhere to the strategy proposed by De Giorgi et al. (2010) in order to identify and estimate endogenous peer effects (i.e., the impact of average peer outcomes on individual outcomes). Our empirical strategy exploits a common feature of social networks: the existence of partially overlapping groups of peers. As De Giorgi et al. (2010) explain, "partially overlapping groups generate peers of peers who act as exclusion restrictions in the simultaneous equation model". Specifically, our strategy uses the exogenous characteristics of peers outside the group as exclusion restrictions. These instruments are valid and relevant because they are correlated with the performance of peers but uncorrelated with the individual group shock.

The paper also explores the mechanism lying behind our results. In particular, we seek to disentangle specific vs. general transmission of knowledge between peers. In addition, we analyse heterogeneous effects involving different types of individuals and the different kinds of links between them.

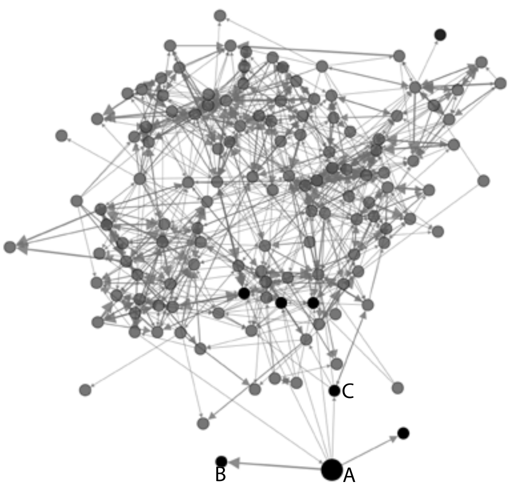
We find that in higher education there are strong peer effects for students who frequently study together. The results show heterogeneous effects of initial academic ability (measured by university entrance exam scores), and suggest that peer effects work through specific rather than general knowledge. We find that the academic performance of low-skilled friends has a negative and significant impact on individual educational outcomes, while the academic performance of medium- and high-skilled friends has a positive and significant effect. In other words, the level of friends' abilities has a positive and significant peer effect. We also find that the academic achievement of the leader has a positive and significant effect on the educational outcomes of other students.

The next section describes the data. The third section details the empirical methodology, while the fourth section presents the results, and the last section draws conclusions.

I. Description of the Data

We used a unique dataset combining network data with the administrative records of first-year Economics students (n = 132) at the Javeriana University who were enrolled in the 2015 "Introduction to Economics" class. Every student in the class was asked to complete an online survey about their study partners, which involved nominating up to five classmates with whom they had studied Economics during the previous month.[3] This information permitted us to reconstruct the geometric structure of the entire study network.

Figure 1. The Study Network



Note: "A" denotes Antonio. Dark spots represent Antonio's peers (with whom A reported having studied during the previous month). The size and darkness of the connecting arrow represents how frequently they studied together. For example, A studied more frequently with B than with C.

Source: Authors' calculations based on student study networks.

As students reported exactly who they study with, we were able to measure the peer group more precisely than previous studies -which used peer groups that were fixed across individuals- had been able to do. A typical example of fixed peer groups is the classroom (i.e., if two students are in the same classroom, then both are assigned the same group of peers). In our case, each peer

3 The survey was carried out half way through the academic semester. In Colombia, each course lasts one semester.

group was individual-specific, and this seems to be a better way of assessing peer effects, because students may study with friends in different classes, a factor that would pose serious identification problems if a fixed group were used. Indeed, Figure 2 shows that students do not restrict their networks to friends in the same economics class.

Figure 2. Network by Economics Class

Students were randomly assigned to four classrooms. Colours represent the classroom that each student belongs to.

Source: Authors' calculations.

We also collected information about each student's individual and household attributes. Individual characteristics included age, gender, the mark obtained in the university entrance exam (known as SABER 11)[4], a dummy for recipients of a government scholarship known as *Ser Pilo Paga* (Hard Work Pays)[5],

4 This examination measures the level of skills developed by students during high school. The exam is composed of a core section and a flexible section. The core section includes language, mathematics, biology, chemistry, physics, philosophy, social sciences, and English. The flexible section evaluates the students' self-selected subject, according to their interests in the following areas: Spanish language, mathematics, natural sciences, a foreign language, social sciences, philosophy, and others.

5 This scholarship aims to help students from low-income backgrounds to access high-quality tertiary education. Students receive a grant that covers their full tuition and living costs. To be eligible, students must have achieved marks above a defined level in their SABER 11 test, and their household must be considered poor.

place of birth, and a dummy for students reporting high levels of self-confidence. Household attributes included maximum parental education, place of residence (i.e., locality fixed effects), and *estrato social*, or social stratum.[6]

Table 1 presents the descriptive statistics for the 87 students for whom we were able to match information. For this subsample, Grade Point Averages (GPA) for the economics class ranged from 0–5 (average 3.15). During the same semester under examination, students were required to enroll in a mathematics class, in which they studied with a different set of classmates. The GPA for the mathematics class was 3.39. The average age in this sample was 17; 42% were female, 66% had been born in Bogotá, 84% described themselves as having high self-esteem, and 65% had parents both of whom had completed at least tertiary education. The students lived in many different areas of the city, 40% of them in the localities of Usaquén and Suba, which are mainly strata 3 and 4.

Our empirical strategy also exploited two unique features of the composition of the classes. First, during the induction week, before teaching started, students were randomly assigned to four groups. The induction week features activities designed to help students meet their peers and learn about the university. Second, students were randomly assigned to their Introduction to Economics classes. Table 2 show that the students were equally distributed between the four economics classes and the four induction week groups.

Tables 3 and 4 present the descriptive statistics of the attributes, differentiated by the groups to which each student was assigned. Appendix Tables A.1. and A.2. present the results of the evaluation of covariate adjustment using bivariate linear regression. The results confirm that there were no important differences across induction week groups. In other words, these groups are balanced in terms of observable characteristics. There were statistical differences between the levels of parental education of the members of the Introduction to Economics groups, in particular the second and fourth groups.

6 The *estrato* (stratum) is the official classification of the socioeconomic characteristics of Colombian households. The system classifies areas on a scale from 1 to 6, with 1 as the lowest income area and 6 as the highest.

Table 1. Descriptive Statistics

Variable	Mean	(SD)
Economics Achievement	3.15	(0.75)
Maths Achievement	3.39	(0.87)
Economic Achievement of Peers	3.28	(0.50)
Entry Exam (Standardized)	0.04	(1.04)
Age	17.20	(2.20)
Female	0.42	(.50)
Born in BogotÃ¡	0.66	(0.48)
Confidence	0.84	(0.37)
Parental Maximal Education		
Primary	0.14	(0.35)
Secondary	0.09	(0.29)
Tecnical	0.11	(0.32)
Tertiary	0.30	(0.46)
Postgraduate	0.35	(0.48)
Locality	10.59	(4.52)
Stratum	3.56	(1.25)
N	87	
[.75ex]		

Source: Authors’ calculations based on 2015 administrative data on first-year Economics students at the Javeriana University.

Table 2. Student distribution between groups

Variable	Mean	(SD)
Induction Week Group		
1	0.29	(0.46)
2	0.29	(0.46)
3	0.23	(0.42)
4	0.19	(0.39)
Economics Classroom		
1	0.28	(0.45)
2	0.24	(0.43)
3	0.22	(0.41)
4	0.27	(0.44)
N	87	
[.75ex]		

Source: Authors’ calculations based on 2015 administrative data on first-year Economics students at the Javeriana University.

