

Goulas, Sofoklis; Megalokonomou, Rigissa; Zhang, Yi

Working Paper

Does the Girl Next Door Affect Your Academic Outcomes and Career Choices?

IZA Discussion Papers, No. 11910

Provided in Cooperation with:

IZA – Institute of Labor Economics

Suggested Citation: Goulas, Sofoklis; Megalokonomou, Rigissa; Zhang, Yi (2018) : Does the Girl Next Door Affect Your Academic Outcomes and Career Choices?, IZA Discussion Papers, No. 11910, Institute of Labor Economics (IZA), Bonn

This Version is available at:

<https://hdl.handle.net/10419/193204>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

DISCUSSION PAPER SERIES

IZA DP No. 11910

**Does the Girl Next Door Affect Your
Academic Outcomes and Career Choices?**

Sofoklis Goulas
Rigissa Megalokonomou
Yi Zhang

OCTOBER 2018

DISCUSSION PAPER SERIES

IZA DP No. 11910

Does the Girl Next Door Affect Your Academic Outcomes and Career Choices?

Sofoklis Goulas

Hoover Institution and Stanford University

Rigissa Megalokonomou

University of Queensland and IZA

Yi Zhang

University of Queensland

OCTOBER 2018

Any opinions expressed in this paper are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but IZA takes no institutional policy positions. The IZA research network is committed to the IZA Guiding Principles of Research Integrity.

The IZA Institute of Labor Economics is an independent economic research institute that conducts research in labor economics and offers evidence-based policy advice on labor market issues. Supported by the Deutsche Post Foundation, IZA runs the world's largest network of economists, whose research aims to provide answers to the global labor market challenges of our time. Our key objective is to build bridges between academic research, policymakers and society.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ABSTRACT

Does the Girl Next Door Affect Your Academic Outcomes and Career Choices?¹

Gender peer effects are potentially important for optimally organizing schools and neighborhoods. In this paper, we examine how the gender of classmates and neighbors affects a variety of high school outcomes and choice of university major. Given that students are assigned to schools based on proximity from their residential address, we define as neighbors all same-cohort peers who attend any other school within a 1-mile radius of one's school. To control for potentially confounding unobserved characteristics of schools and neighborhoods that might be correlated with peer gender composition, we exploit within-school and -neighborhood idiosyncratic variation in gender composition share across consecutive cohorts in the 12th grade. Using data for the universe of students in public schools in Greece between 2004 and 2009, we find that a higher share of females in a school or neighborhood improves both genders' subsequent scholastic performance, increases their university matriculation rates, renders them more likely to enroll in an academic university than a technical school, and affects their choice of university study. In addition, we find that only females are more likely to enroll in STEM degrees and target more lucrative occupations when they have more female peers in school or neighborhood. Based on our back-of-the-envelope calculations, a 10 percentage point increase in the proportion of females in a school or neighborhood reduces the gender gap in STEM enrollments by 2% and 3%, respectively. We also find that (1) neighborhood peer effects are as large as school peer effects, and (2) the effects are nonlinear-namely, the effects are larger for school and neighborhood cohorts with a large majority of female peers.

JEL Classification: J24, J21, J16, I24

Keywords: gender peer effects, neighborhood effects, STEM university degrees

Corresponding author:

Rigissa Megalokonomou
School of Economics
The University of Queensland
St Lucia Qld 4072
Australia

E-mail: r.megalokonomou@uq.edu.au

¹ We are grateful to Sascha Becker, Dan Bernhardt, Clement de Chaisemartin, Victor Lavy and Fabian Waldinger for the insightful discussions and useful comments. We are indebted to Robert Akerlof, Samia Badji, Simon Burgess, Vasilakis Chrysovalantis, Luke Chu, Anne Cumpston, Jan Feld, Thiemo Fetzer, Alessandro Iaria, Asad Islam, Jan Kabatek, Claudio Labanca, Pushkar Maitra, Andrew McLennan, Robin Naylor, Laura Puzzello, Tina Rampino, Heather Royer, Haishan Yuan and Chris Woodruff for useful suggestions. We also thank participants at the University of Warwick, University of Queensland, University of North Carolina at Chapel Hill, University of Macedonia, Monash University, the AUEB Economics and Econometrics Conference, XXVII Meeting of the Economics of Education Association, Athens University in Economics and Business and the CSDA Workshop in Auckland. Any errors or omissions are our own. Rigissa Megalokonomou acknowledges research support from the University of Queensland BEL Early Career Grant (No: UQECR1833757).

1 Introduction

The question of which peers matter for students’ learning has long been of concern to academics, social scientists, and policymakers (Foster 2006, Carrell et al. 2009). A student’s performance not only depends on their ability and characteristics, but also on the ability and characteristics of other students within their peer environment (Manski 1993). In recent literature, social interactions between genders in the classroom or school have been shown to affect students’ academic achievement (Lavy and Schlosser 2011, Hoxby 2000). However, to our knowledge, empirical research has yet to open the “black box” of physical proximity interactions to understand how those peers’ behavior influences pupils’ outcomes and decisions¹.

We examine the belief that peers in other peer environments –*beyond* the scale of the school area– are important for high school students’ performance and decisions. Unlike college students who might move to an area for a predetermined, and typically short, period of time, high school students are likely to live in a specific area for many years². Neighborhoods are particularly important during adolescence, given youths’ limited geographic mobility during this stage of development (Buu et al. 2009). Thus, adolescents are likely to form natural reference groups with the similar-age peers with whom they interact daily for years within relatively contained geographical areas, like their *neighborhood*. Physical proximity amplifies this interplay of utility spillover from other students’ achievements and choices, and the combined effects are *neighborhood specific*.

Physical proximity may affect a pupil’s scholastic outcomes for several reasons. For example, adolescents might have a desire to conform with others in their reference group due to peer pressure, role models, or social norms (Akerlof and Kranton 2002, Mota et al. 2016). Moreover, students might gain from information transmission (Bikhchandani et al. 1992, Xiong et al. 2016) that they would not have received from schoolmates. Thus, although high school students are likely to be affected by their schoolmates, they are also likely to be affected by broader social networks.

In this paper, we examine whether these broader social networks are important for students’ outcomes and choices. Using 12th grade data for the universe of students in public high schools in Greece between 2004 and 2009 and geographical data, we are able to identify another reference group –the neighborhood–, in which students are very likely to interact with each other. Given that students are assigned to public schools based on the school’s proximity to the student’s residential address, we define as *neighbors* all same-cohort peers who attend any *other* school within 1 mile from one’s school. We then examine how idiosyncratic changes in the proportion of females within the school or the neighborhood affect a student’s academic performance in high school and on university entrance exams. Building on and expanding the literature, we then examine the effect of gender within these peer environments on the choice of study at

¹In an experimental setting, some studies have used interventions in rural Mexico that compare the outcomes of ineligible and eligible households within the same village. In particular, Bobonis and Finan (2009) and Lalive and Cattaneo (2009) find evidence of spillover through neighborhood peer effects in school enrollment decisions.

²They may also return to their parental residential address after college graduation.

the university level and occupation-related expected wages. To do this, we use data that span six cohorts of students who took the end-of-high-school exams in academic years 2003/2004 through 2008/2009. We link our data to information on each pupil’s university admission records and labor force survey earnings data, which we exploit to examine how gender peer effects influence a student’s probability of enrolling in a STEM degree and their expected earnings after graduation.

To study gender neighborhood interactions, we exploit an institutional setting in which schools are built very close to each other. Specifically, the median distance of a school from each nearest neighboring school is 0.32 miles. Similar to almost 85% of OECD countries (OECD 2010), in Greece students by law are assigned to public schools based on proximity from their residential address. Due to the geographical density of schools in Greece, it is likely that same-age peers who live nearby are assigned to different public schools. These peers are likely to interact with and affect others’ academic performance and education choices.³ Using data on the universe of schools and conducting cluster analysis, we define a *neighborhood* as a cluster that contains all same-cohort peers who attend any other high school within a 1-mile radius, except for one’s own high school. Thus, we are able to disentangle a student’s schoolmates from their neighbors.

To our knowledge, this is the first paper that separates and compares the size of social interaction effects between genders across distance in space and age. Contributing to the literature, we are the first to examine gender peer effects in neighborhoods using spatial variation and cluster analysis of geographical data. We also expand on the literature of gender peer effects, by studying the longer-term effects of gender interaction on postsecondary education choices—specifically, the decision to pursue a STEM degree. These contributions are well supported by some unique features of our data and institutional setting. First, the availability of student-level data that contain information on exam scores, college enrollment, and choice of university study for every student in the country in six cohorts allows us to test how high school peers affect students’ university outcomes and education choices. Second, the institutional setting in Greece, where high schools are built very close to each other, allows us to exploit rich variation of school characteristics within a relatively constrained geographical area and, in turn, examine the existence of gender peer effects at the neighborhood level. Third, there are particular advantages in having the universe of high school graduates for a country in multiple cohorts, because we can observe the behavior of all students regarding their education decisions rather than only specific groups of students. We also observe and compare the gender peer effects that arise from different reference groups at different spatial and age levels.

