THE UNIVERSITY OF QUEENSLAND
AUSTRALIA
CREATE CHANGE

Discussion Paper Series
UQ School of Economics

The Impact of Peers on Academic Performance: Theory and Evidence from a Natural Experiment
2021-04-29
Diego Carrasco-Novoa, Sandro Díez-Amigo, Shino Takayama

Discussion Paper No.644

The Impact of Peers on Academic Performance: Theory and Evidence from a Natural Experiment [*†]

Diego Carrasco-Novoa[‡] Sandro Díez-Amigo [§] Shino Takayama [¶]

April 29, 2021

Abstract

We introduce a flexible theoretical framework to model the mechanics of peer effects in education. Then we take advantage of a natural experiment in order to illustrate how the proposed model can be used to gain additional empirical insights from reduced-form econometric analysis. Leveraging the exogenous variation in peer characteristics generated by the random assignment of freshman college students to their first semester class groups, we observe a negative impact on academic performance of secondary schoolmate presence and concentration in the first semester college classroom, suggesting that in the study context socialization was in overall terms *distractive*, and that the group structure increased socialization for all students. We also find some evidence of a negative impact of higher average admission scores on academic performance, suggesting that in the study context the direct positive impact of peer mean ability on academic performance was more than eclipsed by the negative effect of higher peer mean ability on self-confidence. Observed peer effects generally persist throughout the duration of undergraduate studies.

[*]A partial version of this paper was previously circulated as Díez-Amigo, S. (2014). "The Impact of College Peers on Academic Performance: Evidence from a Natural Experiment in Chile". MPRA Working Paper No. 62913.

[†]We would like to thank Abhijit Banerjee, Benjamin Olken, Francisco Gallego and Michael Piore for their invaluable contributions to this research. We are also grateful to Benjamin Golub, Gastón Illanes, Jeanne Lafortune, Andrew McLennan, José Miguel Sánchez, José Tessada, Victoria Valdés, Terence Yeo, one anonymous referee and participants in seminars at the MIT Department of Economics for their numerous comments and many suggestions for improving this paper. Funding and institutional support for this research were generously provided by the Caja Madrid Foundation, the Rafael del Pino Foundation, the "la Caixa" Foundation, the MIT Department of Economics, the Abdul Latif Jameel Poverty Action Lab and its office for Latin America and the Caribbean, and the Pontificia Universidad Católica de Chile and its Faculty of Economic and Administrative Sciences. The views represented in this paper and any errors are our own.

[‡]School of Economics, University of Queensland

[§]Department of Economics, Massachusetts Institute of Technology [‖]

[¶](corresponding author) School of Economics, University of Queensland, St Lucia, QLD 4072, Australia; s.takayama1@uq.edu.au; t: +61-7-3346-7379; f: +61-7-3365-7299)

1 Introduction

Small positive impacts of peer ability on academic achievement have been observed using linear-in-means models or fixed effect approaches (Ding and Lehrer 2007; Vigdor and Nechyba 2007; Zabel 2008), and strong positive peer effects have been observed when randomizing primary selection (Duflo et al. 2011; Kang et al. 2007; Whitmore 2005; Graham 2008). However, using this same experimental methodology applied to school roommates, Zárate (2019) finds that higher ability peers have a negative impact on the academic performance of lower-achieving students through a reduction in the latter's self-confidence. Moreover, some studies find peer effects evenly distributed among students (Angrist and Lang 2004; Arcidiacono et al. 2012), whereas other studies observe asymmetric peer effects (Zabel 2008; Rothstein 2009).

One interpretation of this divergence is that peer effects can vary with the environment (e.g., primary vs secondary education, self selection of students vs random assignment). If so, the understanding of the mechanisms through which peer effects operate needs to be further developed. Moreover, according to Epple and Romano (2011), theoretical models need to be extended and refined, providing a more developed framework for the analysis of the impact of peer ability on academic achievement which is grounded on the empirical evidence.

Taking all of these elements into consideration, this paper proposes a flexible theoretical framework to model several factors which empirical evidence suggests that may affect the impact of peer ability on academic performance. In particular, we introduce three additional elements into a baseline model where, in line with the empirical findings in the literature, peer ability has a positive direct impact on academic achievement. Then, we take advantage of a natural experiment in order to illustrate how the proposed theoretical framework can be used to gain additional empirical insights from reduced-form econometric analysis.

First, we consider the pattern of student effort in different peer ability environments. Following Zárate (2019), we introduce into the model the concept of self-confidence, which is positively correlated with effort productivity but negatively associated with the relative own ability of students compared to their peers. Introducing preferences for leisure, we show that in this context an increase in peer ability can result in lower student effort, leading to decreased academic achievement.

Second, we consider the role of social interactions, while allowing for two possibilities: on the one hand, increased socialization may facilitate coordination with peers and increase the marginal return to student effort. On the other hand, socialization inside educational institutions has been associated with higher procrastination in some contexts (see for example Day et al. 2000; Gardner and Steinberg 2005; O'Brien et al. 2011; Gustavson et al. 2014).

Third, we expand the theoretical framework to allow peer effects to persist over time. While the evolution of peer effects has not yet been exhaustively explored in the literature, there is a consensus about its importance for the understanding of the impact of peers on long-term academic performance (see Jaffee, 2007).

Then we take advantage of a natural experiment in order to illustrate how the proposed theoretical framework can be used to gain additional empirical insights from reduced-form econometric analysis. This experimental strategy allows us to overcome the standard problems related to the empirical estimation of peer effects, as discussed in Section 2. In particular, we leverage the exogenous variation in peer characteristics generated by the random assignment of freshman students to their first semester class groups in the Engineering and Commercial Engineering degrees at Pontificia Universidad Católica de Chile, which allows us to robustly estimate the impact of peer characteristics on undergraduate academic performance.

We observe a negative impact on academic performance of secondary schoolmate presence in the first semester college classroom. Assuming that sharing the classroom with peers who attended the same secondary school increased the likelihood of socialization, this finding suggests that in the study context socialization was in overall terms *distractive*, in line with the positive correlation between socialization and procrastination observed by Day et al. (2000), Gardner and Steinberg (2005), O'Brien et al. (2011), or Gustavson et al. (2014) in several other contexts. Similarly, we find that a higher concentration of students from the same secondary schools in the first semester college class group had a negative impact on all students, irrespective of the secondary school they attended, suggesting that in the study context the structure of the group increased (*distractive*) socialization for all students.

We also find some evidence of a negative impact of higher average admission scores on academic performance. This suggests that, in the study context, the direct positive impact of peer mean ability on academic performance was more than eclipsed by the negative effect of higher peer mean ability on self-confidence. This is in line with the findings of Zárate (2019) in the context of selective high schools in Peru.

Finally, first semester college classroom characteristics are estimated to have had a persistent, significant impact on academic performance throughout a student's undergraduate years. This suggests that persistence is indeed a desirable feature of peer effects theoretical frameworks, and that in the study context the first college semester was key for student academic achievement.

While context-specific, these empirical results yield actionable evidence to guide further policy experimentation and research on peer effects in education, and the proposed theoretical framework can be adapted to model the impact of peers on academic performance in different educational contexts.

The rest of this paper is organized as follows: Section 2 further reviews the literature and presents the motivation for the research; Section 3 introduces the proposed theoretical framework; Section 4 presents the natural experiment and empirical results; Section 5 discusses how the proposed theoretical framework can be used to gain additional insights from the empirical results; Section 6 concludes.

2 Literature and Motivation

Theoretical research on peer effects in education has been criticized as lacking an adequate empirical basis (Epple and Romano, 2011). In this paper we consider four main aspects of peer effects in education which have both theoretical and empirical implications.

First, the effort pattern in different peer environments can vary with both observable and unobservable peer characteristics (Graham, 2008; Epple and Romano, 2011; Kimbrough et al., 2017), and this relationship can be complex. For example, in the case of tracking empirical evidence suggests that students in higher tracks exert substantially more effort than students in lower tracks (see Carbonaro, 2005). This positive effect of peer ability can be even more significant if, as Cooley (2009), we consider that higher-achieving peers can positively influence other students through imitation. However, Zárate (2019) randomly varies the characteristics of neighbours in dormitories and observes that higher-achieving peers reduce the productivity of lower-achieving students through a decrease in their self-confidence. Then, even in a context where higher peer ability exerts a positive direct impact on student performance, the indirect effect of self-confidence can result in overall negative peer ability spillovers. Similarly, peers of any ability may also have an indirect negative impact on student performance through social procrastination. Being socially active has been characterised as one of the most common patterns underlying procrastination (Day et al., 2000). Similarly, the presence of peers is associated with increased compulsivity (Gardner and Steinberg 2005; O'Brien et al. 2011), which is positively correlated with procrastination (Gustavson et al. (2014)). Conversely, while this has not been yet documented in the literature, it is conceivable to think that in some contexts increased socialization with peers may also have a positive impact on student performance through other channels, such as better synergies between students arising from additional information exchanges, or improved mental health.

Second, peer effects can persist over time. While this aspect has not yet been exhaustively explored in the literature, Arcidiacono et al. (2012) finds large, persistent effects even after relatively short duration interactions between students.

Third, the formation of peer relationships can be endogenous (*selection problem*), and a student can simultaneously affect and be affected by peers (*reflection problem*) (Carrell et al. (2013)). The established empirical approach to address these issues in the peer effects literature is to follow the methodology introduced by Sacerdote (2001) and Zimmerman (2003), who take advantage of (natural) experiments where students are randomly assigned to peer groups. This methodology has been extensively used to study peer effects on academic performance (Duflo et al. 2011; Garlick 2014, among many), as well as on non-academic dimensions and non-classroom interactions (for example, Boisjoly et al. 2006; Kremer and Levy 2008).

Fourth, the existence of a mechanical relationship between the measures of own ability and peer ability might lead to biased empirical estimates of peer effects. Classical measurement error, typically associated with attenuation bias, can instead lead to substantial overestimation

of peer effects, particularly in a setting where group assignment in systematic Angrist (2014). Although in a context of random group assignment classical measurement has been shown to only lead to typical attenuation bias (Feld and Zölitz (2017)), the Angrist critique is best addressed by introducing covariates in the empirical analysis.