Table 3. Descriptive Statistics by Induction Week Group

	(Induction Week Groups)				
	1	2	3	4	Total
Economics Achievement	3.308 (0.664)	2.843 (0.605)	3.116 (0.900)	3.405 (0.800)	3.147 (0.753)
Economic Achievement of Peers	3.269 (0.488)	3.144 (0.414)	3.290 (0.516)	3.515 (0.577)	3.284 (0.499)
Entry Exam (Standardised)	0.0925 (1.053)	0.0308 (1.044)	-0.0160 (1.109)	0.0549 (1.042)	0.0427 (1.042)
Age	17.78 (1.347)	17.39 (0.941)	17.22 (0.808)	16 (4.456)	17.20 (2.204)
Female	0.261 (0.449)	0.522 (0.511)	0.389 (0.502)	0.533 (0.516)	0.418 (0.496)
Born in BogotÃ¡	0.565 (0.507)	0.609 (0.499)	0.833 (0.383)	0.667 (0.488)	0.658 (0.477)
Confidence	0.870 (0.344)	0.696 (0.470)	0.889 (0.323)	0.933 (0.258)	0.835 (0.373)
Primary	0.174 (0.388)	0.261 (0.449)	0.0556 (0.236)	0 (0)	0.139 (0.348)
Secondary	0.0870 (0.288)	0.174 (0.388)	0 (0)	0.0667 (0.258)	0.0886 (0.286)
Technical	0.130 (0.344)	0.0435 (0.209)	0.222 (0.428)	0.0667 (0.258)	0.114 (0.320)
Tertiary	0.391 (0.499)	0.174 (0.388)	0.333 (0.485)	0.333 (0.488)	0.304 (0.463)
Postgraduate	0.217 (0.422)	0.348 (0.487)	0.389 (0.502)	0.533 (0.516)	0.354 (0.481)
Locality	11.65 (4.376)	9.087 (4.747)	10.50 (4.342)	11.40 (4.388)	10.59 (4.522)
Stratum	3.739 (1.514)	3.043 (1.022)	3.667 (0.840)	3.933 (1.387)	3.557 (1.248)
[.75ex]					

Source: Authors' calculations based on 2015 administrative data on first-year Economics students at the Javeriana University.

Table 4. Descriptive Statistics by Economics Class

	(1)				
	1	2	3	4	Total
Economics Achievement	2.885 (0.638)	2.955 (0.831)	3.366 (0.637)	3.419 (0.779)	3.147 (0.753)
Economic Achievement of Peers	3.081 (0.386)	3.026 (0.443)	3.430 (0.371)	3.611 (0.539)	3.284 (0.499)
Entry Exam (Standardised)	-0.165 (1.046)	-0.0698 (1.127)	0.359 (0.965)	0.106 (1.022)	0.0427 (1.042)
Age	17.41 (1.182)	17.21 (0.918)	17.59 (1.176)	16.67 (3.890)	17.20 (2.204)
Female	0.364 (0.492)	0.579 (0.507)	0.353 (0.493)	0.381 (0.498)	0.418 (0.496)
Born in BogotÃ¡	0.727 (0.456)	0.684 (0.478)	0.529 (0.514)	0.667 (0.483)	0.658 (0.477)
Confidence	0.864 (0.351)	0.737 (0.452)	0.882 (0.332)	0.857 (0.359)	0.835 (0.373)
Primary	0.136 (0.351)	0.368 (0.496)	0.0588 (0.243)	0 (0)	0.139 (0.348)
Secondary	0.136 (0.351)	0 (0)	0.0588 (0.243)	0.143 (0.359)	0.0886 (0.286)
Tecnical	0.227 (0.429)	0 (0)	0.118 (0.332)	0.0952 (0.301)	0.114 (0.320)
Tertiary	0.273 (0.456)	0.158 (0.375)	0.353 (0.493)	0.429 (0.507)	0.304 (0.463)
Postgraduate	0.227 (0.429)	0.474 (0.513)	0.412 (0.507)	0.333 (0.483)	0.354 (0.481)
Locality	10.32 (4.236)	10.79 (4.077)	11.12 (4.833)	10.29 (5.178)	10.59 (4.522)
Stratum	3.500 (1.225)	3.526 (1.307)	3.471 (1.328)	3.714 (1.231)	3.557 (1.248)
[.75ex]					

Source: Authors' calculations based on 2015 administrative data on first-year Economics students at the Javeriana University.

II. Empirical Strategy

In the linear-in-means model, outcome Y is a linear function of a student's background characteristics, the average background characteristics of peers, and peers' average outcomes. The average variables of peers are constructed as the average of self-declared peers. More formally, this can be written as:

$$Y_i = \alpha_0 + \alpha_1 \overline{Y}_{-i} + \alpha_2 X_i + \alpha_3 \overline{X}_{-i} + \in_i \tag{1}$$

where Y_i represents the student's final mark and X_i is a vector of their background characteristics. Each student *i* declares having a specific peer group P_i of size n_i (up to five friends). This reference group contains at most five students whose academic activities or family background might affect *i*'s academic activity. Thus, $\overline{Y}_{-i}$ is measured as *i*-peer's average outcome $\left(i.e., \frac{\sum_{j \in P_i} Y_j}{n_i} \right)$, and is a vector of her peers' average background characteristics $\left(i.e., \frac{\sum_{j \in P_i} X_j}{n_i} \right)$. According to this set-up, α_1 measures the endogenous effect, and α_3 the exogenous effects.[7]

The literature recognises at least three reasons why the linear-in-means model depicted in Eq. (1) might be biased. First, since student *i*'s outcome (Y_i) affects the mean outcome of their peers $\left(, \overline{Y}_{-i}\right)$ and vice versa, is subject to endogeneity bias. In the literature, this is known as the reflection problem. Second, students may self-select into peer groups based on both observed and unobserved attributes. Positive selection, in which similar people tend to associate, occurs frequently. This positive selection might cause upward bias on both endogenous and exogenous peer effects (i.e, α_1 and α_3). Finally, there might be a correlated effect because peer background itself affects peer outcomes.

In our set-up, peer groups were not fixed across individuals (that is, if *i* and *j* were in the same peer group, then the two groups of peers were the same),

7 Initially, in $\overline{Y}_{-i}$ every peer enters with the same weight, but subsequently we use frequency of study to give a higher weighting to those that *i* declares to interact more and a lower weight to those s/he declares to interact with less. It is also important to clarify that the network-related results apply only to students who have chosen to study with others. However, in our case, all the students declare to have done so.

but were individual-specific in that they were created from each student's network description (i.e., peer groups did not fully overlap). As De Giorgi et al. (2010) and Bramoullé et al. (2009) show, using individual-specific groups solves the reflection problem (see Appendix 1 for details). Therefore, we only need to address the two remaining sources of bias: self-selection and correlated effects. On this, De Giorgi et al. (2010) argue that if individual-specific peer groups are randomly assigned, there is no need to overcome the self-selection bias, and that using excluded peers as an instrumental variable may overcome the potential bias caused by correlated effects. In other words, the identification of peer effects required a data structure that satisfied three criteria: (1) an individual-specific network (to solve the reflection problem), (2) a random allocation into that network (to solve the self-selection problem), and (3) excluded peers (to resolve the correlated effect problem). We used peer groups that were individual-specific, guaranteeing the existence of the excluded peer (thereby satisfying criteria 1 and 3). Unfortunately, students are not randomly assigned to their networks, and are therefore able to choose their own peer groups (P_i in this setting), and it was perfectly possible that would they sort into groups of individuals who share the same unobserved individual attributes, even if randomly assigned to larger groups. There might, therefore, be a correlation between the individual effect and any endogenous or exogenous effects.