To overcome selection problems, we rely on within-school and -neighborhood regressions—i.e., on specifications that include school and neighborhoods fixed effects—to control for unobserved characteristics of schools or neighborhoods and students that might be correlated with the proportion of females and could also affect the outcomes. As our variation comes from a cohort-by-cohort comparison, it is important to

³Neighbors are likely to play together in their spare time or they are likely to take part in group activities or sporting activities together.

use multiple cohorts and control for unobserved school or neighborhood time-varying unobserved factors that might confound our identification. The basic idea is to compare the outcomes and choices of pupils from consecutive cohorts who are exposed to the same school or neighborhood environments and have similar characteristics, except for the fact that one cohort has more female students than the other for idiosyncratic reasons. We use a variance decomposition to confirm that there is sufficiently large cohort-to-cohort variation in gender composition within schools and neighborhoods. Our balancing tests also provide evidence that this variation is random and not associated with within-school or -neighborhood variation of students' background characteristics, which allows us to obtain more precise and unbiased estimates of gender peer effects. We also perform a Monte Carlo simulation that verifies that the distribution of school/neighborhood-by-year proportion of female students could be generated by a random sample.

Our results show that there is significant and random variation in the proportion of female students across cohorts within schools and neighborhoods. Overall, an increase in the proportion of girls in a high school or neighborhood has positive and significant effects on various high school scholastic outcomes and college admission outcomes for both genders. We find that the effects that arise from neighbors are as large as those from schoolmates. We also find that gender peer effects are, overall, slightly larger for girls compared to boys. We then examine the effects of gender composition on choice of university study. Interestingly, we find that only females who have more female peers in a school or a neighborhood are more likely to enroll in a STEM post-secondary degree. For boys, estimates of the peer effects on the decision to study STEM are small and insignificant. Considering the effects on females, our back-of-the-envelope calculations suggest that a 10 percentage point increase in the share of females within a school or neighborhood reduces the gender gap in STEM degree enrollments by 2% and 3%, respectively. This is in line with our finding that females choose more lucrative occupations in the workplace when they are surrounded by more females. In particular, using our proxy for wages, we find that a 10 percentage point increase in the share of females within a school or neighborhood will reduce the gender occupation-related expected wage gap by 5% and 9%, respectively.

In the second part of the paper, we explore nonlinearities and heterogeneous effects. First, we find that our effects are nonlinear, and that they are larger when the proportion of girls in a school or a neighborhood is over 58%. Girls are almost 1 percentage point more likely to study STEM and 4 percentage points more likely to follow a more lucrative occupation in schools and neighborhoods in which the percentage of girls is above 63% and 61%, respectively. Second, we investigate whether younger or older cohorts affect peers' outcomes, and find no spillover effects for younger or older cohorts at any level. This is important, as it indicates that our results are driven by interactions among students of the same cohorts. Next, we find that the effects on academic outcomes and education choices are similar for small and big schools, but also neighborhoods. This is important, because a common concern in these studies is that small units drive the results. Our findings, therefore, are relevant for schools and neighborhoods of all sizes.

We then perform a number of falsification and robustness exercises with respect to our identification and directly address various threats to its validity. We construct false peer groups in space with which students are unlikely to interact, and find no gender peer effects at the school or the neighborhood level. Additionally, we perform a number of robustness checks that involve the existence (or openings) of single-sex schools in the area, serial correlation within schools or neighborhoods across time, time-varying dropout and repetition rates, and student mobility based on a school's previous quality or gender composition. Overall, our identification assumptions and main results are highly robust to an extensive battery of robustness exercises.

Increasing or decreasing interactions between genders in the classroom may be an effective and low-cost way to change gender composition across university fields. A recent debate relates to how policymakers would increase female representation in STEM degrees and occupations. Governments have allocated significant budgets to programs or policies that aim to reduce inequalities in gender occupations and income (OECD 2017). One way to do this is to encourage equal representation in specific postsecondary degrees. Not all college degrees offer similar economic returns; for instance, it has been established that STEM degrees are associated with higher lifetime earnings in the labor market (Oreopoulos and Petronijevic 2013)⁴. Nevertheless, a considerable gender difference in enrollment in STEM degrees is still observed. Although there is some increase in girls' enrollment rate in STEM-related degrees over the last years, this convergence is still very slow. Manipulating gender interactions might be a way to change these dynamics, which could potentially increase females' representation in STEM-related occupations and reduce the gender wage gap⁵. Understanding these effects is also important and relevant to the persistent debate regarding single-sex versus coeducational environments (i.e., classes or schooling).⁶ Gender social interactions seem to have important implications for school authorities and policy makers who design each school's catchment area and must determine the assignment of students and resources to schools within neighborhoods, but also the future gender composition across university fields.

The question of which peers actually matter for students' learning and decisions has been almost entirely neglected by the peer effects literature. Almost all peer effects studies have focused on schoolmates; in contrast, the group of same-age peers in students' geographical proximity has been overlooked in the peer effects literature. Several papers use unconventional definitions of schoolmates or classmates as a

⁴For example, in 2012 in the US, a STEM degree was estimated to be worth roughly \$1.5 million, which is 100% more than the premium an arts or humanities degree offers, and almost 50% more than to a social science degree's premium (Webber 2014).

⁵For example, in 2011 in the US, women held close to half of all jobs in the US economy, but they hold less than 25% of STEM jobs (Beede et al. 2011)

⁶In London in 2017, single-sex schools, especially girls' schools, notably outperformed coeducational schools. The *Telegraph* (19 January 2017) reports that in a recent evaluation of British schools, girls' schools outperform boys' schools on average, and both outperform coeducational schools <http://www.telegraph.co.uk/education/2017/01/18/level-school-league-tables-2017-compare-schools-performance/>. This leaves still open questions about the effects of increasing interactions between girls and boys within reference groups.

peer group. Although they do not study gender peer effects, these authors use intriguing definitions of peers. For instance, [Foster \(2006\)](#) highlights the importance of measuring the relevant peer group when examining peer effects. She defines a peer group as “all students residing in rooms that are on the same wing of a residence hall as the given student,” and she finds that the geographical proximity of college students alone yields no evidence of contextual peer effects at the University of Maryland. [Carrell et al. \(2009\)](#) use the random assignment of college students to squadrons at the US Air Force Academy, and find positive peer effects when a college student is randomly assigned to groups of around 30 students who spend most of their time working together. Similar to this study, [Sacerdote 2001](#), [Zimmerman 2003](#), [Stinebrickner and Stinebrickner 2006](#) use the random assignment of students to college dormitories at Dartmouth College, Williams College, and Berea College, respectively, and find only moderate college roommate peer effects. In contrast, [Halliday and Kwak \(2012\)](#) study the role of friends and find large peer effects on educational attainments.

Various studies also examine how changes in gender composition in school affect students’ academic performance and educational choices. At the primary school level, a positive effect on both genders’ test scores is found when there is a higher proportion of girls in the classroom ([Lavy and Schlosser 2011](#), [Hoxby 2000](#)). [Lu and Anderson \(2015\)](#) find that if a female student is surrounded by five females in the class rather than five males, this increases her performance by 0.2 to 0.3 s.d., but has no significant effects on a male’s test scores. At the high school level, [Hu \(2015\)](#) finds that a higher share of female peers in the classroom improves boys’ academic performance in China. [Lavy and Schlosser \(2011\)](#), using Israeli data, conclude that a higher proportion of girls at different stages of the schooling cycle positively influences academic scores for both genders. By contrast, [Schone et al. \(2017\)](#) show that having more female peers in the lower secondary grades renders females more likely to choose male-dominated courses in upper secondary grades.⁷ Taken together, these studies have largely focused on the effect of gender composition at the school or class level on students’ academic performance and educational choices within an elementary or high school. However, there is little evidence of gender peer effects in high schools on students’ choices of university study, and in particular STEM degree programs at the university level. One notable exception is [Anelli and Peri \(2017\)](#), who use Italian high school data and show that male students become more likely to choose a male-dominated college major when they are exposed to a classroom with more than 80% of male students in high school⁸. However, no research has investigated the longer-term effects of high school gender composition, and none has examined the short- and longer-term effects of peer environment based on physical proximity outside of the school and classroom level.

⁷At the university level, some studies have used random assignment of students to groups and obtained mixed results. Other studies have found that the college peer environment has important effects on academic performance and/or on students’ choices (e.g., [Carrell et al. 2009](#), [De Giorgi and Woolston 2012](#), [Zolitz and Feld 2017](#), [Booth et al. 2014](#), [Hill 2015](#)).

⁸[Schneeweis and Zweimuller 2012](#) use variation in the gender composition of consecutive cohorts within schools to show that at the age of 12, girls exposed to a higher share of girls in previous years are less likely to choose a traditionally female-dominated school type and more likely to choose a male-dominated school type.