Based on the above, in this paper we introduce a theoretical framework aimed at explaining the mechanics behind the varied direct and indirect impacts of peer characteristics on student academic achievement in the education literature. This theoretical framework can be expanded to include additional features which may be behind the variation in empirically observed peer effects, such as socialization and self-confidence, and it can also account for the time-persistence of peer effects on academic achievement. We then take advantage of a natural experiment to empirically showcase how this theoretical framework can help to better understand the mechanics of peer effects among freshman college students randomly assigned to their first semester classrooms at Pontificia Universidad Católica de Chile. The random group assignment allows us to circumvent the selection and reflection problems, as well as peer effects overestimation problems arising from classic measurement error, but in order to better address the Angrist critique we also include additional covariates to better capture peer group characteristics. This approach is similar to Shue (2013), who estimates how executive peer networks can affect managerial decision-making and firm policies through random assignment of students to MBA sections at Harvard Business School. However, this paper focuses on college education, instead than on MBA students.

3 Theoretical Framework

3.1 Baseline Effort-Achievement Model

Let's assume that student achievement a is a function of effort e and a vector of other variables denoted by x. Suppose that in each period, a student is allocated 1 unit of time and can choose e units of time to work and l units of time for leisure. Then the maximization problem is

$$\max \sum_{t=1}^{T} u(l_t) + a(\sum_{t=1}^{T} e_t, x)$$
$$\text{subject to } l_t + e_t = 1 \text{ for every } t$$

where $u(\cdot)$ is assumed to be strictly concave. We write the first derivative of a with respect to effort by a'. Thus, $a'(e, x)$ represents the marginal change in achievement when a student with characteristics x changes effort level infinitesimally.

Proposition 1. *In equilibrium, it is held that for every t,*

$$l_t^* = l^*$$

where l^* satisfies

$$u'(l^*) = a'(T \cdot (1 - l^*), x).$$

Further the equilibrium is unique.

Proof. Fix t. By taking the first derivative of the objective function with respect to l_t, when the student chooses the optimal l_t^*, it is held that

$$u'(l_t^*) = a'(\sum_{s \neq t} e_s + (1 - l_t^*), x). \tag{1}$$

Because t is arbitrary, and (1) holds for any t, there must be l^* such that $l_t^* = l^*$ for every t. Therefore we obtain $\sum_{t=1}^{T}(1 - l_t^*) = T \cdot (1 - l^*)$. Further, the solution l^* is unique because u is strictly concave and there is only one point in which $u'(l^*) = a'(T \cdot (1 - l^*), x)$. □

3.2 Introducing Ability

Now let $x = (b, \theta)$ and then an achievement $a(e, b, \theta)$ is a function of own ability b and peer mean ability θ in addition to effort, such that $a(e, b, \theta)$ is strictly increasing and strictly concave in all variables.

Assumption 1. *Let two levels of own ability b_1 and b_2 satisfy $b_1 < b_2$. An achievement function $a(e, b, x)$ satisfies* monotonicity with respect to own ability *if $\frac{\mathrm{d}a(e,b_1,\theta)}{\mathrm{d}e} < \frac{\mathrm{d}a(e,b_2,\theta)}{\mathrm{d}e}$.*

Assumption 1 implies that the marginal (achievement) return to effort is larger for students of higher ability.

Proposition 2. *When $a(e, b, \theta)$ satisfies monotonicity with respect to own ability b and mean pear ability θ is held constant, in equilibrium $l_1^* > l_2^*$ implies $b_1 < b_2$.*

Proof. By the concavity of the achievement function, if $l_1^* > l_2^*$,

$$a'(T \cdot (1 - l_1^*), b_1, \theta) > a'(T \cdot (1 - l_2^*), b_1, \theta). \tag{2}$$

By the concavity of the utility function, $l_1^* > l_2^*$ also implies that $u'(l_1^*) < u'(l_2^*)$. Hence, by Proposition 1:

$$a'(T \cdot (1 - l_1^*), b_1, \theta) < a'(T \cdot (1 - l_2^*), b_2, \theta). \tag{3}$$

Combining 2 and 3, we obtain

$$a'(T \cdot (1 - l_2^*), b_1, \theta) < a'(T \cdot (1 - l_2^*), b_2, \theta)$$

which, by Assumption 1, implies that $b_1 < b_2$. □

Proposition 2 implies that *ceteris paribus* students with lower ability will find it optimal to spend more time on leisure than students with higher ability (i.e., *ceteris paribus* students with lower ability will find it optimal to exert less effort than students with higher ability). This result is supported by physiological research that has found that students with higher levels of perceived ability tend to show higher levels of motivation that is reflected in higher effort (see Schunk and DiBenedetto 2020; Muenks et al. 2016; Wigfield and Eccles 2000).

Assumption 2. *Let two levels of peer mean ability θ_1 and θ_2 satisfy $\theta_1 < \theta_2$. An achievement function $a(e, b, \theta)$ satisfies* monotonicity with respect to mean peer ability *if* $\frac{\mathrm{d}a(e,b,\theta_1)}{\mathrm{d}e} < \frac{\mathrm{d}a(e,b,\theta_2)}{\mathrm{d}e}$.

Assumption 2 establishes that the marginal (achievement) return to effort is larger in environments where the mean peer ability is higher.

Proposition 3. *When $a(e, b, \theta)$ satisfies monotonicity with respect to mean peer ability θ and own ability b is held constant, in equilibrium $l_1^* > l_2^*$ implies $\theta_1 < \theta_2$.*

Proof. The proof is analogous to that of Proposition 2, replacing b_1, b_2 with θ_1, θ_2 and using Assumption 2 instead of Assumption 1. □

Proposition 3 for mean peer ability is analogous to Proposition 2 for own ability. In line with the findings in Carbonaro (2005), it implies that *ceteris paribus* students in an environment of higher mean peer ability (e.g., enrolled in higher tracks) will find it optimal to exert more effort than students in an environment of lower mean peer ability.

The combination of Propositions 2 and 3 implies that, if achievement functions are monotonic with respect to own ability and mean peer ability, *ceteris paribus* student achievement will be strictly increasing in own ability, but students with lower ability can attain a larger achievement in a context of higher mean peer ability than students with higher ability in a context of lower mean peer ability.

3.3 Introducing Socialization

Let ω be the level of student socialization with peers, and assumed to be automatically determined by the student's context (i.e., exogenous). Let $x = (b, \theta, \omega)$ and then an achievement $a(e, b, \theta, \omega)$ is a function of own ability b, peer mean ability θ and the level of student socialization ω in addition to effort.

In order to recognize the complex relationship between socialization and achievement, and so as to maximize applicability of the model to different contexts, socialization with peers can be defined as *distractive* or *productive*.

3.3.1 Distractive Socialization

Assumption 3. *Let two exogenous levels of socialization among peers, ω_1 and ω_2 satisfy $\omega_1 < \omega_2$. Socialization with peers is defined as* distractive *if $\frac{\mathrm{d}a(e,b,\theta,\omega_1)}{\mathrm{d}e} > \frac{\mathrm{d}a(e,b,\theta,\omega_2)}{\mathrm{d}e}$.*

Socialization inside educational institutions has been associated with higher procrastination in some contexts (see for example Day et al. 2000; Gardner and Steinberg 2005; O'Brien et al. 2011; Gustavson et al. 2014). Assumption 3 models this potential socialization dynamic by establishing that spending more time with peers decreases the marginal return to student effort. Ceteris paribus this results in higher levels of *distractive* socialization being correlated with higher levels of leisure and reduced achievement.

Proposition 4. *If socialization is* distractive *and b and θ are held constant, in equilibrium $l_1^* < l_2^*$ implies $\omega_1 < \omega_2$.*

Proof. The proof is analogous to that of Proposition 2, replacing b_1, b_2 with ω_1, ω_2 and using Assumption 3 instead of Assumption 1. □

3.3.2 Productive Socialization

Assumption 4. *Let two exogenous levels of socialization among peers, ω_1 and ω_2 satisfy $\omega_1 < \omega_2$. Socialization with peers is defined as* productive *if $\frac{\mathrm{d}a(e,b,\theta,\omega_1)}{\mathrm{d}e} < \frac{\mathrm{d}a(e,b,\theta,\omega_2)}{\mathrm{d}e}$.*

In other contexts, it is possible that increased socialization may be conducive to higher effort and achievement (e.g., increased socialization may facilitate coordination with peers or decrease informational asymmetries). Assumption 4 models this alternative potential socialization dynamic by establishing that spending more time with peers increases the marginal return to student effort. Ceteris paribus this results in higher levels of *productive* socialization being correlated with higher levels of effort and increased achievement.

Proposition 5. *If socialization is* productive *and b and θ are held constant, in equilibrium $l_1^* < l_2^*$ implies $\omega_1 > \omega_2$.*

Proof. The proof is analogous to that of Proposition 2, replacing b_1, b_2 with ω_1, ω_2 and using Assumption 4 instead of Assumption 1. □

3.4 Introducing Self-Confidence

Let ρ denote student self-confidence. Let $x = (b, \theta, \omega, \rho)$ and then an achievement $a(e, b, \theta, \omega, \rho)$ is a function of own ability b, peer mean ability θ, the level of student socialization ω and self-confidence ρ in addition to effort.

Assumption 5. *An achievement function $a(e, b, \theta, \omega, \rho)$ satisfies* monotonicity with respect to self-confidence *if for $\rho_1 < \rho_2$, $\frac{\mathrm{d}a(e,b,\theta,\omega,\rho_1)}{\mathrm{d}e} < \frac{\mathrm{d}a(e,b,\theta,\omega,\rho_2)}{\mathrm{d}e}$.*

Zárate (2019) introduces the idea of self-confidence to explain academic success, arguing that student self-confidence has a direct positive impact on achievement. Assumption 5 models this proposed relationship, establishing that higher student self-confidence increases the marginal return to student effort. Ceteris paribus this results in higher levels of self-confidence being correlated with higher levels of effort and increased achievement.

Proposition 6. *If an achievement function* $a(e, b, \theta, \omega, \rho)$ *satisfies* monotonicity with respect to self-confidence *and* b, θ *and* ω *are held constant, in equilibrium* $l_1^* > l_2^*$ *implies* $\rho_1 < \rho_2$.