To explore whether positive selection had occurred by analyzing which attributes, if any, affected the probability of a given student being named as a peer. In other words, we were able to assess homophily between first-year students. We used a multiple regression quadratic assignment procedure (MR QAP) to explore the extent to which individual attributes influenced the formation of study networks. The unit of analysis used in the MR QAP is a dyad (i.e., a pair of individuals who may or may not have some sort of connection). The dependent variable is W_{ij}, indicating the network relationship between nodes *i* and *j*, which is equal to one if student *i* names student *j* as belonging to their study group. This variable was regressed on the set of attributes of students *i* and *j*. For example, in order to explore whether gender homophily was present, the following regression would be run:

$$P\left(W_{ij}=1\middle|x_{ij}\right)=G\left(\alpha+\beta x_{ij}\right) \tag{2}$$

The probability that *i* will nominate *j* as a study peer, $y_{ij} = 1$, is a (non-linear) function of x_{ij}, where x_{ij} is equal to one if *i* and *j* are the same sex, and zero otherwise. We used a probit model to estimate the associated parameters; *G*(.) is the normal Cumulative Distribution Function. However, Freeman (1979) has shown that it is incorrect to calculate the standard error of these coefficients. To overcome this problem, our technique repeatedly permuted rows and columns of the matrix representing the dependent variable and recomputed the regression coefficients after each permutation. This step was repeated 500 times in order to determine whether any unbiased standard errors were present (see Freeman, 1979, for more details). There is evidence of positive gender-homophily whenever β is positive and significant. In such cases, there is a non-random association that might lead to biased results for Eq. (1). We argue that in the absence of random assignment, we can overcome the self-selection bias by including the gender variable in the linear-in-means model.

We extended the Eq. (2) model to include all attributes that students might use to sort into study groups by estimating the following model:

$$P\left(W_{ij} = 1 \middle| X_{ij}\right) = G\left(X_{ij}\beta\right) \tag{3}$$

where X_{ij} is a vector of the individual, household, and classroom attributes. The results, discussed below, show that the variables affecting the probability that *i* will name *j* as a study partner are that both students: belong to the same induction group and the same Introduction to Economics and mathematics classes, and that they were born in the same place. We therefore included a set of information on induction group and class fixed effects and birthplace information in Eq. (1) in order to minimise the bias that might arise from self-selection into peer groups even if students are randomly assigned to their induction groups and to their economics and mathematics classes.

Finally, unobservable group shocks that induce endogeneity could still be present. De Giorgi et al. (2010) propose that a possible solution is to use instrumental variables, arguing that this naturally offers valid instruments, namely peers of peers who are not members of *i*'s peer group. The intuition is that the attributes of students who are excluded from *i*'s peer group, but included in the group of one or more of *i*'s peers, are uncorrelated with the group fixed effect of *i* and correlated with the mean outcome of *i*'s group through endogenous interactions.

Identification requires that the only channel through which excluded peers can affect student outcomes is through their academic results. Even though this seems plausible in the case of our set-up, we sought to minimise any kind of relationship by using the attributes of the peer's of peer's of peer's, and assumed that the strength of the interactions within the network declined with distance. The attributes used as exclusion restrictions were the average of the following variables among the excluded peers: standardised test results, gender, and place of birth.

IV. The Mechanism

Peers might affect the academic achievement of other students in different ways. As mentioned above, it is important to identify the mechanism behind these effects because, within the educational environment, there is no immediate monetary incentive that encourages students to help their peers.

The mechanism we propose involves students teaching their peers about specific knowledge that they possess (Griffith et al., 2014). This is a kind of human capital externality in which students increase the productivity of other students through processes of informal learning, without receiving compensation of any kind. Physical proximity to intelligent students may lead to better sharing of ideas and to learning. According to this theory, an increase in the academic achievement of their peers improves the performance of individual students in specific subjects. We might also expect that the effects would be greatest on those who interact most frequently.

Similarly, students might provide their peers with general knowledge of how to succeed in an educational setting (Griffith et al., 2014): for example, transmitting work habits and scheduling tips, or imparting study skills that help improve performance. As in the previous case, an increase in the academic achievement of their peers is also likely to result in students achieving better academic results in any given subject.

To ascertain which of these two mechanisms might be driving our results, we ran the following regression:

$$Z_i = \delta_0 + \delta_1 \overline{Z}_{-i} + \delta_2 X_i + \delta_3 \overline{X}_{-i} + \epsilon_i \tag{4}$$

Where Z_i represents student *i*'s final mark for mathematics, and X_i is a vector of their background characteristics. Each student *i* declares having a specific peer group P_i of size with whom they study economics. Next, $\bar{Z}_{-i}$ is measured as *i*-economics-peer's average outcome $\left(i.e., \frac{\sum_{j \in P_i} Z_j}{n_i} \right)$, and $\bar{X}_{-i}$ is a vector of the average background characteristics of *i*'s peers $\left(i.e., \frac{\sum_{j \in P_i} X_j}{n_i} \right)$.

Parameter estimates from Equations (1) and (4) might potentially help determine which of the two mechanisms is most important. If students teach their peers about specific knowledge that they possess, we might expect positive peer effects for the economics but not for the mathematics class. In other words, the parameter estimates for from Eq. (1) would be positive and significant, and δ_1 from Eq. (4) will be non-statistically different from zero. However, if students transfer general knowledge to their peers, we might expect positive peer effects for both the economics and mathematics classes. This finding would imply positive and significant parameter estimates for both $\alpha_1$and δ_1 .

Our data also allows us to explore whether the peer effect might lead to better or worse outcomes depending on the composition of the peer group. As Hoxby & Weingarth (2005) argue, the weakest and strongest students exert a disproportionate influence on their peers. For example, the shining light model posits that a single student with outstanding outcomes can inspire all others to increase their achievement.[8] By contrast, the bad apple model suggests that the presence of a single student with poor outcomes spoils the outcomes of many others.[9]

8 In the shining light model, the peer effects are those provided by a few outstanding students, who are capable of inspiring all others to raise their achievement. Using data from a Chinese college Han et al. (2009) found evidence that seems to favor the shining light model as it suggested that peer effects are asymmetric: Outstanding students help others, and non-outstanding students do not hurt outstanding students.

9 In contrast to the positive externalities models, Lazear (2001) claims that in the bad apple model the peer effects are those provided by the least academically able students in the classroom. These students provide negative externalities in at least three ways. First, the bad apple peer may distract students from productive tasks. Second, s/he may encourage disruptive behavior among other students. Finally, s/he may simply have low ability and require extra attention. Negative peer effects can also arise from an invidious comparison effect suggested by Hoxby & Weingarth (2005). Students with outstanding academic outcomes are held to depress the performance of everyone who is pushed to a lower rank in the local distribution. This could be a result of depressing their self-esteem. Hoxby & Weingarth (2005)

To explore the existence of heterogeneous effects caused by the composition of the peer group, we ran the following three regressions:

$$Y_i = \beta_0 + \beta_1 \bar{Y}_{S-i} + \beta_2 X_i + \beta_3 \bar{X}_{-i} + \in_i \text{ for } S = low, medium, high \tag{5}$$

where all the variables except the average outcome of peers -which depends on the composition of the peer group- were the same as those included in Eq. (1). Initially, we were interested in exploring whether both the weakest and the strongest students exert a significant influence on individual academic outcomes. To do this, we used a three-step procedure to measure peer effects such as peer academic ability. First, we created a categorical variable of low-, medium-, and high-ability students, using their national university entrance exam scores, a standardised test that is a requisite for admission to higher education institutions in Colombia. Second, we divided *i*'s reference group into those belonging to the low (P_i, low), medium $(P_i, medium)$, and high $(P_i, high)$ tercile, such that $n_i = n_i, low + n_i, medium + n_i, high$.

Finally, we created a set of average peer outcome variables that depended on the tercile of measured ability to which each peer belonged:

$$\bar{Y}_{S-i} = \frac{\sum_{j \in P_i} Y_j}{n_{i,S}} \text{ for } S = low, medium, high \tag{6}$$

A negative effect for the parameter accompanying $\bar{Y}_{low-i}$ would support the bad apple hypothesis while a positive effect of $\bar{Y}_{high-i}$, larger than $\bar{Y}_{medium-i}$, would support the shining light model.

Finally, we estimated three additional regressions to modify the set of friends. We asked students to select from among their study peers: (1) those individuals with whom they would like to study again, (2) those they normally ask for help whenever they do not understand a concept they have come across in their economics classes, and (3) those who have leadership attributes. We created three different sets of study networks for each of these classifications: empathy $(P_{i,emp})$, request $(P_{i,req})$, and leader . With this information, we created three different sets of average outcome variables using the following rule:

also describe other types of peer models that vary depending on group composition. Some examples are: Invidious comparison, boutique, focus, and rainbow models.

$$\bar{Y}_{S-i} = \frac{\sum_{j \in P_i, S} Y_j}{n_{i,S}} \text{ for } S = emp, req, leader \tag{7}$$

We then included each of these variables in a regression that was similar to the one depicted in Eq. (5). These regressions might provide useful empirical insights into the mechanisms lying behind our results.