Students who attend different schools but live nearby likely interact often by playing together, taking part in after-school study activities, or extracurricular activities, or attending schools that share facilities. A few papers provide evidence that having more female students in a school or class affects the learning environment—specifically, by lowering levels of classroom disruption and violence—and improves the teacher-student and inter-student relationships; it also increases student satisfaction with school (Lavy and Schlosser 2011, Schone et al. 2017). This mechanism could explain our results at the school level, but also at the neighborhood level if neighbors’ interaction is through localized after-school classes or other activities that same-cohort neighbors are likely to do together. Another possibility is that females might develop specific noncognitive skills (e.g., confidence) when they are surrounded by a higher share of females. These skills are likely to help females perform better in school, invest more human capital in math and science, and also target more male-dominated fields and occupations. Development of these noncognitive skills is found to positively affect females’ scholastic outcomes and decisions (Murphy and Weinhardt 2016), which could also have positive ramifications for their male peers.

The rest of the paper is organized as follows. Section 2 describes the institutional setting and the data. Section 3 explains the identification strategy. Section 4 presents the results. Section 5 discusses the falsification exercises. Section 6 explores additional results, and Section 7 discusses our robustness checks. Section 8 concludes.

2 Institutional Setting, Data and Descriptive Statistics

2.1 Institutional Setting

Students are assigned to public schools based on proximity from their residential address; those who apply to private schools are exempted from this rule. Most students (92%) in Greece attend public schools. The school authority in each locality informs parents which public school their child has been assigned to, and parents must provide proof of their residential address and utility bills to register their child into that school.⁹ Assignment of students to high schools takes place at the beginning of the 10th grade. Students are required to remain at and graduate from the school they are assigned to. Also, students are by law lexicographically assigned to classes for their core/general education courses. In addition to these classes, students are required to choose a track for specialization in the 11th and 12th grades. Usually, students choose a track based on their desired field of study at the university level. There are three options: (1) classics or humanities, (2) science, or (3) exact science. Teachers are also randomly assigned to classes in public schools by law. The 12th grade is the senior grade in high school and the most important for university admission. Students take school-level exams throughout their senior year, and national exams at the end of 12th grade.

⁹If the family pays rent, the school requires proof of rent payments and official documents from the Tax Authority that certify their status.

The transition from high school to postsecondary education in Greece is based on a systematic and transparent allocation of students to university departments. Many countries have a similar university admissions system, such as Chile, China, Korea, Taiwan, and Turkey. Students are compared with each other based on their admission grade¹⁰, which is the only criterion for university admission. After admission grades are announced, students compile a list of ranked choices of specific departments in universities (degrees) that they submit to the Ministry of Education. A computerized central system at the Ministry of Education ranks all students by their admission score and assigns the highest ranked student to her preferred degree choice¹¹. It then moves to the next student and assigns her to the first degree on her list for which there is an available place, and so on. There are two main postsecondary types of institutions: academic universities and technical schools. Academic universities have, on average, higher admission criteria, and thus are considered to be slightly more prestigious. The duration of studies in both types of degrees is four years. On average, the technical schools have a more applied focus.

2.2 Data Description

To study gender peer effects in schools and neighborhoods, we use data for all high schools in Greece for six consecutive 12th grade cohorts of age-17 pupils in academic years from 2003/2004 through 2008/2009. For our empirical analysis, we construct a unique dataset of all students taking the end-of-secondary-education national exams, and we link this information to data on their postsecondary enrolment. We obtain the information from various sources:

1. Administrative data from the Hellenic Ministry of Education that contain information on all 12th grade students and all schools in the country. For each student, we have information on the high school they attended, their midterm and end-of-year national exam scores, and the track/specialization they choose at the beginning of 12th grade. We also have their gender, year of birth, graduation year, and the quarter of the year in which they were born¹². Importantly, we have the same information for their peers in school. We also obtained information about all students' postsecondary education choices from the Hellenic Ministry of Education, which we link to the main dataset described above using a unique identification code. This dataset contains information on students'

¹⁰National end-of-high-school exams are administered by the Ministry of Education. They take place only once every year and last from late May to early June. The exams are graded blindly and externally. Students usually take exams in six subjects, with a combination of common subjects (language, mathematics, physics, biology or history), four compulsory track-specific subjects, and one elective exam. The university matriculation grade is an average of all of the exams a student takes. Track-related subjects are assigned a greater weight in calculating the admission grade, and special weight is assigned to the track subjects most closely related to the university departments a student plans to apply to.

¹¹Details about this admission system can be found in [Goulas and Megalokonomou \(2018\)](#).

¹²Students who are born in the second, third or fourth quarter of their birth year are allowed to enrol in school "early", which means that they enrol before they turn 5 years old. In all our regressions, we control for a dummy that takes the value of one if a student is born in the first quarter of their birth year.

exact university matriculation grade¹³, whether they are enrolled in a university department (the exact university identifier), a degree identifier, and information about whether a student enrolls in (1) a science or mathematics degree or (2) a STEM degree (science, technology, engineering and mathematics). From this dataset, we also obtain information on each degree’s *admission threshold* or *cutoff grade* for each year. A degree’s threshold or cutoff grade is the grade of the last student admitted in that year, which reflects how selectiveness or prestigious a university department is. More selective/prestigious degrees have higher admission thresholds or cutoff grades.

2. Each postsecondary institution is classified as either an academic university or a technical school (vocational type).
3. We obtained Labor Force Survey data for the year 2003 from the National Statistical Authority. We use this quarterly data to map each college major into the most closely related occupation identifier and then into the annual earnings reported in 2003 for each occupation category. Respondents have 209 occupation categories available to them, and they select their occupation with high precision. Then the earnings data are grouped into 10 bins that represent the 10 national deciles with the highest frequency. For each bin, we use the lowest bound to construct a proxy for the minimum expected annual earnings from each occupation. We consider this to be a proxy for students’ expected annual earnings after graduation from each university degree (in Euros). In other words, this is a proxy for how lucrative, on average, each occupation is. We call this “*Expected Occupation Wages*”.
4. We obtained school-specific information, such as the name of the school each student attended and the type (public, private, or evening¹⁴), as well as the geographical coordinates (latitude and longitude) for each high school using Google Maps.
5. The Ministry of Economy and Finance provided us with average net income information for each postcode in the country in 2009 Euros.

We impose some restrictions on the data to obtain a balanced panel of students and schools. Figure 1 maps all high schools, along with information about their average percentage of females across years. We exclude private (27,784 students) and evening (2,741 students) schools from our analysis to avoid the endogenous selection of students into those schools, and thus endogeneity in the proportion of females. For the same reason, we also remove single-sex schools (427 students in the country attend female single-sex

¹³The university matriculation grade is an average of each student’s grades on the national exams they take throughout the year. Subjects are assigned different weights based on the track students chose and the university department they apply to.

¹⁴Evening schools are designed for working students and classes take place in the evening. These schools are also public, and they are free of fees. In order for a student to enroll in an evening school they must be at least 18 years of age and present an employment certificate.

schools and 363 students attend males single-sex schools). We also drop school cohorts or neighborhood cohorts for which the proportion of females was either 0 (150 students) or 100% (656 students). We also remove 299 students whose ages are not reported in the dataset. Finally, we exclude from the main analysis all students who are observed in a high school or neighborhood whose cohort size in a particular year is smaller than 10, because in these small schools an additional female (or male) student might cause a substantial change in the proportion of females¹⁵. This leaves us with a final sample of 324,451 students and 1,097 schools for the period 2004-2009.

For the neighborhood analysis, we drop clusters that have only one school because, according to our definition of neighbors, these students do not have any neighbors. These areas include, in particular, islands or other remote areas in which only one high school operates (38,333 students). This leaves us with a final sample of 283,733 students and 222 neighborhoods for the period 2004-2009.

2.2.1 Construction of Neighborhoods

To study gender peer effects in space, we widen the reference group and examine not only gender peer effects with respect to same-cohort school peers, but also same-cohort peers in the local residential area (excluding school peers). We exploit the fact that schools are built very close to each other, so students who attend neighboring schools also live very close to each other and interact in daily life. We identify all same-cohort peers who attend different schools within a 1-mile radius of their own school. We call this group of same-cohort peers who attend different schools but live very close to each other *neighbors*. In addition to their schoolmates, students are likely to interact with their neighbors and may also be affected by them, since they attend different schools and face different school environments. In each geographical units/clusters, students attend on average four different schools (Panel C, Table 1). We limit our clusters to a 1-mile radius so that they are big enough to allow for school diversity (with four schools, on average, in each cluster), but also compact enough to capture common behavioral patterns or synergies in the local area.