Proof. The proof is analogous to that of Proposition 2, replacing b_1, b_2 with ρ_1, ρ_2 and using Assumption 5 instead of Assumption 1. □

However, self-confidence may be negatively affected by other factors which on their own may have a direct positive impact on academic performance (e.g., mean peer ability). In order to model this potential dynamic, and following the findings in Zárate (2019), we define self-confidence as the difference between student own ability and mean peer ability:

$$\rho = b - \theta.$$

Then, the impact of ability and peer's ability on achievement is as follows:

$$\frac{\mathrm{d}a}{\mathrm{d}b} = \underbrace{\left(\frac{\partial a}{\partial e}\frac{\partial e}{\partial b} + \frac{\partial a}{\partial b}\right)}_{>0 \text{ by Proposition 2}} + \underbrace{\left(1 + \frac{\partial a}{\partial e}\frac{\partial e}{\partial \rho}\right)}_{>0 \text{ by Proposition 6}} > 0.$$

$$\frac{\mathrm{d}a}{\mathrm{d}\theta} = \underbrace{\left(\frac{\partial a}{\partial e}\frac{\partial e}{\partial \theta} + \frac{\partial a}{\partial \theta}\right)}_{>0 \text{ by Proposition 3}} - \underbrace{\left(1 + \frac{\partial a}{\partial e}\frac{\partial e}{\partial \rho}\right)}_{>0 \text{ by Proposition 6}} \gtrless 0.$$

Ceteris paribus the introduction of self-confidence unambiguously augments the overall positive effect of student own ability b on effort and achievement (i.e., the direct positive impact of higher student own ability is indirectly reinforced through the positive effect of student own ability on self-confidence, which in turn also increases effort and achievement). However, the introduction of self-confidence has an ambiguous impact on the overall effect of peer mean ability θ (i.e., the direct positive impact of higher peer mean ability is indirectly countered through the negative effect of higher peer mean ability on self-confidence, which decreases effort and achievement).

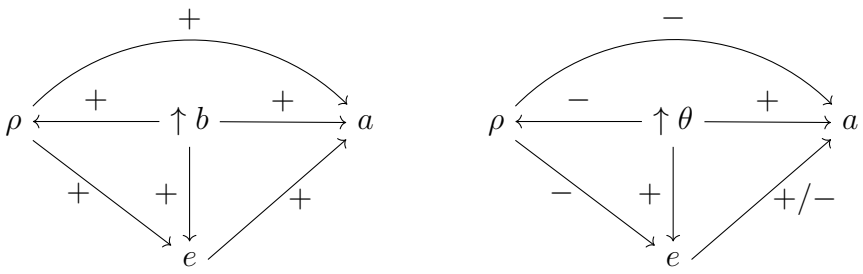


Figure 1: Impact on achievement of increments in b and θ.

3.5 Introducing Persistence

The model can be further expanded by incorporating rounds over an extended period of time over which students accumulate average achievement (i.e., cumulative GPA). The intuition is as follows: students start their program at $t = 1$ and face periodical examinations, the results of which are added to the calculation of their cumulative GPA. This constitutes the relevant variable for students, i.e. the level of student success depends on the average performance over the duration of the program.

Assuming that each semester has a duration of K subperiods and that the total duration of the program is $K \times T$ periods, for each semester t students choose $\{l_{kt}, e_{kt}\}_{k=1}^{K}$ such that:

$$\max \sum_{k=1}^{K} u(l_{k,1}) + a(\sum_{k=1}^{K} e_{k,1}, x) + \sum_{t=2}^{T} \beta^t \left(\sum_{k=1}^{K} u(l_{k,t}) + \left(\frac{\sum_{s=1}^{t} a(\sum_{k=1}^{K} e_{k,s}, x)}{t} \right) \right)$$

subject to $l_{k,t} + e_{k,t} = 1$ for every t, k.

Simplifying and rewriting the Bellman equation, a student that starts period $t \in \{1, \cdots, T\}$ with an average achievement for the $t-1$ periods equal to $\bar{a}$, solves:

$$V(\bar{a}, t) = \max_{\{l_{k,t}\}_{k=1}^{K}} \sum_{k=1}^{K} u(l_{k,t}) + \frac{a(\sum_{k=1}^{K} 1 - l_{k,t}, x) + (t-1)\bar{a}}{t} + \beta V(\bar{a}', t+1)$$

subject to

$$\bar{a}' = \frac{a(\sum_{k=1}^{K} 1 - l_{k,t}, x) + (t-1)\bar{a}}{t}.$$

Proposition 7 follows from solving the above-stated problem.

Proposition 7. *In equilibrium, it is held that for every $t \in \{1, \cdots, T\}$, and every $k \in \{1, \cdots, K\}$,*

$$l_{kt}^* = l_t^*$$

where l_t^* satisfies

$$u'(l_t^*) = a'(K \cdot (1 - l_t^*), x) \left(\frac{1}{t} + \beta \frac{1}{t+1} \right).$$

Further the equilibrium is unique.

Taking Proposition 7 into consideration, it follows that optimal leisure increases with time over the duration of the academic program.

Proposition 8. *In equilibrium, for every $t \in \{1, \cdots, T\}$, $l_{t-1}^* < l_t^* < l_{t+1}^* < \ldots < l_T^*$.*

Proof. By Proposition 7, we have

$$u'(l_t^*) = a'(K \cdot (1 - l_t^*), x) \left(\frac{1}{t} + \frac{\beta}{t+1} \right)$$
$$u'(l_{t+1}^*) = a'(K \cdot (1 - l_{t+1}^*), x) \left(\frac{1}{t+1} + \frac{\beta}{t+2} \right).$$

On the contrary, if $l_t^* \geq l_{t+1}^*$, then by the concavity of u and a, we have $u'(l_t^*) \leq u'(l_{t+1}^*)$ and $a'(K \cdot (1 - l_t^*), x) \geq a'(K \cdot (1 - l_{t+1}^*), x)$. Because $\left(\frac{1}{t} + \frac{\beta}{t+1}\right) > \left(\frac{1}{t+1} + \frac{\beta}{t+2}\right)$, the above two equations would not hold simultaneously, and the proposition is proven by contradiction. □

For each $y \in \{b, \theta, \omega, \rho\}$, $\left|\frac{\partial a_t}{\partial y}\right|$ can be thought of as the impact of y on (current) achievement a_t. Then, by Proposition 8 it follows that the impacts of student own ability, peer mean ability, socialization and self confidence, discussed in previous subsections, are strictly positive for any $T < \infty$ and in this sense, persistent over the duration of the academic program.

Proposition 9. *For each $t \in \{1, \ldots, T\}$, we have $\left|\frac{\partial a_t}{\partial y}\right| < \left|\frac{\partial a_{t+1}}{\partial y}\right|$ for each $y \in \{b, \theta, \omega, \rho\}$.*

Proof. Let $a_t = a\left(K(1 - l_t^*), b, \theta, \omega, \rho\right)$ and $a_{t+1} = a\left(K(1 - l_{t+1}^*), b, \theta, \rho\right)$. By Proposition 8, $K(1 - l_t^*) > K(1 - l_{t+1}^*)$. Then, by the concavity of an achievement function in effort it follows that $\left|\frac{\partial a_t}{\partial y}\right| < \left|\frac{\partial a_{t+1}}{\partial y}\right|$ with each $y \in \{b, \theta, \omega, \rho\}$. □

Proposition 10. *Let $\bar{a} = \sum_{t=1}^{T} \frac{a_t}{T}$ where $a_t = a\left(K(1 - l_t^*), b, \theta, \omega, \rho\right)$. Then, if $\left|\frac{\partial a_0}{\partial y}\right| > 0$ we have $\left|\frac{\partial \bar{a}}{\partial y}\right| > 0$ for each $y \in \{b, \theta, \omega, \rho\}$.*

Proof. For any $y \in \{b, \theta, \omega, \rho\}$, $\frac{\partial \bar{a}}{\partial y} = \frac{1}{T} \sum_{t=1}^{T} \frac{\partial a_t}{\partial y}$. Using Proposition 9, if $\left|\frac{\partial a_0}{\partial y}\right| > 0$, then $\left|\frac{\partial a_t}{\partial y}\right| > 0$ for any $t \in \{1, \cdots, T\}$. As a consequence, for $T < \infty$, $\left|\frac{\partial \bar{a}}{\partial y}\right| > 0$. □

4 Natural Experiment

After proposing a flexible theoretical framework to model the impact of peers on academic performance in different educational contexts, this paper takes advantage of a natural experiment in order to illustrate how the theoretical framework can be used to gain additional empirical insights from reduced-form econometric analysis.

4.1 Context

During the study period freshman undergraduate students in the Engineering and Commercial Engineering (i.e., Business Administration/Economics) degrees at Pontificia Universidad Católica de Chile are ranked according to their weighted admission score, and then randomly assigned to one of the first semester class groups (the latter are referred to as *secciones*, or "sections"). This random assignment generates limited exogenous variation in class composition, allowing to robustly estimate the impact of some peer characteristics on undergraduate academic performance.

In addition, on top of detailed academic performance and socioeconomic status data, the natural experiment also benefits from the availability of a standardized admission score for all students, obtained through a national examination. The introduction of this covariate in the analysis, together with the available information regarding the type of secondary school attended[1], allows to better address the Angrist critique discussed in Section 2 (even if, as mentioned, Feld and Zölitz (2017) show that in a context of random group assignment classical measurement only leads to typical attenuation bias).

Finally, freshman students in each class group share the same classroom when attending their first semester non-elective "core" courses. Core courses are the majority of (and many times the only) courses taken by freshmen during their first college semester, and students in the same class group spend most of their first semester together. It is then plausible that many newly arrived students may form the majority of their college social links during this period, and that assignment to one class group or another may influence the socialization pattern of students during their undergraduate studies (and potentially even after graduation). Also, during the study period it was possible to find (randomly selected) first semester classrooms in both degrees which included up to eight secondary schoolmates. Therefore, the context of the natural experiment is conducive to examine both the potential persistence of peer effects and the impact of peers with a higher likelihood of having an existing social relationship.

4.2 Data

The analysis relies on anonymized administrative data from undergraduate students, collected by the Pontificia Universidad Católica de Chile on a regular basis for academic and administrative purposes. This includes standard administrative data required for the admission process (such as secondary education details and standardized test scores), but also information about socioeconomic status (necessary to determine benefit eligibility). In addition, once students are enrolled in the university their grades, courses taken, class groups and academic status are recorded each semester, in order to allow for the effective monitoring of their progress during

[1] As discussed in Subsection 4.2 below, the type of secondary school attended is considered highly correlated with the quality of secondary education received by students.

their undergraduate studies.

Data is available for all students enrolled in the university from 2000 to 2012.[2] However, the analysis is limited to students admitted from 2000 to 2006, so that a complete record from admission to graduation exists for all the studied cohorts. Also, the analysis is limited to students enrolled in the Engineering and Commercial Engineering (i.e., Business Administration/Economics) degrees. This is because these are among the largest degrees in terms of student enrollment, helping to maximize statistical power[3], but also because other degrees were affected by major curricular changes, not allowing to reliably compare cohorts across time.

Students seeking admission to the university must take a standardized test, administered at the national levels to all Chilean students interested in accessing higher education[4]. The score obtained in this examination was then used to determine admission in a centralized clearing process performed at the national level. First, each university decided whether or not to participate in the centralized admission process (the majority do), and if it did so, it submitted its score-based admission criteria to the centralized authority. Then, after taking the standardized admission examination students were asked to rank their university and degree preferences. Finally, the system cleared in several rounds by allocating the highest ranked students (in terms of admissions score) to their most preferred university-degree choice, and using ranked waiting lists to resolve conflicts (for more details see DEMRE, 2011-2013).