V. Peer Effects

To estimate the effect of peers on student academic performance, we ran the linear-in-means model shown in Eq. (1), where represents the student's final mark, represents the average outcome of their peers, and is a set of controls. Table 5 reports the results of estimating the ordinary least squares (OLS) for two definitions of peer groups: one based on the set of peers that is not weighted by the frequency of interactions, and another that is if the students report interacting more frequently, they receive a higher weight in the average than those who report less frequent interactions. For each of these definitions, we estimated the model using two different sets of regressors: individual and household attributes.

Table 5. Peer Effects OLS Results

	Unweighted		Weighted Frecuency	
	(1)	(2)	(1)	(2)
Achievement of Peers	0.25	0 .29*	0.31*	0.33*
	(0.14)	(0.14)	(0.13)	(0.13)
Constant	2.3**	2.4**	2*	2.1*
	(0.75)	(0.80)	(0.75)	(0.79)
Controls				
Individual Attributes				
Household Attributes				
Mean Dep. Var.	3.1	3.2	3.1	3.2
SD Dep. Var.	0.75	0.75	0.73	0.73
N	83	81	81	79
[.75ex] *** p<0.01, ** p<0.05, * p<0.1				

Robust standard errors in brackets. For the full set of controls, see Table A.3. Column (1) includes individual characteristics such as age, gender, fellowship Ser Pilo Paga and place of birth, while Column (2) includes household attributes including parental education, locality fixed effects, and stratum.

Source: Authors' calculations.

While the unweighted OLS estimates are only significant for the second group of regressors, the weighted OLS estimates are larger and significant for both. Models that assume stronger effects between peers who interact more frequently deliver more robust evidence of peer effects. These results are free of reflection bias, since peer groups are individual-specific. However, they might be biased by self-selection and correlated effects.

Even when students are randomly assigned to induction week groups and economics classes, within-group interactions may adjust endogenously to group composition. In other words, students might be prone to associate and study with similar others. We estimated a multiple regression using a quadratic assignment procedure in order to explore the extent to which individual attributes influence the formation of study networks. Table 6 presents the results. Readers should avoid interpreting individual point estimates in these figures, as they are coefficients and not marginal effects. Students who are assigned to the same induction week group, the same economics and mathematics classes, and who belong to the *Ser Pilo Paga* scholarship program, are more likely to study together. We also found a statistically significant negative relationship in the probability that two individuals will study together if both were born in Bogotá.

Therefore, to minimise the plausible self-selection bias we included additional regressors: a set of dummies for the random assignment and birthplace of peers. Table 7 presents the results. When all variables are controlled for, the point estimates are smaller than those of Table 5 and no longer significant.[10]

10 Table A.4 shows the results of estimating the same model using the entry exam score as dependent variable. This exercise is a robustness check and shows that peers do not have any effect on entry exam score.

Table 6. Homophily

	(1)	(2)	(3)	(4)
Entry Exam (Standardised)	--0.059	--0.059	--0.065	--0.066
	(0.044)	(0.044)	(0.046)	(0.046)
Age	--0.035	--0.035	--0.038	--0.038
	(0.045)	(0.045)	(0.046)	(0.046)
Government Scholarship	0.32*	0.31*	0.31*	0.28
	(0.15)	(0.15)	(0.16)	(0.16)
Female	0.16	0.16	0.16	0.16
	(0.13)	(0.13)	(0.13)	(0.13)
Born in Bogota	--0.26*	--0.26*	--0.29*	--0.29*
	(0.13)	(0.13)	(0.13)	(0.13)
Self-Confidence	0.18	0.17	0.16	0.15
	(0.17)	(0.17)	(0.17)	(0.17)
Parental education		0.052	0.015	--0.0024
		(0.14)	(0.15)	(0.15)
Mathematics Class			0.66***	0.66***
			(0.14)	(0.14)
Economics Class			1.5***	1.6***
			(0.13)	(0.13)
Induction Week Group			1***	1***
			(0.13)	(0.13)
Locality				0.25
				(0.19)
Stratum				0.078
				(0.15)
Constant	--3.8***	--3.8***	--5.1***	--5.1***
	(0.17)	(0.17)	(0.21)	(0.21)
Pseudo R2	0.0066	0.0066	0.12	0.12
Chi2	15	16	273	275
N	8742	8742	8742	8742
[.75ex] *** p<0.01, ** p<0.05, * p<0.1				

This table presents the coefficients obtained from the logit regression (no marginal effects) and standard errors using the MR QAP procedure. Recall that the dependent variable is , indicating the network relationship between nodes *i* and *j*, which is equal to one if student *i* names student *j* as belonging to their study group. Column 1 includes individual controls, Column (2) includes a set of parental attributes, Column (3) controls for occupying the same classroom, and Column (4) adds additional household attributes.

Source: Authors' calculations.

Table 7. Peer Effects OLS Results

	Unweighted		Weighted Frecuency	
	3	4	3	4
Achievement of Peers	0.14	0.13	0.14	0.14
	(0.16)	(0.17)	(0.15)	(0.16)
Constant	3.0**	2.9**	2.8**	2.8**
	(0.85)	(0.88)	(0.86)	(0.87)
Controls				
Individual Attributes				
Household Attributes				
Maths, Econ, and Induction Group FE				
Peers Attributes				
Mean Dep. Var.	3.2	3.2	3.2	3.2
SD Dep. Var.	0.75	0.75	0.73	0.73
N	81	81	79	79
[.75ex] *** p<0.01, ** p<0.05, * p<0.1				

Notes: Standard errors in brackets. For the entire set of controls refer to Table 9 in Appendix

Robust standard errors in brackets. Column (1) includes individual characteristics such as age, gender, Ser Pilo Paga scholarship and place of birth, while Column (2) includes household attributes such as parental education, locality fixed effects, and stratum.

Source: Authors' calculations.

As mentioned above, the linear-in-means model might still be biased because of the presence of correlated effects. We used an IV strategy to minimise this potential bias. Table 8 reports the marginal effects computed for the average student and the average peer using a two-stage least squares methodology (2SLS), which uses the exogenous characteristics of excluded peers as instruments. In particular, we used excluded the average scores of peers in the entry exam, and the fraction of peers who were male, and who were born in Bogotá.

We report the results for two different measures of peer academic achievement: unweighted and weighted by the frequency of studying together. We also present the results for the four different sets of regressors. We do not reject the Sargan test of over-identification, but do reject the Hausman Test of exogeneity in all specifications. Although we reject the hypothesis that all instruments are jointly equal to zero in the first stage, the F-statistic is small, which may suggest that our instruments are weak. If this is the case, the test of significance is the incorrect size, and confidence intervals are wrong. We use the Conditional Likelihood Ratio methodology proposed by Moreira (2003) to correct both the individual p-value and the confidence intervals of peers' economic achievement. The results suggest that the peer effects are positive and statistically significant even after controlling for the weak instruments problem. Our preferred specification is the one in Column 4, in which we used the weighted-by-frequency definition of peers and conditioned on the entire set of regressors. In this specification, the estimates indicate that increasing the peer's average by one unit increases the mark achieved by 0.48, representing approximately 0.65 standard deviations of the dependent variable.