Figure 2 presents a simple hypothetical example that illustrates how neighborhoods are constructed in a dense area in Athens, the capital of Greece, where two schools are located within 1 mile of each other. Students who reside within the pink polygon are assigned to school 1, and all other same-cohort students who attend school 1 are considered to be their schoolmates. Students who reside within the green rectangle are assigned to school 2. For students who attend school 1, their neighbors are considered to be students attending school 2. All other same-cohort students who attend school 2 are considered to be their schoolmates. For students who attend school 2, their neighbors are considered to be students attending school 1. The median distance of a school from each nearest neighboring school is 0.32 miles¹⁶.

¹⁵If we include them the estimations are very similar, only varying at the second decimal point. Contact the corresponding author for further results.

¹⁶Mean of distance from nearest neighbor: 1.85 miles. Standard deviation: 18.37 miles. 25th percentile: 0.07 miles. 75th

We use cluster analysis to define and construct the geographical units of the neighborhoods within 1 mile of each school.¹⁷ We construct 392 clusters that cover the whole country. We remove 142 neighborhoods that contain only one school and 17 clusters that contain only schools whose size is less than 10 students. Our final sample contains 233 neighborhoods. Every cluster is a neighborhood that contains all 12th grade students who attend any other high school within an 1-mile radius of one's high school.¹⁸

Students who attend different schools but live nearby, might play or study together.¹⁹ Figure A.3 reports the average time that 15 year-old students spend each week on after-school study activities using PISA data for 2012. We notice that these after-school classes or study activities are popular in OECD countries (OECD 2013).

2.3 Descriptive Statistics

Table 1 describes our pooled data across cohorts. Panel A presents descriptive statistics at the pupil level. Students are on average 17 years old, and 57% are females. The school midterm score is around 87% and is on average 23 percentage points higher than the national exam score, and almost 81% of students enroll in some postsecondary institution. We also calculate the midterm ordinal rank of students within a school and cluster, which is increasing in midterm test score. The mean ordinal rank of students within a school is around 38, and within a cluster is around 209. The high school exact science track is the most popular (almost 49% of students enroll in this track), while the second most popular high school track is classics or humanities. Only 15% of students enroll in the science track. Panel B refers to school-level variables. All schools in the sample are public, and 81% are located in urban areas. Urban areas are those with more than 20,000 inhabitants.²⁰ On average, 53 students are enrolled in each school. Panel C refers to neighborhood/cluster-level variables. Each geographical cluster contains, on average, four schools and 231 students.

Table 2 displays various demographics for our sample, separately for each year from 2003-2004 to 2008-2009. The last row reports the average of these demographics across all years. In a typical year, the average

percentile: 0.77 miles.

¹⁷We apply a single-linkage hierarchical clustering that calculates the Euclidean distance between every two points and uses the cutoff rule of 0.106 to determine the radius of the clusters.

¹⁸We exploit the fact that many schools were built very close to each other in most urban settings in Greece. To give an example, Figure A.1 depicts the complex of Grava in Athens, where there are six high schools and several elementary and middle schools form a massive complex and share various facilities. Another example is Kaisariani, in Athens, where two high schools share a yard and, in the communal area in front of both schools, a basketball court (Figure A.2).

¹⁹Students might attend after-school classes or study activities (i.e., cramming schools). Cramming schools are popular in many OECD countries (OECD 2013), including Greece, given the intense competition for university admission. These are private businesses that operate locally and prepare students for university admission exams. In some cases, students have even during the weekend take classes and do revision tests. So students take classes and practice tests on weekends, and are very likely to participate in after-school activities or attend a cramming school together in their neighborhood.

²⁰This definition is used by the Hellenic Ministry of Internal Affairs.

proportion of females within a school or neighborhood is 57%. We also observe that in every year, female students on average outperform males in all academic outcomes. In particular, females’ national exam scores, matriculation status, and matriculation scores are higher than those of males. Also, females are on average (63.4-57.8=) 5.6 percentage points more likely than males to enroll in an academic university, rather than a technical school. What is interesting here is that although females on average outperform males in all academic outcomes (columns 7-16), males follow more lucrative occupations, and thus their expected occupation wages (column 17) are higher than those of females (column 18) in each year.

3 Empirical Strategy

Our goal is to estimate the effect of peer gender within a school or a neighborhood on a variety of outcomes, including subsequent score and degree choices. To overcome all endogeneity issues related to sorting of students across schools or neighborhoods, or unobserved correlated factors (Manski 1993), we rely on within-unit (school or neighborhoods) variation. We exploit this within-school or -neighborhood variation in the proportion of females in 12th grade across consecutive cohorts, in order to examine whether there is a systematic association between cohort-to-cohort changes in the proportion of females in schools or neighborhoods and cohort-to-cohort changes in students’ outcomes within the same school or neighborhood.

We use the following equation to estimate gender peer effects for boys and girls separately:

$$Outcomes_{i,u,t+1} = \alpha_u + \beta_u year + \gamma X_i + \delta Prop.Females_{u,t} + \zeta X_{u,t} + \psi_t + \epsilon_{i,u,t} \quad (1)$$

where i denotes student, u denotes geographical unit (school or neighborhood), and t denotes time. $Outcomes_{i,u,t+1}$ is the outcome variable of student i , at school/neighborhood u and time $t+1$. We use the following outcome variables: national exams score across subjects, postsecondary matriculation status (takes the value of 1 if a student enrolls in some postsecondary institution), university matriculation score, enrollment in an academic university (a dummy that takes the value of 1 if a student enrolls in an academic university vs a technical school), quality of postsecondary degree, enrollment in a mathematics or science degree at the university level (takes the value of 1 if a student enrolls in a mathematics or science degree), enrollment in a STEM degree (takes the value of 1 if a student enrolls in a STEM degree), and expected occupation wage. The quality of postsecondary degree expresses how prestigious/selective a degree is, is increasing in quality, and is calculated as a degree’s ranking based on the average (across the sample years) threshold each university department imposes for admission.²¹

α_u is a unit (school or neighborhood) fixed effect, and ψ_t is a year fixed effect. X_i consists of two components: (a) some student level controls that are post-determined, such as dummies for the track chosen at the 12th grade, and the 12th grade midterm ranking across all subjects (“student-level

²¹The quality of post-secondary degree takes values from 0 to 100, with the mean of 50.

controls”) derived at the unit level (school or neighborhood) and (b) a vector that contains student-level predetermined characteristics, including age and a dummy for quarter of birth (“student characteristics”). X_{ut} is a vector of unit (school or neighborhood)-by-year characteristics of unit u at time t , which contains: (a) the 12th grade unit (school or neighborhood) enrollment size and (b) a set of student characteristics averaged at the unit level and year. ϵ_{iut} is the error term, which is composed of a unit-specific random element that allows for any type of correlation within observations of the same unit across time and an individual random element. In all cases, we cluster standard errors at the school level to allow for heteroskedasticity and serial correlation among students within each school. The regression is run separately at each unit level—namely, the school and neighborhood—and all variables are unit specific.

Given that our identification relies on cohort-to-cohort variation, it is important to control for any unobserved unit-specific time-varying heterogeneity that might be correlated with the share of females. Thus we include $\beta_u year$, which is a unit (school or neighborhood)-specific linear time trend. $Prop.Females_{ut}$ is the proportion of 12th grade female students in unit u =school, neighborhood and time t . The school-specific female share excludes the own student, and the neighborhood-specific female share excludes the own school. The coefficient of interest is δ , which captures the effect of having a higher share of females in 12th grade in a school or neighborhood on a student’s educational outcomes and choices. As discussed above, we are interested in separating and comparing the estimates derived at the school and neighborhood level to determine the size of social interactions between genders across distance in space. To examine peer effects on choice of university study, we focus on the two choices of degrees for which we have data: the decision to study mathematics or science²² and the decision to study STEM²³.

3.1 Validity of Identification Strategy

As discussed above, we rely on cohort-to-cohort variation in the share of female students within schools or neighborhoods to estimate gender peer effects in space. For our identification to be valid, two important assumptions must hold. First, the within-unit (school or neighborhood) variation must be considerably large, which is a sufficient condition for precision. Second, the within-unit (school or neighborhood) gender composition must be (conditionally) random, which is the sufficient condition for unbiasedness.

To test the first assumption, we use two approaches. First, we decompose the variation of the proportion of female students in the sample into within-unit variation and between-unit variation. Results are reported in Table 3. We notice that there is a considerable amount of within-school variation in the female share of students (sum of squares: 35.97), which represents 69% of the total variation of gender composition at the school level. Between-school variation (sum of squares: 15.97), on the other hand, is much smaller than within-school variation. Similarly, the between-neighborhood variation is small,

²²Science degrees are all degrees offered by physical and earth sciences, biology, veterinary science, medicine and pharmacy departments.

²³STEM degrees include all degrees in science, technology, engineering and mathematics.

whereas within-neighborhood variation accounts for 63% of the overall variation. To see this visually, we depict the density of the proportion of females within units in Figure 3.