Both the Engineering and the Commercial Engineering degrees at the Pontificia Universidad Católica de Chile were highly competitive during the study period. Weighted admission scores were invariably very close to the maximum possible (850 points), with minimum and maximum admission scores ranging between 705-710 and 825-830 points, respectively. The Engineering degree generally took 6 years to complete, while the Commercial Engineering degree usually took 5 years to complete. There were several financial aid options available for students who met both socioeconomic and academic criteria.

During their secondary schooling students may have attended a public (fully paid for by the government), subsidized (partly paid for by the government) or private (paid for in full by the student) school. Over the study period the quality of secondary education in subsidized schools was considered generally lower than in private ones, while the quality of secondary education in public schools was in turn considered generally lower than in subsidized ones (with

[2]Records exist before the year 2000. However, due to the use of outdated databases and inconsistent data gathering and storage protocols, they were not deemed reliable by the university, and were not made available for the purpose of this analysis.

[3]Ceteris paribus a smaller population size will result in more limited statistical power in hypothesis tests, i.e., an increase in the minimum effect size which can be inferred to be significantly different from zero. Pooling degrees may alleviate this problem, this would only allow to estimate aggregate peer effects, as opposed to degree-specific peer effects.

[4]The standardized examination format changed in 2003 from the former PAA (*Prueba de Aptitud Academica* or "Academic Aptitude Test"), to the new PSU (*Prueba de Selección Universitaria* or "University Selection Test"

a few but important exceptions, most notably the *Instituto National* or "National Institute", an elite secondary school funded by the government).. This, together with other factors such as the high cost of test preparation courses, resulted in students from private secondary schools usually obtaining substantially higher scores in the standardized admission test. Therefore, the majority of students admitted to the most demanded degrees at elite universities (including the two studied in this paper) attended private secondary schools. Moreover, many students who attended the same secondary school then enrolled in the same degree at the same university, sometimes even sharing their first semester college classroom (e.g., in the database used in this analysis some students shared their first semester college classroom with up to eight secondary school classmates).

Table I provides a summary of the number of students by school type in each degree. Each year there were usually four first semester class groups in the Commercial Engineering degree, and at least six class groups in the Engineering degree. Each first semester class group usually featured 50-60 students in the Commercial Engineering degree, and 40-50 students in the Engineering degree. Each year between 200 and 250 students were admitted to the Commercial Engineering degree, while between 250-300 were admitted to the Engineering school.

4.3 Estimation

The exogenous variation in peer characteristics generated by the random assignment of freshman students to their first semester class groups allows to robustly estimate the impact of peer characteristics on undergraduate academic performance.

Linear specifications with fixed effects and clustered standard errors at the class group and year level are used, in order to correct for potential correlation within each class group (class groups are considered distinct across admission years).[5] [6]

Academic performance outcome variables of interest, y_{ijkl}, include: (1) *Graduation* (i.e., whether students graduated); (2) *Drop Out* (i.e., whether students decided to abandon their studies); (3) *Dismissal* (i.e., whether students were dismissed due to poor academic performance); (4) *First Semester GPA*; (5) *First Year GPA*; (6) *Final Undergraduate GPA*.[7]

In all specifications the parameters of interest are the δ coefficients, which identify the impact of independent variables in each specification on the academic performance outcome

[5] More sophisticated specifications were also explored, but the structure of the data is such that, given the available population size, the simpler linear specifications already provide limited statistical power. Therefore, more sophisticated specifications quickly run into standard small population size problems.

[6] In addition, as discussed in Section 4.1 above, additional covariates (e.g. standardized admission score, secondary school type) are included in the analysis in order to better the Angrist critique discussed in Section 2.

[7] A student dropping out during the study period was considered a likely proxy for lack of adaptation to the new environment, while a student being dismissed was considered a likely proxy for gaps academic readiness.

variables of interest.[8]

Six additional student-level individual control variables, $x_{hijkl} h = \{1, ..., 6)\}$, are included for robustness purposes in the extended specifications: (i) *Gender* (1 = Male); (ii) *Weighted Admission Score*; (iii) *Mother's Educational Level*; (iv) *Father's Educational Level*; (v) *Housing Status* (1 = Student lives with both parents); and (vi) *Region* (1 = Santiago Metropolitan Region). Table II provides an overview of the distribution by first semester class groups of these six additional control variables.[9]

4.3.1 Estimating the Impact of Secondary Schoolmate Presence

We first estimate the impact on academic performance of secondary schoolmate presence in a student's first semester college class group. For this purpose two linear functional forms are specified: a baseline specification (which includes only the independent variable of interest and the appropriate fixed effects), and an extended specification (which for robustness purposes includes the six additional control variables). These two linear regression models are respectively represented as

$$y_{ijkl} = \beta_0 + \delta_1 m_{ijkl} + \gamma_{jl} + \mu_{kl} + e_{ijkl} \tag{4}$$

$$y_{ijkl} = \beta_0 + \delta_1 m_{ijkl} + \sum_{h=1}^{6} \beta_h x_{hijkl} + \gamma_{jl} + \mu_{kl} + e_{ijkl} \tag{5}$$

where y_{ijkl} is one of the six college academic performance outcome variables of interest, and m_{ijkl} is an indicator variable equal to one if student i in first semester college class group j shares the classroom with any other students from their same secondary school k who were also admitted to the same degree during academic year l.[10]

Two sets of fixed effects are also specified in the functional form: (a) γ_{jl} accounts for any (observable or unobservable) idiosyncratic characteristics of first semester college class group j in academic year l; (b) μ_{kl} is the number of students from secondary school k admitted in academic year l to the same degree.[11]

[8] $\delta_1 \neq \delta_1'$, so that while for functional forms 4 and 5 it is the case that $\mathbb{E}(y_{ijkl}|m_{ijkl} = 1) - \mathbb{E}(y_{ijkl}|m_{ijkl} = 0) = \delta_1$, for functional forms 6 to 9 it is the case that $\mathbb{E}(y_{ijkl}|m_{ijkl} = 1) - \mathbb{E}(y_{ijkl}|m_{ijkl} = 0) = \delta_1' + \delta_2 s_{ijkl}$.

[9] While, as mentioned, the natural experiment takes advantage of the exogenous variation in peer characteristics generated by the random assignment of freshman students to their first semester class groups to robustly estimate the impact of peer characteristics on undergraduate academic performance, balance in student-level individual control characteristics is desirable for robustness purposes. As discussed in Section 4.5, the distribution of student-level individual control characteristics across first semester class groups appears to be reasonably well, if not perfectly, balanced.

[10] As discussed in Section 4.5, the results of the analysis are robust to the substitution of the binomial presence variable for the actual number of secondary schoolmates in the first semester college class group.

[11] μ_{kl} is included in place of η_{kl}, secondary school fixed effects, because although this identification strategy

4.3.2 Estimating the Impact of Secondary Schoolmate Average Admission Score

We also estimate the impact on academic performance of the average weighted admission score of any secondary schoolmates in a student's first semester college class group.

Two linear functional forms are again specified: a baseline specification (which includes only the independent variable of interest and the appropriate fixed effects), and a extended specification (which for robustness purposes includes the six additional control variables). These two linear regression models are respectively represented as

$$y_{ijkl} = \beta_0 + \delta_1' m_{ijkl} + \delta_2 s_{ijkl} m_{ijkl} + \gamma_{jl} + \eta_{kl} + e_{ijkl} \tag{6}$$

$$y_{ijkl} = \beta_0 + \delta_1' m_{ijkl} + \delta_2 s_{ijkl} m_{ijkl} + \sum_{h=1}^{6} \beta_h x_{hijkl} + \gamma_{jl} + \eta_{kl} + e_{ijkl} \tag{7}$$

where as before y_{ijkl} is one of the six college academic performance outcome variables described above, and m_{ijkl} is an indicator variable equal to one if student i in first semester college class group j shares the classroom with any other students from their same secondary school k who were also admitted to the same degree during academic year l.

However, in these specifications the variable of interest is s_{ijkl}, which represents the average weighted admission score of secondary school k mates of student i in their first semester college class group j in academic year l. This is interacted with m_{ijkl} to account for the fact that many students do not share their first semester college class group with any secondary schoolmates.

In addition, in this case two slightly different sets of fixed effects are specified in these functional forms: (a) γ_{jl} again accounts for any (observable or unobservable) idiosyncratic characteristics of first semester college class group j in academic year l; (b) η_{kl} accounts for any (observable or unobservable) idiosyncratic characteristics of students from secondary school k in academic year l.[12]

4.3.3 Estimating the Impact of Overall Peer Characteristics

Finally, we also estimate the impact on academic performance of characteristics of *all* first semester college class group peers (i.e., as opposed to considering only the characteristics of first semester college class group peers who also attended the same secondary school as the student).

As usual, two linear functional forms are specified: a baseline specification (which includes only the independent variables of interest and the appropriate fixed effects), and a extended

also guarantees that m_{ijkl} satisfies the standard exogeneity assumption, it decreases the number of required fixed effect terms in the specification, increasing the precision of the estimation.

[12] In this case it is not possible to substitute η_{kl} with μ_{kl} to increase precision, given that the latter does not guarantee that s_{ijkl} satisfies the standard exogeneity assumption

specification (which includes the six additional controls for robustness purpose)s. These two linear regression models are respectively represented as

$$y_{ijkl} = \beta_0 + \delta_1' m_{ijkl} + \delta_2 s_{ijkl} m_{ijkl} + \delta_3 s_{ijl} + \delta_4 H_{ijkl} + \delta_5 p_{1ijl} + \delta_6 p_{2ijl} + \eta_{kl} + e_{ijkl} \quad (8)$$

$$y_{ijkl} = \beta_0 + \delta_1' m_{ijkl} + \delta_2 s_{ijkl} m_{ijkl} + \delta_3 s_{ijl} + \delta_4 H_{ijkl} + \delta_5 p_{1ijl} + \delta_6 p_{2ijl} + \sum_{h=1}^{6} \beta_h x_{hijkl} + \eta_{kl} + e_{ijkl} \quad (9)$$

where as before y_{ijkl} is one of the six college academic performance outcome variables described above, m_{ijkl} is an indicator variable equal to one if student i in first semester college class group j shares the classroom with any other students from their same secondary school k who were also admitted to the same degree during academic year l, and s_{ijkl} represents the average weighted admission score of secondary school k mates of student i in their first semester college class group j in academic year l.