Table 8. Peer Effects 2SLS Results

	Unweighted				**Weighted Frecuency**			
	1	**2**	**3**	**4**	**1**	**2**	**3**	**4**
Achievement of Peers	0.66** (0.25)	0.67* (0.26)	0.46 (0.26)	0.44 (0.25)	0.79** (0.26)	0.77** (0.26)	0.49* (0.23)	0.48* (0.23)
CLR Confidence Interval	[.16,1.7]	[.18,2]	[-.11,1.6]	[.3,2]	[.3,2]	[.3,2.3]	[.03,1.7]	[.2,1.7]
CLR p value	0.013	0.015	0.095	0.11	0.0034	0.0045	0.037	0.042
Controls								
Individual Attributes								
Household Attributes								
Math, Econ, and Ind. Group FE								
Others								
Mean Dep Var.	3.1	3.1	3.1	3.1	3.1	3.2	3.2	3.2
SD Dep Var.	0.76	0.75	0.75	0.75	0.74	0.73	0.73	0.73
N	81	79	79	79	79	77	77	77
Sargan	4.7	7.4	8	8	5.3	7.7	9.6	9.7
Sargan P	0.58	0.28	0.24	0.24	0.51	0.26	0.14	0.14
F-statistic*	4	3.6	4	4	3.5	3.3	4	3.9
F-pval	0.0011	0.0025	0.0013	0.0014	0.003	0.0043	0.0013	0.0015
[.75ex] *** p<0.01, ** p<0.05, * p<0.1								

Robust standard errors in brackets. For the entire set of controls, refer to Table A.6. Column (1) includes individual characteristics such as age, gender, Ser Pilo Paga scholarship and place of birth, while Column (2) includes household attributes including parental education, locality fixed effects, and stratum. F-Statistics refer to the first stage regressions.

Source: Authors' calculations.

Our 2SLS estimates are larger than the OLS estimates. However, we would have expected the OLS results to over-estimate the true endogenous peer effect, because they cumulate the impact of endogenous and correlated effects. Nonetheless, De Giorgi et al. (2010), whose IV estimates were also larger than their OLS equivalents, argued that this interpretation rests on the implicit assumption that the two effects influence the dependent variable in the same direction for all students. In this setting, as in theirs, the excluded peers may be exposed to different common shocks (e.g., teacher or friend effects), and might influence the endogenous variable in different ways. This result is, therefore, not unexpected.

Table 9 shows that economics class peers do not seem to influence mathematics class performance. Recall that peers were identified by each student as individuals with whom they had studied economics during the previous month. This implies that although study networks for mathematics and economics might be correlated, there is evidence of peer effects only for the economics class. This finding suggests that peer effects work through specific rather than general knowledge. In other words, for specific subjects, physical proximity to intelligent students leads to an improved sharing of ideas and learning. We might argue that this is true because the results are significant only for those who interact more frequently and on a specific subject. Under the general knowledge mechanism, students learn work habits and scheduling techniques, or actual study skills, which enable them to improve their performance, regardless of the subject.

We next explore how low-, medium-, and high-skilled students affect individual academic outcomes. Table 10 shows the estimation results when peer effects are measured by the academic ability of peers, which is measured as a categorical variable of low-, medium-, and high-ability students. We find that the academic performance of low-skilled friends has a negative and significant impact on individual educational outcomes, while the academic performances of medium- and high-skilled friends have a positive and significant effect on a student's own academic achievement. In terms of magnitude, the effects of both medium- and high-skilled individuals are similar and close to those obtained for the entire sample.

Table 9. Peer Effects Two Stage OLS Results for Mathematics Scores

	Unweighted				Weighted Frecuency			
	1	2	3	4	1	2	3	4
Achievement of Peers	0.46	0.42	0.15	0.01	0.54	0.48	0.24	0.19
	(0.39)	(0.40)	(0.38)	(0.37)	(0.36)	(0.37)	(0.35)	(0.35)
Controls								
Individual Attributes								
Household Attributes								
Maths, Econ, and Ind Group FE								
Others								
Mean Dep. Var.	3.4	3.4	3.4	3.4	3.4	3.4	3.4	3.4
SD Dep. Var.	0.86	0.87	0.87	0.87	0.83	0.84	0.84	0.84
N	81	79	79	79	79	77	77	77
Sargan	4.5	4.3	3.6	3.5	4.4	4	4.2	4.3
Sargan P	0.61	0.63	0.73	0.74	0.62	0.68	0.65	0.63
F-Statistic	1.7	1.5	1.3	1.3	1.8	1.6	1.3	1.4
F-pval	0.13	0.17	0.26	0.25	0.11	0.15	0.26	0.24
[.75ex] *** p<0.01, ** p<0.05, * p<0.1								

Robust standard errors in brackets. For the entire set of controls, refer to Table A.3. Column (1) includes individual characteristics such as age, gender, fellowship and place of birth, while Column (2) includes household attributes including parental education, locality fixed effects, and stratum. F-Statistics refer to the first stage regressions.

Source: Authors' calculations.

The point estimates suggest that increasing the peer achievement of medium-skilled students by one unit raises student *i*'s marks by 0.39 units, while increasing the achievement of high-skilled students by the same amount increases their marks by 0.21 units. Surprisingly, a one-unit increase in the average grade of low-skilled students decreases one's own grade by 0.29 units. Although this result is suspicious, it echoes the results of previous analyses. For example, Foster (2006) explored the presence of peer effects among socially proximate peers at the University of Maryland, where some students are randomly assigned housing. She found a statistically significant negative peer effect on median students that stems from their peers' median SAT score. Carrell et al. (2011) took cohorts of first-year students at the United States Air Force Academy and assigned half to peer groups that had been designed to maximise the academic performance of the lowest ability students. They found a negative and significant treatment effect for the students they aimed to help and suggested that homophily effects explained their results. High and low ability students in the treatment groups appear to have segregated themselves into separate social networks, resulting in decreased beneficial social interaction between group members. Brady et al. (2015) claim that there can be a range of peer ability that increases the average and may hurt individual performance for both high-ability and low-ability students. They show that bad apples can pull down performance, even as the apples become less bad; they also show that these effects are sensitive to the degree to which homophily preferences matter.

Table 10. Heterogeneous Effects

	Low Skilled	Medium Skilled	High Skilled	Empathy	Request	Leader
	(1)	(2)	(3)	(4)	(5)	(6)
Achievement of Peers	--0.29**	0.39*	0.21*	0.36	0.43	0.46**
	(0.1)	(0.18)	(0.1)	(0.23)	(0.34)	(0.15)
Constant	3.8***	3.4***	2.7***	1.7	.91	2*
	(0.62)	(0.69)	(0.6)	(1)	(1.5)	(0.81)
Mean Dep. Var.	3.2	3.2	3.2	3.2	3	3.2
SD Dep. Var.	0.73	0.73	0.73	0.72	0.72	0.65
CLR Confidence Interval	[-.7,-.1]	[.1,2.8]	[-.04,1]	[.03,2.6]		[.19,1.2]
Cond LR p value	0.011	0.015	0.077	0.042	0.3	0.002
N	77	77	77	73	49	53
Sargan	6.3	5.6	11	15	9.8	9.2
Sargan P	0.39	0.47	0.1	0.017	0.14	0.17
Fstat	5	1.9	3	3.5	1.3	5.7
Fpval	0.00022	0.09	0.01	0.0038	0.28	0.00033
[.75ex] *** p<0.01, ** p<0.05, * p<0.1						

Robust standard errors in brackets. All columns control for individual characteristics such as age, gender, fellowship, place of birth, and household attributes like parental education, locality fixed effects, and stratum. F-Statistics refer to the first stage regressions. CLR confidence interval and Cond LR p value refer to the confidence interval and p-value estimated using the procedure developed by Moreira (2003) to correct for weak instruments.

Source: Authors' calculations.

We estimated three additional regressions that modify the set of friends and that might provide useful empirical insights into the mechanism behind our results. Columns (4) to (6) of Table 10 present the results of these regressions using three different sets of study networks for each of the previously identified classifications: empathy, request, and leader.

Peer effects are positive and statistically significant for the leader network. We did not find any significant result for the empathy network. We were, furthermore, unable to interpret the request network result, because the response rate was low, and the F test for the first stage was not statistically significant. These results suggest that peer effects operate through imitation. We argue that students imitate their peers because they view them as an important point of reference and use them as a benchmark for their own academic behaviours. In both these cases, peer effects are more likely to arise when the study friend is seen as a leader.