To further support our claim that the within-unit variation is considerably large, we calculate the percentage difference in the share of females within schools and neighborhoods from one cohort to the consecutive one and draw the distributions for each pair of years (Figure 4). We notice that they follow similar patterns, and that most of the variation lies between -20% and +20% difference in the share of female students from one year to the next. Figures 3 and 4 reach the same conclusion: There is significant cohort-to-cohort variation in the proportion of females within units from one year to the next.

To test the second assumption, we use two approaches: First, we deploy balancing tests to examine whether these cohort-to-cohort changes in the share of female students are systematically associated with cohort-to-cohort changes in students' characteristics. To do this, we regress students' characteristics on the share of girls in the school or neighborhood. Columns (1) and (2) in Table 4 present school-level estimates, and columns (3) and (4) present neighborhood-level estimates. The school fixed effects specification in column (1) indicates that the female share at the school level is negatively associated with school enrollment. After adding school-specific linear time trends (column 2), these associations are eliminated. The neighborhood fixed-effects specification in column 3 indicates a high, yet insignificant, association between female share and neighborhood size. Again, once we include neighborhood-specific linear time trends (column 4), the coefficient drops dramatically and remains statistically insignificant. Thus, for the rest of the analysis, we will control for unit (school or neighborhood)-specific fixed effects and we are going to add a unit-specific linear trend to wipe out any observed association between school or neighborhood gender composition and students' characteristics.

Second, we perform a Monte Carlo simulation at the school and neighborhood level to check whether the actual proportion of female students at the school/neighborhood level is consistent with a random process. To do this, we first histogram the actual within-unit standard deviation of the proportion of female students into a probability density plot. Second, for each school and neighborhood in each cohort, we randomly generate the gender of each student using a binomial distribution function with p equal to 0.57—namely, the average proportion of female students in the sample—and then compute the within-school/neighborhood standard deviation of this artificially generated proportion of female students. We use a kernel density to visualize the simulated standard deviation proportion of female students, and superimpose it on the actual probability density histogram. The top (bottom) panel in Figure 5 shows the distribution of the within-school (neighborhood) actual and simulated standard deviation. As this figure shows, the pattern between the actual standard deviation of female share within a school and the simulated standard deviation is similar, which further indicates that the actual within-school gender composition is indeed random. The bottom panel in Figure 5 demonstrates that at the neighborhood level, the distribution of the standard deviation of female share and the simulated randomly generated

standard deviation exhibit a similar pattern.²⁴

Deviations in the share of females from one cohort to another might be generated by a differential fertility gender pattern 17 years earlier. However, the sex ratio at birth, which is defined as male births per female births, for the period from 1986 to 1991 seems to be constant and is equal to 1.067²⁵. However, the pattern we observe in terms of the share of female students cannot be explained by the sex ratio of males and females at birth. In general, any factors that affect the proportion of females in the period from 2004 to 2009 in a similar way should not threaten our identification strategy. Given that our data refer to the 12th grade, one could be sceptical about the existence of a higher female share, which indicates a differential gender attrition pattern in the previous grades. We investigate this empirically in our robustness checks exercise.

4 Results

In this section, we present our main results for gender composition on different students' outcomes in high school and university. In Subsection 4.1 we show how gender interactions affect high school and university admissions-related outcomes. In Subsection 4.2 we look at the effect of gender interactions on choice of university study, and in Subsection 4.3 we consider nonlinearities.

4.1 Main findings on high school and university admission outcomes

Estimated gender peer effects at the school and neighborhood level on a student's subsequent scholastic performance, based on estimating regression (1), are shown in Table 5. We look at gender school peer effects (columns 1-6) and neighborhood (columns 7-12) gender peer effects separately. Each cell in Table 5 presents the estimated coefficient of the female share from a separate regression. In the same table we also report the outcome means (columns 1, 4, 7 and 10) as a benchmark. We present two main specifications. The first, in columns (2), (5), (8), and (11), controls for year fixed effects, unit fixed effects, unit-specific linear trends, and students' background characteristics. In the full specification, in columns (3), (6), (9), and (12), we add controls for students' post-determined characteristics (students' choice of track at the beginning of 12th grade and their mid-term performance ranking), as well as for unit (school or neighborhood)-by-year characteristics (which we explained in detail in Section 3). The estimates of interest follow a similar pattern in both specifications.

At the school level (columns 1-6), all estimates in Table 5 for boys and girls are positive and statistically significant. In particular, when there is a higher proportion of female peers in a high school, both male

²⁴ In Figure 5, when we compute the within-school and -neighborhood standard deviation of the proportion female, we repeat this simulation process 30 times. Further increasing the number of simulations requires an extremely high large programming space.

²⁵World Bank data: <https://data.worldbank.org/indicator/SP.POP.BRTH.MF?locations=GR>

and female students (a) perform better on the subsequent national examination, (b) are more likely to enroll in some postsecondary institution, (c) obtain a higher university matriculation score, (d) are more likely to enroll in an academic university departments (rather than a technical school), and (e) are more likely to enroll in a more selective or prestigious university department. In particular, we find that a 10 percentage point increase in the female share in school increases the national exam score of boys and girls by almost 0.29 and 0.34 percentage points, respectively. Moreover, we find that a similar increase in the female share increases the matriculation likelihood for boys and girls by 0.5 and 0.3 percentage points, respectively. Our findings are slightly smaller than those of [Lavy and Schlosser 2011](#), who find that a 10 percentage point increase in the proportion of females increases the likelihood of matriculation by almost 1 percentage point among girls, and by 0.5 percentage point among boys in Israeli high schools. We also find that a 10 percentage point increase in the female share in school increases the matriculation score by 1.4% and 2.1% of a s.d. for boys and girls respectively. Although at a different education level, our finding here is of the same magnitude as that of [Hoxby 2000](#), who found that a 10 percentage point increase in the proportion of females increases students' mathematics scores by 1-2% of a standard deviation in Texas elementary schools. In Table 5, we also find that a 10 percentage point increase in the female share in school renders boys 0.6 and girls 0.8 percentage points more likely to enroll in an academic university rather than a technical school, and also makes boys and girls more likely to enroll in a university department that is 0.4 and 0.7 percentage points more selective or prestigious, respectively.

At the neighborhood level (columns 7-12) in Table 5, our estimates follow a very similar pattern: Having a higher proportion of female peers in the neighborhood increases both genders' national exam scores, matriculation rates, and matriculation scores and makes both genders more likely to enroll in (a) an academic university (versus a technical school) and (b) a more selective/prestigious university department. All estimated coefficients—except for one that is insignificant—are positive and statistically different from zero for both genders. Focusing on the estimates from the full specifications (columns (9) and (12)), we notice that a 10 percentage point increase in the share of females in the neighborhood increases boys' and girls' national exam performance by 0.30 and 0.35 percentage points, respectively, while the university matriculation score increases by 3.5% of a s.d. for boys and 2% of a s.d. for girls. A 10 percentage point increase in the share of females in the neighborhood also makes boys and girls 2 and 1 percentage points, respectively, more likely to enroll in a university department rather than a technical school.

Focusing on our estimates of gender composition on students' outcomes, both male and female students benefit from a higher share of females in their two potential peer groups (schoolmates and neighbors). This could be explained by the fact that female-heavy environments are found to lead to improved learning conditions. In particular, a high share of females in the peer group is found to lead to an improved learning environment, which is reflected by a lower level of violence and disruption ([Lavy and Schlosser 2011](#)). In female-heavy environments, students report higher levels of well-being and discipline and less

bullying and discrimination (Schone et al. 2017). Thus both genders could benefit by a higher proportion of females in the group. Another interesting pattern in our results is that on average, our effects seem to be higher for girls at the school level and for boys at the neighborhood level. The fact that boys seem to be more affected by gender composition within a neighborhood rather than a school is compatible with the idea that males seem to prefer larger and looser social networks (Lindenlaub and Prummer 2013).

We then repeat this exercise while controlling for the proportion of females in the other geographical unit. In Table A.5 (columns 1-2 for boys and 3-4 for girls), we look at gender peer effects at the school level. Columns 1 and 3 present the main estimates for boys and girls (these estimates come from Table 5), respectively, while in columns 2 and 4 we also control for the proportion of females in the neighborhood. In Table A.5 (columns 5-6 for boys and 7-8 for girls), we examine gender peer effects at the neighborhood level. In columns 5 and 7, we present the main estimates for boys and girls (these estimates come from Table 5), respectively, while in columns 6 and 8 we also control for the proportion of females in the school. Estimated effects change only slightly when we include the proportion of females in the other geographical unit.

4.2 Main findings on postsecondary degree choices

In the previous section, we find that a higher share of females improves both genders' academic outcomes. This might differently impact boys and girls' postsecondary decisions. Previous research suggests that girls could more easily develop noncognitive skills (i.e., self-confidence) in male-dominated subjects when they are surrounded by females (Schneeweis and Zweimuller 2012). This could happen for at least two reasons. First, males—especially in male-dominated subjects—might have a dominating behavior in the class, and girls might be inclined to hold back. This could be driven by the different levels of competitiveness that boys and girls exhibit when surrounded by their peers (Niedelre and Vesterlund 2007, Gneezy et al. 2003). Second, the behavior of practices of teachers, educators, sport coaches etc. might reinforce gender stereotypes in mixed-gender environments (Lavy and Sand 2015, Lavy and Megalokonomou 2017).