Additional explanatory variables are included in these specifications, capturing relevant characteristics of all first semester college class group peers: (a) s_{ijl}, *Average Classmate Admission Score*, represents the average weighted admission score of *all* student i's first semester college class group j mates in academic year l;[13] (b) H_{ijkl}, *Secondary School Concentration*, measures the concentration of secondary schools in class group j in academic year l, excluding student i's secondary school k;[14] (c) p_{1ijl} and p_{2ijl}, *Public Secondary School Classmates* and *Subsidized Secondary School Classmates*, represent the percentage of student i's first semester college class group j mates in academic year l who attended a public or subsidized secondary school, respectively .[15]

In this case only one set of fixed effects is specified in the functional form[16]: η_{kl} accounts for any (observable or unobservable) idiosyncratic characteristics of students from secondary school k in academic year l.[17]

[13] I.e., s_{ijl} refers to the average weighted admission score of all first semester college classmates, while s_{ijkl} refers only to secondary schoolmates in the first semester class group)

[14] Concentration of secondary schools in the class group is measured as the Herfindahl index of secondary school share, constructed by squaring and adding each secondary school's share in the classroom (i.e. the percentage of students who attended it before being admitted to the university). In particular,

$$H_{ijkl} = \sum_{k} \left\{ \frac{n_{jkl}}{\sum_{k'} n_{jk'l}} \right\}^2 \text{ for } k \neq k_i \text{ and } k' \neq k_i'$$

where and s_{kj} is the share of students in class group j who attended secondary school k.

[15] Neither s_{ijl}, H_{ijkl}, p_{1ijl} nor p_{2ijl} include student i.

[16] In this case γ_{jl} class group fixed effects are not included because although neither s_{ijl}, H_{ijkl}, p_{1ijl} nor p_{2ijl} include student i, and therefore differ across students within the same class group, the variation at that level is not enough to avoid acute multi-collinearity problems.

[17] In this case it is again not possible to substitute η_{kl} with μ_{kl} to increase precision, as the latter does not guarantee that s_{ijl}, H_{ijkl}, p_{1ijl} and p_{2ijl} satisfy the standard exogeneity assumption.

4.4 Results

4.4.1 Impact of Secondary Schoolmate Presence

We find that there was a significant negative, persistent impact of secondary schoolmate presence in the first semester college classroom, both in terms of lower grades in the medium and long term, and of an increased likelihood of being dismissed due to poor academic performance.

As shown in Table III.1, ceteris paribus students who shared their first semester college class group with a secondary schoolmate were 2.7% less likely to graduate (in particular, because they were 1.5% more likely to be dismissed due to poor academic performance). Also, students who shared their first semester college class group with a secondary schoolmate had first year and final undergraduate GPAs which were respectively 0.39 and 0.5 points lower.[18]

Estimated coefficients are significant with a 90% confidence, and as shown in Table III.2, these results are qualitatively robust to the inclusion of additional student-level individual control variables.[19]

No significant impact is observed on the likelihood of a student choosing to drop out. Although this may be attributable to limited statistical power, it suggests that the negative impact of secondary schoolmate presence in the first semester college class group may be attributable to academic adaptation problems, which resulted in poor academic performance (as opposed to social adaptation problems, which were considered to be more likely to result in students choosing to abandon their undergraduate studies).

4.4.2 Impact of Secondary Schoolmate Average Admission Score

We observe that a higher average admission score of secondary schoolmates (if any) in the first semester college class group may have had a negative, persistent impact on short, medium and long term grades. This would suggest that the negative impact of secondary schoolmate presence in the first semester college class group was augmented when those secondary schoolmates attained higher admission scores.

As shown in Table IV.1, ceteris paribus an additional average weighted admission score point of secondary schoolmates in the first semester class group translated into between 0.003 and 0.002 less GPA points in the short and medium/long term, respectively. These coefficients are significant with a 95% confidence in the case of short and medium term grades, and with a 90% confidence in the case of long term grades.

[18] In the Chilean educational system grades range from 1 to 7, which are respectively the lowest and highest possible scores, and 4 generally is the lowest passing grade.

[19] As expected, the precision of the analysis decreases with the inclusion of additional student-level individual control variables, but estimated coefficients remain significant, their sign does not change, and magnitudes are roughly comparable.

However, in this case the results are not robust to the inclusion of additional student-level individual control variables.[20]

4.4.3 Impact of Overall Peer Characteristics

Average Classmate Admission Score

We find some evidence of a significant and persistent negative impact of the average admission score of first semester college classmates on academic performance.

As shown on Table V.1, ceteris paribus an additional average weighted admission score point of first semester college classmates translated into a reduction of between 0.041 and 0.026 GPA points in the short and medium term, and into 0.018 less GPA points in the long term. These coefficients are significant with a 99% confidence in the case of short and medium term grades, and with a 95% confidence in the case of long term grades.

However, in this case the results are again not robust to the inclusion of additional student-level individual control variables.[21]

Secondary School Concentration

We find a persistent and significant negative impact of secondary school concentration in the first semester college classroom on both short/long term grades and the likelihood of dismissal. This suggests that having a large proportion of first semester college classmates who attended the same few secondary schools was detrimental to academic performance.

As shown on Table V.1, ceteris paribus an increase of one decimal point in secondary school concentration increased a student's likelihood to be dismissed by 14 percentage points. Also, it resulted in 0.97 and 0.79 less GPA points in the short and medium term, respectively.

These coefficients are significant with a 99% and 90 % confidence for short/long term grades and dismissal likelihood, respectively. As shown on Table V.2 these results are qualitatively robust to the inclusion of additional student-level individual control variables.[22]

4.5 Robustness

As discussed in Section 4.4, empirical results regarding the impact of *Secondary Schoolmate Presence* and *Secondary School Concentration* on undergraduate academic performance are qualitatively robust to the inclusion of additional student-level individual control variables

[20] With the inclusion of additional student-level individual control variables coefficients of interest become insignificant, and in the case of short term grades they change sign.

[21] With the inclusion of additional student-level individual control variables the coefficients of interest again become insignificant, although they do not change sign.

[22] Again, as expected, the inclusion of additional student-level individual control variables decreases the precision of the analysis, but estimated coefficients remain significant and do not change sign, and their magnitudes are roughly comparable.

(i.e., estimated coefficients remain significant and do not change sign, and their magnitudes are roughly comparable). However, empirical results concerning the impact of *Secondary Schoolmate Average Admission Score* and *Average Classmate Admission Score* are not robust to the inclusion of additional student-level individual control variables (i.e., estimated coefficients become not significant, and in some instances change sign).

In addition, the main results presented in this paper are robust to the use of Huber-White heteroskedasticity-consistent estimation, instead of clustered standard errors at the class group level. Also, results regarding the impact of *Secondary Schoolmate Presence* on academic performance are robust to alternative specifications.[23] Results regarding the impact of *Average Classmate Admission Score* are robust to the inclusion of *Secondary Schoolmate Presence* as an independent variable in the specification. Results regarding the impact of *Secondary School Concentration* are also robust to the use of alternative measures of school concentration.[24]

When looking separately at students admitted to each of the two degrees under study, the results are generally consistent with the presented overall findings.[25] Similarly, when allowing for the coefficients of interest to vary with the type of secondary school attended the results are qualitatively comparable to the presented findings.[26]

Despite the multidimensionality of the data and the reduced size of first semester class groups, the distribution of student-level individual control characteristics across first semester class groups appears to be reasonably well, if not perfectly, balanced (see Table II). These additional control variables are not guaranteed to be exogenous in the context of the analysis. Therefore, their estimated coefficients are not discussed, although they are generally consistent with relevant existing findings in the literature and anecdotal evidence.[27]

Finally, although this paper analyzes a population consisting of several thousand observations, the complexity of the data structure still results in limited statistical power. Therefore, the discussion of empirical results focuses on significant impacts found, as opposed to observed non-significant relationships. Similarly, point estimates should be used with caution, as they they are generally more sensitive to small population problems than the sign of the

[23] E.g., including school*year*degree fixed effects, or using the number of secondary school class mates instead of a binomial simply denoting presence or absence.

[24] E.g., number of schools with more than a 5% or 10% share of students in the classroom, or the share of students in the class group belonging to the top 1, top 2, top 3, top 5 and top 10 most represented schools.

[25] Cross-degree comparisons are limited by the non-random assignment of students to each degree. In addition, small population size problems are aggravated when looking at each degree separately.

[26] This line of analysis is limited by the reduced statistical power to detect differentiated impacts by school type, and near-complete determination problems.

[27] E.g., during the study period the weighted average admission score was positively correlated with academic performance. Also, students from public and subsidized secondary schools were less likely to graduate and have lower grades, while male students attained lower grades than female ones. The educational level of parents was positively correlated with academic performance, and students who lived with their parents performed better during their undergraduate studies. Finally, students from the Santiago metropolitan region achieved a better academic performance than their peers from other areas of the country.

estimated coefficients. And, while the identification strategy based on the natural experiment assures the internal validity of the empirical results, caution should be used when attempting to extrapolate results to other contexts.

5 Theoretical Interpretation of Empirical Findings

The proposed theoretical framework can be used to gain additional insights from the empirical results obtained through the reduced-form econometric analysis of the natural experiment. However, as with the empirical results, these additional insights are context specific and may not necessarily extrapolate to other countries, to other Chilean universities, or even to other degrees at the Pontificia Universidad Católica de Chile.

5.1 Socialization

Assuming that sharing the classroom with peers who attended the same secondary school increased the likelihood of socialization, the empirically observed negative impact on academic performance of secondary schoolmate presence in the first semester college classroom suggests that, in the context of the Engineering and Commercial Engineering degrees at the Pontificia Universidad Católica de Chile during the study period, socialization was in overall terms *distractive* (as opposed to *productive*). I.e., in terms of the proposed theoretical framework, the empirical results from the natural experiment suggest that in this context spending more time with peers decreased the marginal return to student effort, validating Assumption 3 over Assumption 4. This resulted in higher levels of leisure and reduced achievement, in line with the positive correlation between socialization and procrastination observed by Day et al. (2000), Gardner and Steinberg (2005), O'Brien et al. (2011), or Gustavson et al. (2014) in several other contexts.

Moreover, the fact that a higher concentration of students from the same secondary schools in the first semester college class group is also estimated to have a negative impact on all students, irrespective of the secondary school they attended, suggests that in this context the structure of the group increased *distractive* socialization for all students. However, even if this negatively impacted the overall academic performance of the first semester class group (i.e. grades and likelihood of dismissal), no significant effect on adaptation (i.e. likelihood of abandonment) is observed.