VI. Conclusions

There is no consensus in the literature about whether peers affect a student's academic performance. Many studies have found evidence of significant, albeit small, peer effects, while others have found no evidence of any such effect. This paper contributes to this discussion by estimating the effect of peer ability using an identification strategy that permitted us to determine causal peer effects by using the structure of the study network. We used data for one cohort of first-year students at a selective Colombian university.

We find that strong peer effects only occurred in higher education for frequent study peers only for the economics class. This finding suggests that peers work through human capital externalities. In particular, we argue that this type of externality only involves specific knowledge and does not occur in the case of general knowledge. The results also show heterogeneous effects of initial academic ability (measured by university entrance exam scores). We find that the academic performance of low-skilled friends has a negative and significant impact on individual educational outcomes, while the academic performance of medium- and high-skilled friends has a positive and significant effect on academic achievement. In other words, the ability level of friends has a positive and significant impact on a student's own academic achievement. When the mechanism behind these results is examined, it appears that the academic achievement of emphatic ties has no significant effect on a student's academic achievement, while the success of a leader has a positive and significant effect on the educational outcomes of students in the same group. This result suggests that, in our setting, peer effects work through imitation.

Our results imply that specific knowledge is the mechanism behind peer effects. They also suggest that the bad apple and the shining light models can coexist in the same environment. The finding that who you know is more important than what you know has important implications for the way in which universities could improve academic performance by paying attention to student study networks.

Acknowledgment

This paper benefited greatly from discussions with the members of the *Semillero de Econometría Aplicada de la Pontificia Universidad Javeriana* [Javeriana University Seedbed for Applied Econometrics].

References

1. Babcock, P., Bedard, K., Charness, G., Hartman, J., & Royer, H. (2015). Letting down the team? Social effects of team incentives. *Journal of the European Economic Association, 13*(5), 841-870.

2. Brady, R., Insler, M., & Rahman, A. (2015). *Bad company: Reconciling negative peer effects in college achievement* (MPRA Paper 68354). Munich, Germany.

3. Bramoullé, Y., Djebbari, H., & Fortin, B. (2009). Identification of peer effects through social networks. *Journal of Econometrics, 150*(1), 41-55.

4. Carrell, S. E., Sacerdote, B. I., & West, J.E. (2011). *From natural variation to optimal policy? The Lucas critique meets peer effects.* Washington D. C.: National Bureau of Economic Research.

5. De Giorgi, G., Pellizzari, M., & Redaelli, S. (2010). Identification of social interactions through partially overlapping peer groups. *American Economic Journal: Applied Economics, 2*(2), 241-275.

6. Duflo, E., & Saez, E. (2003). The role of information and social interactions in retirement plan decisions: Evidence from a randomized experiment. *The Quarterly Journal of Economics, 18*(3), 815-842.

7. Foster, G. (2006). It's not your peers, and it's not your friends: Some progress toward understanding the educational peer effect mechanism. *Journal of Public Economics, 90*(8), 1455-1475.

8. Freeman, L. C. (1979). Centrality in social networks conceptual clarification. *Social Networks, 1*, 215-239.

9. Griffith, A. L., & Rask, K. N. Peer effects in higher education: A look at heterogeneous impacts. *Economics of Education Review, 39*, 65-77.

10. Han, L. & Li, T. (2009). The gender difference of peer influence in higher education. *Economics of Education Review, 28*(1), 129-134.

11. Hoxby, C. M., & Weingarth, G. (2005). *Taking race out of the equation: School reassignment and the structure of peer effects* (Working paper).

12. Lazear, E. P. Educational production. (2001). *The Quarterly Journal of Introduction to Economics, 16*(3), 777-803.

13. Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The Review of Economic Studies, 60*(3), 531-542.

14. McEwan, P. J., & Soderberg, K. A. (2006). Roommate effects on grades: Evidence from first-year housing assignments. *Research in Higher Education, 47*(3), 347-370.

15. Moreira, M. J. (2003) A conditional likelihood ratio test for structural models. *Econometrica, 71*(4), 1027-1048.

16. Sacerdote, B. (2001). Peer effects with random assignment: Results for Dartmouth roommates. *The Quarterly Journal of Introduction to Economics, 16*(2), 681-704.

17. Sacerdote, B. (2011). Peer effects in education: How might they work, how big are they and how much do we know thus far? *Handbook of the Economics of Education* (chap. 3, pp. 249-277). Amsterdam, The Netherlands: Elsevier.

18. Zimmerman, D. J. (2003). Peer effects in academic outcomes: Evidence from a natural experiment. *Review of Economics and Statistics, 85*(1), 9-23.

Appendices

Appendix 1. Solution to the reflection problem

We use an example provided by De Giorgi et al. (2010) to illustrate this point. Consider the simple case of three students Ana, Ben, and Cesar. Students Ana and Ben study together. However, Ben also studies with Cesar. Ana's peer group includes only Ben $\left(i.e., \ G_A = \{B\}\right)$, while Ben's peer group includes both Ana and Cesar $\left(i.e., \ G_B = \{A,C\}\right)$, and Cesar's peer group includes only Ben $\left(i.e., \ G_C = \{B\}\right)$. This identification can be seen as a case of triangularization.

In the standard simultaneous equation model, at least one exogenous variable is excluded from each equation. This implies that Eq.1 is equal the following system of three equations:

$$Y_A = \alpha + \beta Y_B + \gamma X_A + \delta X_B + \in_A$$

$$Y_B = \alpha + \beta\left(\frac{Y_A + Y_C}{2}\right) + \gamma X_B + \delta\left(\frac{X_A + X_C}{2}\right) + \in_B$$

$$Y_C = \alpha + \beta Y_B + \gamma X_C + \delta X_B + \in_C$$

The reduced form equations are:

$$Y_A = \pi_{A0} + \pi_{A1} X_B + \pi_{A2}\left(\frac{X_A + X_C}{2}\right) + \pi_{A3} + \pi_{A3} + \mu_A$$

$$Y_B = \pi_{B0} + \pi_{B1} X_B + \pi_{B2}\left(\frac{X_A + X_C}{2}\right) + \mu_B$$

$$Y_C = \pi_{C0} + \pi_{C1} X_B + \pi_{C2}\left(\frac{X_A + X_C}{2}\right) + \pi_{C3} + \mu_C$$

where the error terms are linear combinations of the structural error terms. Moreover, $\pi_{A0} = \pi_{C0} = \alpha + \frac{\alpha\beta(1+\beta)}{1-\beta^2}$, $\pi_{A1} = \pi_{C1} = \frac{\beta(\beta\delta+\gamma)}{1-\beta^2} + \delta$, $\pi_{A2} = \pi_{C2} = \frac{\beta(\beta\gamma+\delta)}{1-\beta^2}$, $\pi_{A3} = \pi_{C3} = \gamma, \pi_{B0} = \left(\frac{\alpha(1+\beta)}{1-\beta^2}\right)$, $\pi_{B1} = \frac{(\beta\delta+\gamma)}{\left(1-\beta^2\right)}$,

$\pi_{B2} = \frac{(\beta\gamma + \delta)}{(1-\beta^2)}$, . There are four reduced form parameters and four structural ones. Therefore, it is possible to identify all structural parameters. This identification strategy rests on the assumption that the excluded peer Cesar does not interact with Ana directly. In our setting, this seems a plausible assumption.