In Table 6 we present the estimated effect of changes in gender composition within a school or a neighborhood on the following postsecondary choices: enrollment in a math or science degree (1=yes), enrollment in a STEM degree (1=yes), and the expected wages from pursuing a given occupation.²⁶ These “*Expected Occupation Wages*” indicate how lucrative each occupation is.

We find that in a cohort in which—simply for idiosyncratic reasons—there is a higher percentage of females within a school or neighborhood, there are long-lasting effects on students' career choices, especially for girls. These results demonstrate an interesting pattern: An increase in the share of females at both levels—school and neighborhood—positively and significantly affects the probability that girls will study a traditionally male-dominant college degree—that is, a STEM degree. Estimates for boys do not follow

²⁶These occupations are linked to postsecondary degrees.

the same pattern. Interestingly, we find that only females who have more female peers in a school or a neighborhood are more likely to enroll in a STEM postsecondary degree.

Table 6 presents the estimated effect of the proportion of girls on enrollment rates in these university fields using a similar specification to that used in Table 5. In particular, focusing on columns 5 and 11, we find that enrollment rates for girls in math and science degrees increase by 0.2 and 0.4 percentage points when they are exposed to a 10 percentage point increase in the share of girls in their school or neighborhood, respectively. This is a sizeable effect, because on average only 8% of girls in a school and 6% of girls in a neighborhood enroll in math and science degrees. Our estimates remain almost identical when we add student characteristics and school-/neighborhood-by-year-characteristics as controls (columns 6 for girls at the school level and 12 for girls at the neighborhood level).

Moreover, we find that a 10 percentage point increase in the share of females within a school or a neighborhood increases the proportion of females who enroll in STEM postsecondary degrees by 0.4 and 0.5 percentage points, respectively. Given the low percentage of girls who enroll in STEM, a 10 percentage point increase in the share of females in a school or a neighborhood increases females' probability of enrolling in STEM postsecondary degrees by almost 1.8% $[(0.038/0.217)*10]$ and 2.7% $[(0.048/0.178)*10]$, respectively. Our estimated results imply that increasing the share of females within a school or a neighborhood in a given cohort will compress the gender gap in STEM degrees' enrollment, because the effects are positive and statistically significant only for females, while they are not statistically different from zero for boys. Considering the strong peer effects on females and based on our back-of-the-envelope calculation, a 10 percentage point increase in the proportion of females in a school or a neighborhood reduces the gender gap in STEM enrollments by 2% and 3% at the school and neighborhood levels, respectively.

That is in line with our finding that females choose occupations that pay better in the workplace when they are surrounded by females²⁷. In particular, a 10 percentage point increase in the proportion of females within a school or a neighborhood increases girls' expected wages by 1.3% and 2.1% of a standard deviation, respectively. The fact that we do not find a similar pattern for boys is interesting, because females on average are paid less than males. In particular, girls' average standardized wage is -0.101, while boys' average standardized wage is 0.136. This implies that a higher share of females within a school or a neighborhood will reduce gender wage differences. In particular, based on our back-of-the-envelope calculations, a 10 percentage point increase in the proportion of females in a school or neighborhood reduces the occupation-related expected wage gap by 5% and 9%, respectively.

²⁷When the outcome variable is the expected occupation wages and to increase precision, we also control for a dummy that indicates whether a student graduated from a university department (versus a technical school), as well as a variable that indicates how male-dominated this degree is. The latter is calculated based on the proportion of male students who enroll in each degree every year. If a degree is male-dominated, this variable becomes closer to 1. Our results also hold when we do not control for those variables.

Students who attend different schools but live nearby might often interact, as described previously. Several papers provide evidence that having more female students in the school or class affects the learning environment, namely by lowering the level of classroom disruption and violence, and improves teacher-student and inter-student relationships and caused students to report greater satisfaction with school (Lavy and Schlosser 2011, Schone et al. 2017). This mechanism could explain our results at the school level, but also at the neighborhood level if neighbors' interaction is through localized after-school classes or other learning activities that same-cohort neighbors are likely to do together.

Another possible channel is that females might develop specific noncognitive skills (e.g., confidence), when they are surrounded by a higher share of females. Development of these noncognitive skills is found to positively affect scholastic outcomes and choices (Murphy and Weinhardt 2016). These skills are likely to help females perform better in school, but also target more male-dominated fields and occupations. There are at least two possible reasons why this might happen. First, males tend to dominate discussions and females are inclined to hold back (Karpowitz et al. 2012)²⁸. If females are surrounded by a higher proportion of females, they are more likely to not only adopt a more collaborative and discussion-based approach, but also develop their leadership skills. A higher share of females might also have positive ramifications for their male peers, because males are likely to be less disruptive, more focused, and better behaved in peer environments in which they are outnumbered by girls, which is in line with the findings of Lavy and Schlosser (2011) and Figlio (2007). Second, the behavior and practices of teachers, educators, sport managers etc. in mixed environments might reinforce gender stereotypes and enlarge the existing gender performance gaps (Lavy and Sand 2015, Lavy and Megalokonomou 2017). For example, teachers might favor boys more in mathematics courses in male-heavy classes, and thus reduce girls' probability of targeting male-dominated fields and occupations.

4.3 Nonlinear Effects

Our specifications thus far have assumed that the effects are linear, but it is possible that the effects are higher when the share of females is larger. To address this, we split our variable of interest at the school or neighborhood level into five quintiles and we replace this variable in the regression with a set of quintile dummies. The first quintile dummy (the one for the lowest proportion of females) is our reference group, and is thus omitted from the regressions. Schools and neighborhoods are observed in multiple years, and the proportion of girls in schools/neighborhoods could switch across quintiles frequently throughout all years. In a within-school/-neighborhood regression model, this dynamic transition across quintiles within sample years allows us variation for identifying non-linear gender peer effects. We first show that there is enough transition of schools and neighborhoods from one quintile to another within the sample period

²⁸Using experimental evidence, they find that men dominate 75% of the conversation during conference meetings, and even when women have something unique and important to add to the group, that is being lost.

(Table A.3, Panels B and D).²⁹ This approach was also followed by [Lavy and Schlosser \(2011\)](#). Table A.3 presents some descriptive statistics for each quintile of the proportion of females in the school (Panel A) and neighborhood (Panel C).

Tables 7 and 8 present the estimated gender peer effects from switching to an environment with only a very low proportion of females (quintile 1) to an environment in which the proportion of females is increasingly higher (quintiles 2, 3, 4, and 5). The results largely indicate a positive relationship between academic outcomes and percentage of females in the peer group, with students performing better in cohorts with a high share of females. Table 7 presents the school-level results and shows that most of the effects appear to increase with the quintiles. In particular, the effects become more evident after the fourth and fifth quintiles, in which the proportion of female students exceeds 59% or 63%, respectively. For example, in school cohorts with a female share above 0.632 (quintile 5), a 10 percentage point increase in the share of female schoolmates causes boys to experience an increase in their national exam score by 0.7 percentage points, and a higher matriculation score by 0.4% of a standard deviation, and renders them 0.2 percentage points more likely to enroll in an academic university (rather than a technical school), compared to boys in the first quintile. Similarly, in cohorts with a female share above 0.632 (quintile 5), a 10 percentage point increase in the share of female schoolmates increases girls' national exam score by 0.7 percentage points and girls' matriculation score by 0.4% of a standard deviation, and renders girls 0.4 percentage points more likely to enroll in some university and 0.2 percentage points more likely to enroll in an academic university, relative to cohorts with a female share below 0.509 (quintile 1). Table 8 presents estimates of the nonlinear effects at the neighborhood level. The table shows a similar pattern as Table 7—namely, switching to the fourth (proportion: 0.581-0.607) or fifth quintile (proportion: 0.608-0.917) compared to the first (proportion: 0.200-0.532) is beneficial for both boys and girls.

5 Falsification Exercise

Our falsification exercise demonstrates that students' geographical proximity alone does not generate positive gender peer effects. It also establishes that identifying the relevant peer group is important.

²⁹The cells in the diagonal of the matrix show the number of schools/neighborhoods whose proportion of girls remain in the same quintile across the sample period. This is the case for only a few schools and neighborhoods throughout the years: 86 out of 1,097 schools, and 7 out of 233 neighborhoods. The cells in the off-diagonal of the matrix display the number of schools/neighborhoods that are observed in two different quintiles during the six sampling years. For example, the first row of the school panel shows that 413 schools moved from quintile 1 to quintile 2; 350 schools from quintile 1 to quintile 3; 359 schools from quintile 1 to quintile 4; and 428 schools from quintile 1 to quintile 5. From this pattern of transitions across quintiles, we conclude that there is a considerable amount of within-school and within-neighborhood switching across quintiles. For schools, 5,485 potential transitions could occur, and 3,688 switches actually occurred. For neighborhoods, 1,165 potential transitions could occur, and 750 actually occurred. In addition, the probability of transition between any two quintiles is relatively comparable at both school and neighborhood levels, suggesting that this quintile transition is close to a random process.