5.2 Peer Ability and Self-Confidence

The empirically observed negative impact of a higher average admission score (of both all classmates and only secondary schoolmates in the first semester college class group) suggests that, in the context of the natural experiment, the direct positive impact of peer mean ability

on academic performance was more than eclipsed by the negative effect of higher peer mean ability on self-confidence. I.e., in terms of the proposed theoretical framework, peer mean ability $\hat{\theta}^*$ is such that the negative effect of peer mean ability through decreased self-confidence is larger than its direct positive effect.

This is in line with the findings of Zárate (2019) in the context of selective high schools in Peru. However, because this proposed theoretical relationship is not linear, in some contexts a lower $\hat{\theta}^*$ could be such that the direct positive effect of peer mean ability is larger than its indirect negative effect through self-confidence (i.e., a small deterioration in achievement due to decreased peer mean ability could be more than compensated by the indirect improvement through increased self-confidence).

5.3 Persistence

The fact that first semester college classroom characteristics (presence/concentration of secondary schoolmates, average admission scores) are estimated to have had a significant impact on academic performance throughout a student's undergraduate years indicates that, in the context of the natural experiment, peer effects were persistent and the first college semester was key for student academic achievement. Therefore, in terms of the proposed theoretical framework, the empirical evidence suggests that persistence is indeed a desirable feature of peer effects models.

6 Conclusion

First, this paper introduces a flexible theoretical framework to model the persistent effect of peers on academic performance in different educational contexts. In line with the main body of the existing literature, we propose a direct positive impact of peer ability on academic achievement. However, following Zárate (2019), we also incorporate an indirect negative impact of peer ability through reduced self-confidence. In addition, we introduce two potential types of socialization, *distractive* and *productive*, which are respectively predicted to decrease and increase effort and academic achievement. We further expand the theoretical framework to model persistence by incorporating rounds over an extended period of time.

Then we take advantage of a natural experiment in order to illustrate how the proposed theoretical framework can be used to gain additional empirical insights from reduced-form econometric analysis. In particular, we leverage the exogenous variation in peer characteristics generated by the random assignment of freshman students to their first semester class groups in the Engineering and Commercial Engineering degrees at Pontificia Universidad Católica de Chile, which allows us to robustly estimate the impact of peer characteristics on undergraduate academic performance.

We observe a negative impact on academic performance of secondary schoolmate presence in the first semester college classroom. Assuming that sharing the classroom with peers who attended the same secondary school increased the likelihood of socialization, this finding suggests that in the study context socialization was in overall terms *distractive*. This is in line with the positive correlation between socialization and procrastination observed by Day et al. (2000), Gardner and Steinberg (2005), O'Brien et al. (2011), or Gustavson et al. (2014) in several other contexts. Similarly, we find that a higher concentration of students from the same secondary schools in the first semester college class group had a negative impact on all students, irrespective of the secondary school they attended. In terms of the proposed theoretical framework, this suggests that in the study context the structure of the group increased *distractive* socialization for all students.

We also find some evidence of a negative impact of higher average admission scores (of both all classmates and only secondary schoolmates in the first semester college class group) on academic performance. This suggests that, in the study context, the direct positive impact of peer mean ability on academic performance was more than eclipsed by the negative effect of higher peer mean ability on self-confidence. This is in line with the findings of Zárate (2019) in the context of selective high schools in Peru.

Finally, first semester college classroom characteristics are estimated to have had a persistent, significant impact on academic performance throughout a student's undergraduate years. This suggests that persistence is indeed a desirable feature of peer effects theoretical frameworks, and that in the study context the first college semester was key for student academic achievement.

While context-specific, these empirical results yield actionable evidence to guide further policy experimentation and research on peer effects in education, particularly on the impact of socialization, the role of self-confidence, and the persistence of these factors. Most importantly, the proposed theoretical framework reflects emerging empirical findings in the literature, and it can be adapted to model the impact of peers on academic performance in different educational contexts.

TABLE I

DISTRIBUTION OF FIRST SEMESTER COLLEGE STUDENTS

	Commercial Engineering Degree					Engineering Degree				
		Secondary School Type					Secondary School Type			
Year	Group	Public	Subsidized	Private	Total	Group	Public	Subsidized	Private	Total
2000	1	1	3	51	55	1	3	2	45	50
	2	4	7	41	52	2	3	2	43	48
	3	1	6	45	52	3	4	1	42	47
	4	1	9	43	53	4	3	7	40	50
	…	…	…	…	…	5	7	4	37	48
	…	…	…	…	…	6	1	3	44	48
	…	…	…	…	…	7	2	5	38	45
	…	…	…	…	…	8	1	5	39	45
	All	7	25	180	212	All	24	29	328	381
2001	1	6	3	45	54	1	5	4	41	50
	2	2	1	50	53	2	3	4	39	46
	3	2	3	49	54	3	3	5	38	46
	4	4	2	47	53	4	5	3	35	43
	…	…	…	…	…	5	7	2	41	50
	…	…	…	…	…	6	2	1	45	48
	All	14	9	191	214	All	25	19	239	283
2002	1	6	3	42	51	1	0	4	44	48
	2	2	12	38	52	2	4	6	39	49
	3	5	4	44	53	3	6	5	35	46
	4	3	3	51	57	4	4	5	40	49
	…	…	…	…	…	5	3	3	39	45
	…	…	…	…	…	6	1	2	47	50
	All	16	22	175	213	All	18	25	244	287
2003	1	4	4	43	51	1	4	3	42	49
	2	0	3	48	51	2	3	8	36	47
	3	1	7	45	53	3	5	4	37	46
	4	4	4	41	49	4	5	6	35	46
	…	…	…	…	…	5	6	3	36	45
	…	…	…	…	…	6	1	2	45	48
	All	9	18	177	204	All	24	26	231	281
2004	1	0	2	52	54	1	6	3	39	48
	2	1	6	49	56	2	9	4	32	45
	3	2	5	43	50	3	7	7	35	49
	4	2	3	53	58	4	4	5	39	48
	…	…	…	…	…	5	5	2	40	47
	…	…	…	…	…	6	4	4	41	49
	All	5	16	197	218	All	35	25	226	286
2005	1	3	3	50	56	1	1	8	41	50
	2	0	5	49	54	2	7	3	35	45
	3	3	6	45	54	3	3	5	41	49
	4	1	2	53	56	4	4	4	34	42
	…	…	…	…	…	5	3	4	40	47
	…	…	…	…	…	6	4	5	37	46
	All	7	16	197	220	All	22	29	228	279
2006	1	0	5	51	56	1	6	2	40	48
	2	2	3	47	52	2	4	6	36	46
	3	3	0	47	50	3	3	6	39	48
	4	2	4	48	54	4	8	6	35	49
	…	…	…	…	…	5	4	3	40	47
	…	…	…	…	…	6	7	4	34	45
	…	…	…	…	…	7	2	5	38	45
	…	…	…	…	…	8	7	3	35	45
	All	7	12	193	212	All	41	35	297	373
Total		123	224	2427	2774	Total	317	326	3142	3785

NOTES. Distribution of first semester college students in the Commercial Engineering and Engineering degrees at one of the leading Chilean universities. The data set has been constructed using the administrative data routinely gathered by the university from 2000 to 2012, and it includes all students who entered the Commercial Engineering and Engineering degrees at the university via ordinary admission process between 2000 and 2006.

TABLE II

BALANCE BY ADMISSION YEAR AND CLASS GROUP

II.A. Commercial Engineering Degree										
Year	Group	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	Obs.
2000	1	727.37	0.44	8.24	8.44	0.85	0.84	0.02	0.05	55
	2	726.47	0.58	7.69	8.17	0.77	0.73	0.08	0.13	52
	3	727.61	0.56	8.06	8.44	0.73	0.85	0.02	0.12	52
	4	726.02	0.66	8.19	8.58	0.77	0.83	0.02	0.17	53
	p-value	0.97	0.13	0.23	0.42	047	0.40	0.25	0.31	
2001	1	730.52	0.61	7.94	8.20	0.87	0.85	0.11	0.06	54
	2	731.72	0.62	7.77	8.60	0.87	0.92	0.04	0.02	53
	3	732.07	0.44	7.91	8.24	0.74	0.83	0.04	0.06	54
	4	729.24	0.58	7.96	8.23	0.73	0.85	0.08	0.04	52
	p-value	0.84	0.22	0.91	0.43	0.12	0.53	0.35	0.75	
2002	1	734.08	0.55	7.90	8.00	0.90	0.86	0.12	0.06	51
	2	732.70	0.50	8.04	8.17	0.77	0.73	0.04	0.23	52
	3	730.85	0.49	8.11	8.42	0.83	0.87	0.09	0.08	53
	4	731.82	0.40	8.10	8.33	0.83	0.84	0.05	0.05	58
	p-value	0.78	0.38	0.90	0.54	0.36	0.21	0.39	0.01	
2003	1	732.32	0.69	8.08	8.43	0.86	0.88	0.08	0.08	51
	2	732.57	0.39	8.06	8.49	0.78	0.92	0.00	0.06	51
	3	730.76	0.53	7.60	8.19	0.79	0.83	0.02	0.13	53
	4	730.08	0.57	7.96	8.35	0.84	0.86	0.08	0.08	49
	p-value	0.86	0.03	0.47	0.73	0.70	0.56	0.10	0.60	
2004	1	744.05	0.52	8.07	8.48	0.80	0.87	0.00	0.04	54
	2	744.76	0.52	7.89	8.25	0.73	0.86	0.02	0.11	56
	3	744.92	0.60	7.86	8.54	0.82	0.90	0.04	0.10	50
	4	744.53	0.57	7.76	8.29	0.79	0.84	0.03	0.05	58
	p-value	1.00	0.80	0.78	0.66	0.72	0.86	0.51	0.41	
2005	1	741.34	0.50	8.27	8.63	0.80	0.79	0.05	0.05	56
	2	742.86	0.54	8.17	8.44	0.87	0.80	0.00	0.09	54
	3	742.54	0.52	7.76	8.33	0.85	0.85	0.06	0.11	54
	4	739.62	0.59	8.50	8.64	0.71	0.80	0.02	0.04	56
	p-value	0.92	0.81	0.05	0.46	0.16	0.83	0.27	0.41	
2006	1	747.90	0.55	8.32	8.61	0.82	0.80	0.00	0.09	56
	2	746.31	0.48	7.69	8.48	0.73	0.88	0.04	0.06	52
	3	746.25	0.62	7.98	8.70	0.72	0.84	0.06	0.00	50
	4	746.67	0.54	8.09	8.74	0.81	0.89	0.04	0.07	54
	p-value	0.98	0.57	0.20	0.59	0.46	0.55	0.38	0.22	

NOTES. The data set has been constructed using the administrative data routinely gathered by the university from 2000 to 2012, and it includes all students who entered the Commercial Engineering degree at the university via ordinary admission process between 2000 and 2006. Assignment of students to their first semester college class group was random. Each cell presents the mean of the balance variable (column) in each class group (row). Balance variables are: (1) weighted admission score, (2) gender (1 = male), (3) mother's educational level, (4) father's educational level, (5) housing status (1 = student lives with both parents), (6) region (1 = Santiago Metropolitan Region), (7) secondary school type (1 = public), (8) secondary school type (1 = subsidized). Reported p-values are for joint orthogonality test across class groups during the same admission year for each of the corresponding balance variables..