Table A.1. Balance test induction week groups

	(Induction Week Groups)			
	(1)	**(2)**	**(3)**	**(4)**
Entry Exam (Standardised)	0.012	0.025	--0.029	--0.0088
	(0.049)	(0.052)	(0.057)	(0.047)
Age	0.021	0.015	0.013	--0.049***
	(0.017)	(0.014)	(0.015)	(0.0094)
Female	--0.21*	0.096	--0.021	0.14
	(0.1)	(0.11)	(0.11)	(0.1)
Born in Bogota	--0.22	0.1	0.14	--0.029
	(0.12)	(0.11)	(0.11)	(0.11)
Confidence	0.022	--0.12	0.017	0.079
	(0.14)	(0.19)	(0.17)	(0.12)
Secondary	--0.21	0.14	--0.11	0.18
	(0.23)	(0.24)	(0.16)	(0.15)
Technical	--0.074	--0.32	0.31	0.08
	(0.23)	(0.23)	(0.25)	(0.16)
Tertiary	--0.19	--0.2	0.19	0.2
	(0.19)	(0.23)	(0.2)	(0.16)
Postgraduate	--0.45*	--0.0095	0.21	0.25
	(0.18)	(0.23)	(0.21)	(0.16)
Locality	0.012	--0.014	--0.01	0.012
	(0.013)	(0.013)	(0.012)	(0.011)
Stratum	0.1	--0.068	--0.023	--0.013
	(0.054)	(0.048)	(0.051)	(0.058)
Constante	--0.11	0.49	--0.057	0.68**
	(0.33)	(0.29)	(0.28)	(0.22)
Mean Dep Var.	0.29	0.29	0.23	0.19
SD Dep Var.	0.46	0.46	0.42	0.39
N	79	79	79	79
[.75ex]				

This table presents the coefficients obtained for the linear probability model. The dependent variable is Y_{ij} indicating the group that each student belongs to. In column (1), for example, the dependent variable is equal to one if student *i* belongs to induction week group 1 and it is equal to zero if he or she belongs to any of the other three groups. Robust Standard Errors in parenthesis.

Source: Authors' calculations.

Table A.2. Balance test economics class groups

	(Economics Class Groups)			
	(1)	(2)	(3)	(4)
Entry Exam (Standardised)	--0.042	--0.0026	0.046	--0.0013
	(0.052)	(0.05)	(0.052)	(0.056)
Age	0.013	0.0013	0.014	--0.028
	(0.016)	(0.019)	(0.016)	(0.017)
Female	--0.073	0.14	--0.059	--0.0077
	(0.11)	(0.099)	(0.1)	(0.11)
Born in Bogota	0.037	0.16	--0.18	--0.017
	(0.11)	(0.1)	(0.11)	(0.13)
Confidence	0.063	0.027	0.058	--0.15
	(0.17)	(0.16)	(0.11)	(0.17)
Secondary	0.085	--0.64***	0.069	0.49*
	(0.28)	(0.16)	(0.17)	(0.24)
Technical	0.21	--0.77***	0.27	0.28
	(0.24)	(0.17)	(0.17)	(0.2)
Tertiary	--0.064	--0.56**	0.19	0.43*
	(0.19)	(0.19)	(0.13)	(0.17)
Postgraduate	--0.16	--0.39	0.27	0.28
	(0.2)	(0.2)	(0.15)	(0.18)
Locality	--0.0058	0.0038	0.0058	--0.0038
	(0.013)	(0.0096)	(0.013)	(0.013)
Stratum	0.023	0.015	--0.062	0.023
	(0.056)	(0.049)	(0.056)	(0.057)
Constant	0.035	0.39	0.038	0.53
	(0.34)	(0.38)	(0.29)	(0.33)
Mean Dep Var.	0.28	0.29	0.23	0.19
SD Dep Var.	0.45	0.46	0.42	0.39
N	79	79	79	79
[.75ex]				

This table presents the coefficients obtained for the linear probability model. The dependent variable is *Y_{ij}* indicating the group that each student belongs to. In column (1), for example, the dependent variable is equal to one if student *i* belongs to the Eeconomics class group 1 and it is equal to zero if he belongs to any of the other three groups. Robust Standard Errors in parenthesis.

Source: Authors' calculations.

Table A.3. Peer Effects OLS Results

	Unweighted				Weighted Frecuency			
	1	2	3	4	1	2	3	4
Achievement of Peers	0.25 (0.14)	0.29* (0.14)	0.14 (0.16)	0.13 (0.17)	0.31* (0.13)	0.33* (0.13)	0.14 (0.15)	0.14 (0.16)
Scholarship	--0.088 (0.21)	--0.19 (0.24)	--0.052 (0.24)	--0.022 (0.25)	--0.032 (0.2)	--0.11 (0.23)	0.093 (0.23)	0.11 (0.23)
Entry Exam (Standardised)	0.45*** (0.051)	0.44*** (0.052)	0.45*** (0.052)	0.46*** (0.056)	0.44*** (0.051)	0.43*** (0.052)	0.44*** (0.05)	0.45*** (0.054)
Age	--0.0039 (0.03)	--0.0055 (0.03)	0.0015 (0.03)	0.00084 (0.03)	0.0025 (0.032)	0.0013 (0.032)	0.01 (0.033)	0.0093 (0.033)
Female	0.034 (0.13)	0.047 (0.13)	0.084 (0.14)	0.082 (0.14)	0.07 (0.13)	0.083 (0.13)	0.14 (0.13)	0.14 (0.13)
Born in Bogota	--0.033 (0.17)	--0.027 (0.17)	0.013 (0.18)	0.0097 (0.18)	0.029 (0.16)	0.035 (0.16)	0.11 (0.17)	0.1 (0.17)
Confidence	0.16 (0.21)	0.21 (0.21)	0.16 (0.24)	0.14 (0.24)	0.061 (0.19)	0.1 (0.19)	0.031 (0.22)	0.016 (0.21)
Maximum parental education		--0.06 (0.061)	--0.051 (0.061)	--0.069 (0.059)		--0.042 (0.06)	--0.021 (0.06)	--0.036 (0.063)
Math Class			--0.018 (0.17)	0.0093 (0.18)			--0.0038 (0.16)	0.015 (0.17)
Econ Professor 1			--0.26 (0.2)	--0.27 (0.21)			--0.39* (0.2)	--0.4* (0.2)
Econ Professor 2			--0.23 (0.23)	--0.25 (0.23)			--0.35 (0.21)	--0.37 (0.21)
Econ Professor 3			--0.24 (0.21)	--0.22 (0.21)			--0.31 (0.21)	--0.29 (0.21)
Induction Week Group 2			--0.39* (0.16)	--0.36 (0.19)			--0.42** (0.16)	--0.4* (0.18)
Induction Week Group 3			--0.16 (0.23)	--0.14 (0.24)			--0.15 (0.22)	--0.13 (0.23)
Induction Week Group 4			0.039 (0.24)	0.051 (0.24)			--0.0073 (0.23)	0.0033 (0.23)
Locality				0.00049 (0.012)				0.0016 (0.012)
Stratum				0.048 (0.077)				0.035 (0.079)
Constant	2.3** (0.75)	2.4** (0.8)	3.0** (0.85)	2.9** (0.88)	2* (0.75)	2.1* (0.79)	2.8** (0.86)	2.8** (0.87)
Mean Dep Var.	3.1	3.2	3.2	3.2	3.1	3.2	3.2	3.2
SD Dep Var.	0.75	0.75	0.75	0.75	0.73	0.73	0.73	0.73
N	83	81	81	81	81	79	79	79

[.75ex] *** p<0.01, ** p<0.05, * p<0.1

Table A.4. Robustness check: Dependent variable entry exam (standardized)

	Unweighted		Weighted Frecuency	
	3	4	3	4
Achievement of Peers	0.09	0.12	0.09	0.12
	(0.24)	(0.23)	(0.24)	(0.23)
Constant	-1.3	-0.92*	-1.4	-0.95
	(1.1)	(1.0)	(1.0)	(1.1)
Controls				
Individual Attributes				
Household Attributes				
Maths, Econ, and Induction Group FE				
Peers Attributes				
Mean Dep. Var.	0.038	0.038	0.044	0.044
SD Dep. Var.	1.0	1.0	1.0	1.0
N	81	81	79	79
[.75ex] *** p<0.01, ** p<0.05, * p<0.1				

Notes: Standard errors in brackets. For the entire set of controls refer to Table 9 in Appendix.