In this exercise, we replace our gender composition variable within a school and a neighborhood with a *false gender composition* and check whether the false gender ratio can generate results similar to those of our main analysis. To test this, we construct two false peer groups: (1) all same-cohort students in one’s postcode, excluding one’s own school³⁰ and (2) all same-cohort students in one’s periphery, excluding one’s own neighborhood.³¹ Within each of those false peer groups we construct the annual gender composition and replace our main variable with this.

Table 9 presents results for this analysis. Columns 1 and 2 show results for the first false peer group and use the proportion of same-cohort females in all other schools in one’s postcode, except for their own school. Columns 3 and 4 present results for the second false peer group, which uses the proportion of same-cohort females in all other neighborhoods in one’s periphery, except for their own neighborhood. Similar to the main analysis, we present results for boys and girls separately. In all four specifications, out of the 32 estimates only one estimate is statistically significant for boys at the 10% level and is negative. No other academic peer effect variable has a statistically significant effect on students’ academic outcomes or decisions, indicating that geographical proximity alone does not generate the pattern we observe in our main analysis.

6 Additional Results

In this section, we investigate whether gender peer effects vary by school and neighborhood size, and also whether older or younger peers affect students’ performance and decisions.

6.1 Heterogeneity by School and Neighborhood Size

First, to account for the large changes in the share of females that might occur in small schools and neighborhoods due to a small change in gender composition, we have removed from the main analysis all schools and neighborhoods that have an enrollment smaller than 10 students.³²

Table 10 presents the effects of variations in gender composition on scholastic outcomes and education decisions for samples stratified by school and neighborhood size: below average (below average school-/neighborhood-cohort size of 53/232 students, respectively) and above average (above average school-/neighborhood-cohort size of 53/232 students, respectively). For high school-related outcome variables, the effects of gender composition are similar for samples of both small and large schools and neighborhoods.

³⁰There are 661 postcodes included in our analysis. In Greece, postcodes are not very narrowly defined. Some postcodes are even nongeographic, in the sense that they relate to post office boxes where mail is retained for collection rather than to actual locations. Also, for each island there is usually only one postcode, although schools are located in different parts of the island. In our data, the smallest postcode contains 2 schools, while the largest contains 10 schools.

³¹There are 13 peripheries in total.

³²Estimations when we include those very small schools and neighborhoods are very similar only varying at the second decimal point. Contact the corresponding author for further results.

In particular, the effects coming from large schools and neighborhoods seem to be larger than those coming from smaller ones. For the remaining outcomes (university enrollment decisions and expected wage), estimates are less precise and seem to be less powerful in large schools and neighborhoods. Overall, there is a loss in precision in both subsamples, as expected.

Therefore, we conclude that our estimated effects are relevant overall for units (schools and neighborhoods) of all sizes, and not only relevant for small units.

6.2 Are students affected by older or younger cohorts?

Next, we examine the existence of gender peer effects generated by the previous and following cohorts. We examine whether students are affected by changes in the proportion of female students in the younger (in $t-1$), or the older cohort ($t+1$). To study this, we replace the actual measure of treatment (proportion of female students at time t) within a school or a neighborhood with the proportion of females within the same school or neighborhood in the previous year ($t-1$) or in the following year ($t+1$).

Results are presented in Table 11. In columns 1-2 and 5-6, we report estimates of the proportion of females in cohort $t-1$ in a school and a neighborhood, respectively. In columns 3-4 and 7-8 we report the related coefficient of the proportion of females for the cohort $t+1$ in a school and neighborhood, respectively. We report the estimates for male and female students separately. Almost all of the estimated coefficients are statistically insignificant, and are all quantitatively much smaller than the corresponding coefficients in the main analysis. Only a very few estimates appear to be negative and significant, but they do not indicate any persistent pattern. This is not very surprising for older cohorts, as they have already graduated from high school when our students are in the 12th grade. These results imply that our gender peer effects operate mainly at the grade level and that boys and girls are not affected by a higher share of females in the younger or older cohort.

7 Robustness Checks

In this section, we present a set of robustness exercises that support the causal interpretation of our findings. Some nontrivial empirical challenges arise, when estimating gender peer effects at the school and neighborhood levels, because of the potential existence or openings of single-sex schools in the area, serial correlation within schools or neighborhoods across years, time-varying dropout and repetition rates, and the mobility of students across schools that might affect a specific cohort but not others.

7.1 Single-sex and/or private schools in the neighborhood

In some areas, single-sex or coeducational private schools operate, and students with specific characteristics might select to attend those schools. For example, if a single-sex (or a private) school for boys (or

girls) is in a close proximity to a public coeducational school, then boys (or girls) with certain characteristics are likely to choose to attend this single-sex (or private) school. That would affect the gender ratio and result in a smaller share of boys attending coeducational schools in that specific neighborhood. Although there are only a few single-sex schools in Greece, their existence would not invalidate our strategy if these schools operate for the full sample period and thus affect the cohort-to-cohort calculation of the proportion of females in a similar way. A more credible concern for our setting is if those single-sex or private schools begin or cease to operate in some year during our sample period, as this would disproportionately affect our cohort-to-cohort variation in the proportion of female students in the 12th grade. If the composition of schools in a neighborhood changes over time, that could disproportionately affect the gender composition in some schools and years.

To capture this potential threat, we perform two robustness exercises. First, we control (in all specifications) for school- or neighborhood-specific linear time trends to control for any unobserved factors that might confound school or neighborhood gender peer effects. Second, we include controls for the cohort-to-cohort enrollment of boys and girls in these single-sex or private schools in the area. In Table 12 columns (1), (3), (5), and (7) we present the main estimated coefficients (that come from Table 5), and in columns (2), (4), (6), and (8) we present the estimated coefficient when we control for the cohort-to-cohort enrollment of males and females in single-sex or private schools in the neighborhood. Estimates with and without these controls are very similar, and vary only at the second decimal point.³³

7.2 Controlling for the lag and lead proportion of girls

We then examine whether there is serial correlation in the proportion of female students within schools or neighborhoods from one year to another. Our results in Table 11 imply that gender peer effects operate among peers within the same cohort rather than across consecutive cohorts, so we do not expect to find a significant effect from the proportion of females in other cohorts. In Tables A.1 and A.2 we present the simple correlation coefficients between the 12th grade female share within schools or neighborhoods, respectively, in year t and $t+1$ for every possible combination of years in our sample. This correlation is always below 0.20, indicating that there is no evidence of strong serial correlation across time within schools or neighborhoods. This is in line with our balancing tests in Table 4, which presents evidence that the variation we notice in the proportion of females within schools and neighborhoods seems to be uncorrelated with any observed characteristics and seems to occur for purely idiosyncratic reasons.

In the absence of serial correlation in the gender composition within the same unit (school or neighborhood), we do not expect our results to be driven by future or past values in the share of females. To further support our findings, we present evidence in Table 13, in which we focus again on the estimates of the female share in cohort t , while we control in the same regression for the proportion of females in

³³ We also re-run our main regressions while we exclude the areas in which single-sex schools (or private schools) operate. Estimations remain very similar. We do not report these for space reasons.

cohort $t-1$, cohort t and cohort $t+1$. We present the coefficients of the lagged, current, and the lead value of the proportion of females within a school (columns 1-6) and neighborhood (columns 7-12). The main gender peer effect estimates remain positive and significant in almost all cases, while the estimated effects of future and lagged values are only occasionally positive and rarely significant. Overall, controlling for the future and lagged value of the proportion of females within a school or a neighborhood does not seem to affect our actual treatment variable much, as our results show almost the same pattern as before.

7.3 Dropout and repetition rates

Only variables that follow a differential trend from one year to the next and affect our variable of interest might confound our identification strategy. One concern would be if for some unobserved reasons school dropout or grade-repetition trends follow a different trend from one year to the next. That could directly or indirectly affect our gender composition variable. Our main variable of interest (namely, the share of females) might be affected by either (1) a direct change in the number of girls in a school or neighborhood, or (2) an indirect change in the school or neighborhood enrollment size initiated by a change in the number of boys in a school or a neighborhood.

A gender differential dropout or grade-repetition rate pattern that affects all sample years in a similar way would not invalidate our identification strategy. Boys are more prone to drop out from school and repeat a grade more often. This affects our female share variable in a school or neighborhood indirectly, but this would only invalidate our identification strategy if it affects treated and control years in a different manner; there is no institutional reason why one might expect a time-specific differential trend. However, to further assess the robustness of our findings against the possibility of differential dropout and repetition trends, we use a smaller sample of schools for which we have annual data about dropout and repetition rates for the whole sample period. We provide evidence that this sample of 144 is a random sample in terms of the share of females in Table A.4.