II.B. Engineering Degree

Year	Group	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	Obs.
2000	2	739.56	0.86	8.28	8.32	0.70	0.72	0.06	0.04	50
	3	738.34	0.85	7.85	8.13	0.75	0.75	0.06	0.04	48
	4	740.70	0.83	8.26	8.38	0.74	0.85	0.09	0.02	47
	5	739.35	0.80	7.98	8.18	0.84	0.84	0.06	0.14	50
	6	739.01	0.81	8.17	8.50	0.79	0.81	0.15	0.08	48
	7	739.82	0.79	7.69	8.06	0.85	0.92	0.02	0.06	48
	8	738.85	0.76	7.78	8.11	0.78	0.80	0.04	0.11	45
	9	739.41	0.87	7.73	8.24	0.80	0.80	0.02	0.11	45
	p-value	1.00	0.86	0.40	0.83	0.64	0.32	0.26	0.30	
2001	1	753.33	0.92	7.72	7.82	0.86	0.80	0.10	0.08	50
	2	752.57	0.85	7.43	8.39	0.85	0.85	0.07	0.09	46
	3	751.14	0.83	8.07	8.26	0.78	0.85	0.07	0.11	46
	4	749.02	0.91	7.77	8.40	0.86	0.84	0.12	0.07	43
	5	750.31	0.88	7.52	7.84	0.86	0.86	0.14	0.04	50
	7	749.99	0.83	8.08	8.54	0.85	0.79	0.04	0.02	48
	p-value	0.92	0.67	0.32	0.07	0.90	0.93	0.54	0.57	
2002	1	738.03	0.90	7.63	8.13	0.77	0.73	0.00	0.08	48
	2	738.72	0.92	8.04	8.29	0.82	0.84	0.08	0.12	49
	3	738.20	0.87	7.87	8.24	0.76	0.72	0.13	0.11	46
	4	737.26	0.88	7.71	7.96	0.78	0.84	0.08	0.10	49
	5	740.09	0.80	7.69	8.53	0.73	0.78	0.07	0.07	45
	7	737.96	0.78	8.18	8.34	0.84	0.78	0.02	0.04	50
	p-value	0.99	0.32	0.49	0.51	0.84	0.61	0.11	0.73	
2003	1	738.23	0.88	7.84	8.16	0.76	0.82	0.08	0.06	49
	2	739.12	0.79	7.89	8.00	0.70	0.74	0.06	0.17	47
	3	736.76	0.65	8.00	8.30	0.78	0.80	0.11	0.09	46
	4	738.45	0.89	7.33	7.67	0.70	0.67	0.11	0.13	46
	5	738.15	0.87	7.53	8.36	0.76	0.64	0.13	0.07	45
	7	736.55	0.79	8.38	8.56	0.71	0.73	0.02	0.04	48
	p-value	0.99	0.03	0.08	0.24	0.91	0.35	0.44	0.26	
2004	1	757.16	0.88	8.00	8.52	0.83	0.85	0.13	0.06	48
	2	756.79	0.98	7.31	7.73	0.73	0.80	0.20	0.09	45
	3	757.44	0.80	7.67	8.31	0.80	0.76	0.14	0.14	49
	4	755.68	0.90	8.06	8.06	0.75	0.77	0.08	0.10	48
	5	757.35	0.83	8.06	8.28	0.79	0.77	0.11	0.04	47
	7	756.69	0.84	7.96	8.04	0.71	0.80	0.08	0.08	49
	p-value	1.00	0.14	0.24	0.24	0.76	0.87	0.52	0.61	
2005	1	767.66	0.92	7.92	7.96	0.76	0.84	0.02	0.16	50
	2	768.24	0.96	7.89	8.27	0.69	0.80	0.16	0.07	45
	3	766.14	0.88	7.61	7.82	0.80	0.78	0.06	0.10	49
	4	767.75	0.81	7.93	8.31	0.81	0.76	0.10	0.10	42
	5	765.61	0.74	7.87	8.26	0.91	0.74	0.06	0.09	47
	7	767.03	0.89	7.91	8.17	0.76	0.80	0.09	0.11	46
	p-value	0.99	0.04	0.94	0.64	0.17	0.90	0.26	0.77	
2006	1	774.80	0.85	8.10	8.38	0.67	0.85	0.13	0.04	48
	2	774.96	0.74	8.24	8.35	0.70	0.85	0.09	0.13	46
	3	772.46	0.79	7.88	8.58	0.83	0.77	0.06	0.13	48
	4	774.47	0.86	7.61	8.37	0.80	0.76	0.16	0.12	49
	5	774.14	0.83	8.00	8.17	0.79	0.79	0.09	0.06	47
	6	772.69	0.82	8.00	8.20	0.71	0.73	0.16	0.09	45
	7	773.30	0.84	8.07	8.24	0.73	0.73	0.04	0.11	45
	8	774.16	0.93	8.18	8.38	0.71	0.82	0.16	0.07	45
	p-value	1.00	0.41	0.66	0.89	0.57	0.72	0.43	0.74	

NOTES. The data set has been constructed using the administrative data routinely gathered by the university from 2000 to 2012, and it includes all students who entered the Engineering degree at the university via ordinary admission process between 2000 and 2006. Assignment of students to their first semester college class group was random. Each cell presents the mean of the balance variable (column) in each class group (row). Balance variables are: (1) weighted admission score, (2) gender (1 = male), (3) mother's educational level, (4) father's educational level, (5) housing status (1 = student lives with both parents), (6) region (1 = Santiago Metropolitan Region), (7) secondary school type (1 = public), (8) secondary school type (1 = subsidized). Reported p-values are for joint orthogonality test across class groups during the same admission year for each of the corresponding balance variables..

TABLE III

IMPACT ON ACADEMIC PERFORMANCE OF PRESENCE OF SECONDARY SCHOOL MATES IN FIRST SEMESTER COLLEGE CLASS GROUP

		(1)	(2)	(3)	(4)	(5)	(6)
		Graduation	Drop Out	Dismissal	1st Semester GPA	1st Year GPA	Final GPA
(III.1)	1 = Secondary School Mate(s) in Group	-0.027	0.009	0.015	-0.033	-0.039	-0.050
		(0.015)*	(0.012)	(0.009)*	(0.031)	(0.023)*	(0.021)**
	Other Student Characteristics	No	No	No	No	No	No
	R^2	0.04	0.03	0.04	0.07	0.05	0.05
	Observations	3,634	3,634	3,634	3,660	3,625	3,660
(III.2)	1 = Secondary School Mate(s) in Group	-0.025	0.009	0.014	-0.015	-0.026	-0.037
		(0.015)	(0.012)	(0.009)	(0.026)	(0.019)	(0.019)*
	Other Student Characteristics	Yes	Yes	Yes	Yes	Yes	Yes
	R^2	0.07	0.04	0.07	0.31	0.30	0.23
	Observations	3,634	3,634	3,634	3,660	3,625	3,660

* $p<0.1$; ** $p<0.05$; *** $p<0.01$

NOTES. This table analyzes the impact of the presence of secondary school mates in a student's first semester college class group in the Commercial Engineering and Engineering degrees at one of the leading Chilean universities. The assignment to these first semester college class groups is random, allowing to treat peer characteristics as exogenous. The data set has been constructed using the administrative data routinely gathered by the university from 2000 to 2012, and it includes all students who entered the Commercial Engineering and Engineering degrees at the university via ordinary admission process between 2000 and 2006. The impact of the presence of secondary school mates in the first semester college class group is evaluated on six academic performance measures, presented horizontally, in particular: (1) graduation likelihood, (2) drop out likelihood, (3) dismissal likelihood, (4) 1st semester GPA, (5) 1st year GPA and (6) final GPA. Two sets of specifications are presented stacked over each other, with the one above (III.1) including only the variable of interested and the one below (III.2) including other student characteristics as additional controls, namely: gender (1 = male), weighted admission score, mother's educational level, father's educational level, housing status (1 = student lives with both parents) and region (1 = Santiago Metropolitan Region). Both specifications include two sets of fixed effects: (i) admission year * first semester college class group * degree, (ii) number of same secondary school students admitted the same year to the same degree. The latter allows to increase precision by excluding secondary school fixed effects while still ensuring that the secondary school mate presence variable satisfies the exogeneity condition. Standard errors are clustered by admission year * first semester college class group * degree.

TABLE IV

IMPACT ON ACADEMIC PERFORMANCE OF AVERAGE WEIGHTED ADMISSION SCORE OF SECONDARY SCHOOL MATES IN FIRST SEMESTER COLLEGE CLASS GROUP

		(1)	(2)	(3)	(4)	(5)	(6)
		Graduation	Drop Out	Dismissal	1st Semester GPA	1st Year GPA	Final GPA
(IV.1)	Avg. Admission Score of Secondary School Mate(s) in Group (If Any)	0.000	-0.000	0.000	-0.003	-0.002	-0.002
		(0.001)	(0.000)	(0.000)	(0.001)**	(0.001)**	(0.001)*
	1 = Secondary School Mate(s) in Group	-0.210	0.314	-0.114	1.944	1.793	1.383
		(0.419)	(0.371)	(0.264)	(0.954)**	(0.781)**	(0.756)*
	Other Student Characteristics	No	No	No	No	No	No
	R^2	0.59	0.55	0.67	0.56	0.54	0.59
	Observations	3,634	3,634	3,634	3,660	3,625	3,660
(IV.2)	Avg. Admission Score of Secondary School Mate(s) in Group (If Any)	0.001	-0.001	-0.000	0.002	0.001	0.001
		(0.001)	(0.000)	(0.000)	(0.001)*	(0.001)	(0.001)
	1 = Secondary School Mate(s) in Group	-0.691	0.615	0.067	-1.460	-0.961	-0.853
		(0.402)*	(0.371)	(0.251)	(0.781)*	(0.592)	(0.559)
	Other Student Characteristics	Yes	Yes	Yes	Yes	Yes	Yes
	R^2	0.61	0.56	0.67	0.68	0.68	0.68
	Observations	3,634	3,634	3,634	3,660	3,625	3,660

* $p<0.1$; ** $p<0.05$; *** $p<0.01$

NOTES. This table analyzes the impact of the average weighted admission score of secondary school mates in a student's first semester college class group in the Commercial Engineering and Engineering degrees at one of the leading Chilean universities. The assignment to these first semester college class groups is random, allowing to treat peer characteristics as exogenous. The data set has been constructed using the administrative data routinely gathered by the university from 2000 to 2012, and it includes all students who entered the Commercial Engineering and Engineering degrees at the university via ordinary admission process between 2000 and 2006. The impact of the average weighted admission score of secondary school mates in the first semester college class group is evaluated on six academic performance measures, presented horizontally, in particular: (1) graduation likelihood, (2) drop out likelihood, (3) dismissal likelihood, (4) 1st semester GPA, (5) 1st year GPA and (6) final GPA. Two sets of specifications are presented stacked over each other, with the one above (IV.1) including only the variable of interested and the one below (IV.2) including other student characteristics as additional controls, namely: gender (1 = male), weighted admission score, mother's educational level, father's educational level, housing status (1 = student lives with both parents) and region (1 = Santiago Metropolitan Region). Both specifications include two sets of fixed effects: (i) admission year * first semester college class group * degree, (ii) admission year * secondary school * degree. Standard errors are clustered by admission year * first semester college class group * degree.