Source: Authors' calculations. Column (1) includes individual characteristics such as age, gender, fellowship and place of birth, while Column (2) includes household attributes such as parental education, locality fixed effects, and stratum.

Table A.5. Peer effects two stage OLS results

	Unweighted				Weighted Frecuency			
	1	2	3	4	1	2	3	4
Achievement of Peers	0.66** (0.25)	0.67* (0.26)	0.46 (0.26)	0.44 (0.25)	0.79** (0.26)	0.77** (0.26)	0.49* (0.23)	0.48* (0.23)
CLR Confidence Interval	[.16,1.7]	[.18,2]	[-.11,1.6]	[.3,2]	[.3,2]	[.3,2.3]	[.03,1.7]	[.2,1.7]
CLR p value	0.013	0.015	0.095	0.11	0.0034	0.0045	0.037	0.042
Scholarship	--0.0093	--0.15	--0.066	--0.051	0.052	--0.077	0.064	0.068
	(0.19)	(0.2)	(0.19)	(0.2)	(0.19)	(0.2)	(0.19)	(0.19)
Entry Exam (Standardised)	0.42*** (0.066)	0.42*** (0.067)	0.44*** (0.067)	0.44*** (0.069)	0.4*** (0.066)	0.4*** (0.066)	0.43*** (0.063)	0.43*** (0.065)
Age	0.018	0.012	0.013	0.014	0.028	0.022	0.023	0.024
	(0.032)	(0.031)	(0.029)	(0.031)	(0.032)	(0.031)	(0.028)	(0.029)
Female	--.0041	0.014	0.056	0.055	0.049	0.062	0.12	0.12
	(0.14)	(0.14)	(0.13)	(0.13)	(0.13)	(0.14)	(0.12)	(0.12)

Table A.5. Peer effects two stage OLS results (continued)

	Unweighted				Weighted Frecuency			
	1	2	3	4	1	2	3	4
Born in Bogota	0.017	0.018	0.034	0.034	0.096	0.091	0.13	0.13
	(0.15)	(0.15)	(0.14)	(0.14)	(0.15)	(0.15)	(0.13)	(0.14)
Confidence	0.14	0.21	0.16	0.15	0.023	0.087	0.035	0.032
	(0.21)	(0.21)	(0.2)	(0.2)	(0.21)	(0.21)	(0.19)	(0.19)
Maximum parental education		--0.085 (0.064)	--0.063 (0.059)	--0.071 (0.065)		--0.07 (0.063)	--0.038 (0.056)	--0.039 (0.062)
Matsh Class			--0.0021	0.011			0.018	0.021
			(0.16)	(0.16)			(0.15)	(0.15)
Econ Professor 1			--0.12 (0.21)	--0.13 (0.21)			--0.26 (0.2)	--0.26 (0.2)
Econ Professor 2			--0.043 (0.23)	--0.055 (0.24)			--0.16 (0.22)	--0.16 (0.22)
Econ Professor 3			--0.19 (0.19)	--0.18 (0.19)			--0.27 (0.18)	--0.27 (0.18)
Induction Week Group 2			--0.41* (0.18)	--0.4* (0.19)			--0.45** (0.17)	--0.45* (0.18)
Induction Week Group 3			--0.21 (0.2)	--0.2 (0.2)			--0.2 (0.19)	--0.2 (0.2)
Induction Week Group 4			0.0055 (0.22)	0.01 (0.22)			--0.045 (0.21)	--0.044 (0.21)
Locality				--0.0014				--0.001
				(0.015)				(0.014)
Stratum				0.024 (0.074)				0.0059 (0.069)
Mean Dep Var.	3.1	3.1	3.1	3.1	3.1	3.2	3.2	3.2
SD Dep Var.	0.76	0.75	0.75	0.75	0.74	0.73	0.73	0.73
N	81	79	79	79	79	77	77	77
Sargan	4.7	7.4	8	8	5.3	7.7	9.6	9.7
Sargan P	0.58	0.28	0.24	0.24	0.51	0.26	0.14	0.14
Fstat	4	3.6	4	4	3.5	3.3	4	3.9
Fpval	0.0011	0.0025	0.0013	0.0014	0.003	0.0043	0.0013	0.0015
[.75ex] *** p<0.01, ** p<0.05, * p<0.1								

Source: Authors' calculations.

Table A.6. Peer effects two stage OLS results for mathematics scores

	Unweighted				Weighted Frecuency			
	1	2	3	4	1	2	3	4
Achievement of Peers	0.46 (0.39)	0.42 (0.40)	0.15 (0.38)	0.01 (0.37)	0.54 (0.36)	0.48 (0.37)	0.24 (0.35)	0.19 (0.35)
Scholarship	0.016 (0.27)	0.047 (0.3)	--0.13 (0.26)	--0.096 (0.26)	0.1 (0.26)	0.17 (0.29)	0.06 (0.25)	0.078 (0.25)
Entry Exam (Standardised)	0.21* (0.094)	0.21* (0.098)	0.36*** (0.089)	0.38*** (0.091)	0.21* (0.091)	0.2* (0.094)	0.35*** (0.085)	0.35*** (0.087)
Age	0.011 (0.044)	0.012 (0.044)	--0.0091 (0.039)	--0.013 (0.04)	0.018 (0.042)	0.021 (0.043)	0.0039 (0.037)	--0.00048 (0.038)
Female	--0.049 (0.19)	--0.036 (0.2)	0.024 (0.17)	0.021 (0.17)	0.00071 (0.19)	0.014 (0.19)	0.091 (0.16)	0.09 (0.16)
Born in Bogota	0.21 (0.21)	0.22 (0.22)	0.24 (0.2)	0.24 (0.2)	0.31 (0.21)	0.33 (0.21)	0.36 (0.19)	0.35 (0.19)
Confidence	0.32 (0.32)	0.29 (0.34)	0.33 (0.29)	0.28 (0.3)	0.21 (0.32)	0.15 (0.33)	0.18 (0.28)	0.15 (0.29)
Maximum parental education		0.026 (0.085)	0.05 (0.078)	0.027 (0.084)		0.053 (0.084)	0.075 (0.077)	0.062 (0.082)
Maths Class			0.94*** (0.24)	1*** (0.26)			0.92*** (0.24)	0.95*** (0.26)
Econ Professor 1			0.18 (0.24)	0.16 (0.24)			0.04 (0.24)	0.027 (0.24)
Econ Professor 2			--0.053 (0.24)	--0.081 (0.24)			--0.22 (0.23)	--0.24 (0.23)
Econ Professor 3			0.25 (0.3)	0.26 (0.29)			0.19 (0.28)	0.19 (0.27)
Induction Week Group 2			--0.24 (0.3)	--0.22 (0.3)			--0.24 (0.29)	--0.23 (0.28)
Induction Week Group 3			--0.19 (0.33)	--0.18 (0.32)			--0.084 (0.34)	--0.08 (0.33)
Induction Week Group 4			--0.13 (0.29)	--0.13 (0.29)			--0.15 (0.28)	--0.15 (0.28)
Locality				0.0046 (0.02)				0.0058 (0.019)
Stratum				0.058 (0.11)				0.03 (0.1)
Mean Dep Var.	3.4	3.4	3.4	3.4	3.4	3.4	3.4	3.4
SD Dep Var.	0.86	0.87	0.87	0.87	0.83	0.84	0.84	0.84

Table A.6. Peer effects two stage OLS results for mathematics scores (continued)

	Unweighted				Weighted Frecuency			
	1	**2**	**3**	**4**	**1**	**2**	**3**	**4**
N	81	79	79	79	79	77	77	77
Sargan	4.5	4.3	3.6	3.5	4.4	4	4.2	4.3
Sargan P	0.61	0.63	0.73	0.74	0.62	0.68	0.65	0.63
Fstat	1.7	1.5	1.3	1.3	1.8	1.6	1.3	1.4
Fpval	0.13	0.17	0.26	0.25	0.11	0.15	0.26	0.24
[.75ex] *** p<0.01, ** p<0.05, * p<0.1								

Source: Authors' calculations.