TABLE V

IMPACT ON ACADEMIC PERFORMANCE OF FIRST SEMESTER COLLEGE CLASS GROUP PEER CHARACTERISTICS

		(1)	(2)	(3)	(4)	(5)	(6)
		Graduation	Drop Out	Dismissal	1st Semester GPA	1st Year GPA	Final GPA
(V.1)	Avg. Admission Score of Other Students in Group	-0.003	0.001	0.002	-0.041	-0.026	-0.018
		(0.003)	(0.003)	(0.002)	(0.010)***	(0.008)***	(0.007)**
	Secondary School Concentration in Group (Herfindahl Index)	-1.255	-0.316	1.401	-9.774	-3.890	-7.973
		(1.156)	(1.156)	(0.796)*	(3.677)***	(3.003)	(2.329)***
	% Public School Students in Group	-0.297	0.074	0.218	-0.507	-0.570	-0.337
		(0.137)**	(0.146)	(0.084)**	(0.486)	(0.370)	(0.362)
	% Subsidized School Students in Group	-0.294	0.083	0.160	0.943	0.668	0.017
		(0.205)	(0.184)	(0.089)*	(0.484)*	(0.383)*	(0.341)
	Avg. Admission Score of Secondary School Mate(s) in Group (If Any)	0.000	-0.000	0.000	-0.002	-0.002	-0.001
		(0.000)	(0.000)	(0.000)	(0.001)	(0.001)**	(0.001)
	1 = Secondary School Mate(s) in Group	-0.320	0.355	-0.049	1.239	1.288	0.900
		(0.344)	(0.308)	(0.225)	(0.780)	(0.647)*	(0.624)
	Other Student Characteristics	No	No	No	No	No	No
	R^2	0.33	0.31	0.33	0.37	0.33	0.33
	Observations	2,677	2,677	2,677	2,690	2,671	2,690
(V.2)	Avg. Admission Score of Other Students in Group	0.002	-0.001	-0.000	-0.009	0.000	0.003
		(0.003)	(0.003)	(0.002)	(0.006)	(0.006)	(0.005)
	Secondary School Concentration in Group (Herfindahl Index)	-0.354	-0.843	1.026	-3.922	0.570	-3.972
		(1.232)	(1.220)	(0.791)	(2.440)	(2.211)	(1.834)**
	% Public School Students in Group	-0.274	0.058	0.211	-0.433	-0.501	-0.281
		(0.141)*	(0.152)	(0.078)***	(0.353)	(0.268)*	(0.273)
	% Subsidized School Students in Group	-0.341	0.117	0.173	0.634	0.393	-0.185
		(0.199)*	(0.184)	(0.086)**	(0.354)*	(0.288)	(0.266)
	Avg. Admission Score of Secondary School Mate(s) in Group (If Any)	0.001	-0.001	-0.000	0.002	0.001	0.001
		(0.000)**	(0.000)*	(0.000)	(0.001)**	(0.001)**	(0.001)**
	1 = Secondary School Mate(s) in Group	-0.720	0.593	0.113	-1.625	-1.061	-1.008
		(0.325)**	(0.306)*	(0.213)	(0.638)**	(0.499)**	(0.473)**
	Other Student Characteristics	Yes	Yes	Yes	Yes	Yes	Yes
	R^2	0.35	0.32	0.34	0.54	0.53	0.47
	Observations	2,677	2,677	2,677	2,690	2,671	2,690

* $p<0.1$; ** $p<0.05$; *** $p<0.01$

NOTES. This table analyzes the impact of first semester college class group peer characteristics on students in the Commercial Engineering and Engineering degrees at one of the leading Chilean universities. The assignment to these first semester college class groups is random, allowing to treat peer characteristics as exogenous. The data set has been constructed using the administrative data routinely gathered by the university from 2000 to 2012, and it includes all students who entered the Commercial Engineering and Engineering degrees at the university via ordinary admission process between 2000 and 2006. Four peer characteristics are studied: (a) average weighted admission score of other students in group, (b) secondary school concentration in the class group, measured as the sum of the square of the percentage of students from each secondary school in the class group, (c) percentage of public secondary school students in the class group, and (d) percentage of subsidized secondary school students in the classroom. The impact of these variables is evaluated on six academic performance measures, presented horizontally, in particular: (1) graduation likelihood, (2) drop out likelihood, (3) dismissal likelihood, (4) 1st semester GPA, (5) 1st year GPA and (6) final GPA. Two sets of specifications are presented stacked over each other, with the one above including only the variable of interested and the one below including other student characteristics as additional controls, namely: gender (1 = male), weighted admission score, mother's educational level, father's educational level, housing status (1 = student lives with both parents) and region (1 = Santiago Metropolitan Region). Both specifications include admission year * secondary school * degree fixed effects. Standard errors are clustered by admission year * first semester college class group * degree.

References

Angrist, J. D. (2014). The perils of peer effects. *Labour Economics*, 30:98–108.

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

Arcidiacono, P., Foster, G., Goodpaster, N., and Kinsler, J. (2012). Estimating spillovers using panel data, with an application to the classroom. *Quantitative Economics*, 3(3):421–470.

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

Carbonaro, W. (2005). Tracking, students' effort, and academic achievement. *Sociology of Education*, 78(1):27–49.

Carrell, S. E., Sacerdote, B. I., and West, J. E. (2013). From natural variation to optimal policy? the importance of endogenous peer group formation. *Econometrica*, 81(3):855–882.

Cooley, J. (2009). Can achievement peer effect estimates inform policy? a view from inside the black box. *Review of Economics and Statistics*, 93.

Day, V., Mensink, D., and O'Sullivan, M. (2000). Patterns of academic procrastination. *Journal of College Reading and Learning*, 30(2):120–134.

Ding, W. and Lehrer, S. F. (2007). Do peers affect student achievement in china's secondary schools? *The Review of Economics and Statistics*, 89(2):300–312.

Duflo, E., Dupas, P., and Kremer, M. (2011). Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in kenya. *American Economic Review*, 101(5):1739–74.

Epple, D. and Romano, R. E. (2011). *Peer effects in education: A survey of the theory and evidence*, volume 1. Elsevier.

Feld, J. and Zölitz, U. (2017). Understanding peer effects: On the nature, estimation, and channels of peer effects. *Journal of Labor Economics*, 35(2):387–428.

Gardner, M. and Steinberg, L. (2005). Peer influence on risk taking, risk preference, and risky decision making in adolescence and adulthood: an experimental study. *Developmental psychology*, 41(4):625.

Garlick, R. (2014). Academic peer effects with different group assignment rules: Residential tracking versus random assignment. Technical report, Mimeo, Duke University.

Graham, B. S. (2008). Identifying social interactions through conditional variance restrictions. *Econometrica*, 76(3):643–660.

Gustavson, D. E., Miyake, A., Hewitt, J. K., and Friedman, N. P. (2014). Genetic relations among procrastination, impulsivity, and goal-management ability: Implications for the evolutionary origin of procrastination. *Psychological science*, 25(6):1178–1188.

Jaffee, D. (2007). Peer cohorts and the unintended consequences of freshman learning communities. *College Teaching*, 55(2):65–71.

Kang, C. et al. (2007). Classroom peer effects and academic achievement: Quasi-randomization evidence from south korea. *Journal of Urban Economics*, 61(3):458–495.

Kimbrough, E. O., McGee, A. D., and Shigeoka, H. (2017). How do peers impact learning? an experimental investigation of peer-to-peer teaching and ability tracking. Technical report, National Bureau of Economic Research.

Kremer, M. and Levy, D. (2008). Peer effects and alcohol use among college students. *Journal of Economic perspectives*, 22(3):189–206.

Muenks, K., Miele, D. B., and Wigfield, A. (2016). How students' perceptions of the source of effort influence their ability evaluations of other students. *Journal of Educational Psychology*, 108(3):438.

O'Brien, L., Albert, D., Chein, J., and Steinberg, L. (2011). Adolescents prefer more immediate rewards when in the presence of their peers. *Journal of Research on adolescence*, 21(4):747–753.

Rothstein, J. (2009). Student sorting and bias in value-added estimation: Selection on observables and unobservables. *Education finance and policy*, 4(4):537–571.

Sacerdote, B. (2001). Peer effects with random assignment: Results for dartmouth roommates. *The Quarterly journal of economics*, 116(2):681–704.

Schunk, D. H. and DiBenedetto, M. K. (2020). Motivation and social cognitive theory. *Contemporary Educational Psychology*, 60:101832.

Shue, K. (2013). Executive networks and firm policies: Evidence from the random assignment of mba peers. *The Review of Financial Studies*, 26(6):1401–1442.

Vigdor, J. and Nechyba, T. (2007). Peer effects in north carolina public schools. *Schools and the equal opportunity problem*, pages 73–101.

Whitmore, D. (2005). Resource and peer impacts on girls' academic achievement: Evidence from a randomized experiment. *American Economic Review*, 95(2):199–203.

Wigfield, A. and Eccles, J. S. (2000). Expectancy–value theory of achievement motivation. *Contemporary educational psychology*, 25(1):68–81.

Zabel, J. E. (2008). The impact of peer effects on student outcomes in new york city public schools. *Education Finance and Policy*, 3(2):197–249.

Zárate, R. A. (2019). Social and cognitive peer effects: Experimental evidence from selective high schools in Peru. *Working Paper at briq/IZA Workshop on Behavioral Economics of Education*.

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