Using this sample of 144 schools, we plot the year-to-year dropout rates from 10th to 11th grade (Figure 6) and from 11th to 12th grade (Figure 7), the year-to-year repetition rates, and the year-to-year dropout and repetition rates by gender. In Figures 6 and 7 we notice that the dropout rates between 10th -11th and 11th -12th grades (Panel A in these two figures), as well as the proportion of females within a school (incorporated in the same figure) follow a very flat trend for our sample period, without any unexpected peaks. The same holds for the 10th-11th and 11th -12th dropout rates for boys and girls (Panel B). We observe a similar pattern when we look at the percentage of repeaters from one year to the next from 2003 to 2009 (Panel C) and the percentage of repeaters in 11th and 10th grades for boys and girls, separately (Panel D). From this evidence we conclude that our findings are unlikely to be affected by differential dropout and repetition rates.

7.4 Mobility across schools

One could still be concerned that students might respond to changes in cohort gender composition or school quality, and thus switch from one school to another. To lend additional credibility to our causal interpretation of the results, we check whether changes in the proportion of females or changes in a school’s quality are associated with changes in a school’s enrollment.

The institutional setting here does not leave a lot of space for this kind of behavior. First, students are not allowed to enroll in a school based on their preference. Second, schools’ quality measures or other statistics (such as gender composition) are not publicly available. Thus, parents or students are unlikely to be aware of how a school ranks relative to others or its gender composition. It would be even more difficult for them to know in advance the gender composition of a cohort that enters the school in the next year. Third, unlike other countries, it is not the norm in Greece for families to move to another neighborhood because one school performs better than another.

To empirically check these concerns, we examine whether a future school’s enrollment is likely to change based on last year’s school quality or gender composition. In particular, we run the following regressions:

$$\begin{aligned}
 Enrollment_{s,t+1} &= \alpha_1 + \lambda_s + \kappa_t + \gamma_s year + \delta_1 SchoolRank_{s,t} + \theta_1 Enrollment_{s,t} \\
 &+ NeighborhoodCharacteristics_t + u_{s,t} \quad (2) \\
 Enrollment_{s,t+1} &= \alpha_2 + \zeta_s + \beta_t + \gamma_s year + \delta_2 ProportionFemales_{s,t} + \theta_2 Enrollment_{s,t} \\
 &+ NeighborhoodCharacteristics_t + e_{s,t} \quad (3)
 \end{aligned}$$

where s denotes schools and t denotes time. $Enrollment_{s,t+1}$ is the outcome variable and denotes the enrollment size of school s in time $t+1$. The coefficients of interest are δ_1 and δ_2 , and they examine whether a schools’ future enrolment is affected by that school’s last-year ranking or a school’s last-year proportion of females, respectively, *ceteris paribus*. To construct a measure for school’s quality, we rank schools based on their average performance on the national exams across all nationally examinable subjects³⁴. If students respond to changes in school rankings or gender composition, we would expect δ_1 and δ_2 to be statistically different than zero.

Table 14 reports the estimates of the effects of a school’s quality (columns 1 and 2) and proportion of females in year t (columns 3 and 4) on its future enrollment. Columns 2 and 4 include school fixed effects (ζ_s), year fixed effects (β_t), a school linear time trend ($\gamma_s year$), controls for prior and current enrollment size, and a neighborhood’s characteristics in year t . All estimates appear to be statistically insignificant, regardless of whether we control for year and school fixed effects. Overall, these results suggest that changes in a school’s enrollment are uncorrelated to the school’s quality or proportion of females.

³⁴The school-ranking variable takes values from 0 to 100 and it is increasing in performance, where the highest performing school has a ranking equal to 100 and the lowest performing school gets a ranking equal to 0.

8 Conclusion

In this paper, we study nonconventional peer effects and examine whether social networks broader than schoolmates matter for students' learning and post-secondary choices. A common thread in the literature is the inclusion of irrelevant peers. We believe that an obvious peer group for high school students is their same-cohort peers who reside near them, but attend different schools in the local area. High school students, unlike college students, are very likely to interact with their physical proximity peers, as it is likely that there have been many occasions in their lives as adolescents that they have participated in neighborhood group-specific activities (i.e., play sports together, participate in after-school learning activities, interact at public spaces in the locality, participate in community activities etc.). Thus, we separate and compare the social interaction effects between genders that come from schoolmates and neighbors. We examine the effects of these gender interactions on high school academic performance, postsecondary admission outcomes, postsecondary degree choices and expected occupation earnings.

We use unique administrative data for the universe of students in Greece from 2004 to 2009 and geographical data to open the black box of relevant peers and gender peer effects. The data allow us to identify all same-cohort students who attend any other school within 1 mile of one's school, which we define as one's neighbors. We exploit an institutional setting in which (1) students are assigned to schools based on proximity from their residential address and (2) schools are built very close to each other. Thus, students are likely to live close to each other, but attend different schools. We exploit cohort-to-cohort variation in the gender composition of 12th-grade students within schools and neighborhoods to deal with the usual sorting and endogeneity problems. We compare the outcomes and choices of students from consecutive cohorts that have similar characteristics and face the same environment, except for the fact that one cohort has a higher female share than the other.

The evidence we provide in this paper suggests that a higher female share in students' environment increases both genders' scholastic end-of-high-school outcomes, as well as their university admission outcomes. We also find that gender composition affects girls' choices of university major. In particular, we find that female students in cohorts that have more girls in their school or neighborhood are more likely to enroll in STEM degrees at the university level and pursue a more lucrative occupation. A 10 percentage point increase in the share of females within a school or neighborhood increases the share of females who enroll in STEM postsecondary degrees by 0.4 and 0.5 percentage points, respectively, while males' decision to enroll in STEM is not significantly affected by the share of girls in the school or the neighborhood. Focusing on the strong effects we find on females, and in an attempt to put our effects into perspective, we find that a 10 percentage point increase in the share of females within a school or a neighborhood reduces the gender gap in STEM enrollment by 2% and 3%, respectively. Then, we explore the nonlinearities in our effects, and our effects seem to be larger for higher proportions of girls in a student's environment—specifically, above 58%. We find that our results overall are relevant both small and

large schools and neighborhoods. Moreover, we find no spillover effects from younger or senior cohorts, indicating that these gender peer effects mainly operate through same-cohort peers.

We believe that this paper contributes significantly to our understanding of which peers are relevant for a high school student. Our findings emphasize the importance of social networks that are broader than schoolmates alone. They also highlight the fact that these social networks directly affect not only students' high school and university admission outcomes, but also their decisions to specialize in STEM fields and occupations. Our findings are relevant for assessing the consequences of imbalances in gender composition in coeducational environments and for determining an optimal allocation of students and resources in schools and neighborhoods. Moreover, this study has important implications for school authorities and policy makers who are designing the schools' local catchment areas and who must determine the assignment of students based on geographical areas and residential addressees to schools within their neighborhoods. School authorities can well predict the gender share in high schools based on present and past trends in gender share in the elementary schools in the neighborhood. Our findings suggest that communal areas or other neighborhood-specific initiatives between schools with very different gender compositions might be beneficial for students. Our research provides evidence that manipulating the gender composition of a given environment may have direct consequences for future gender imbalances in college majors and labor market occupations.

References

- Akerlof, G. and R. Kranton (2002). Identity and Schooling: Some Lessons for the Economics of Education. *Journal of Economic Literature*, 40(4), 1167–1201.
- Anelli, M. and G. Peri (2017). The Effects of High School Peers' Gender on College Major College Performance and Income. *The Economic Journal*, Forthcoming.
- Beede, D., T. Julian, D. Langdon, G. McKittrick, B. Khan, and M. E. Doms (2011). Women in STEM: A Gender Gap to Innovation.
- Bikhchandani, S., D. Hirschleifer, and I. Welch (1992). A Theory of Fads, Fashion, Custom, Cultural Changes as Informational Cascades. *Journal of Political Economy*, 100(4), 992–1026.
- Bobonis, G. J. and F. Finan (2009). Neighborhood Peer Effects in Secondary School Enrollment Decisions. *The Review of Economics and Statistics*, 91(4), 695–716.
- Booth, A., L. Sosa-Cardona, and P. Nolen (2014). Gender Differences in Risk Aversion: Do Single-Sex Environments Affect their Development. *Journal of Economic Behavior and Organizations*, 99, 126–154.
- Buu, A., C. DiPiazza, L. Puttler, H. Fitzgerald, and R. Zucker (2009). Parent, Family, and Neighborhood Effects on the Development of Child Substance Use and Other Psychopathology From Preschool to the Start of Adulthood. *Journal of Studies on Alcohol and Drugs*, 70(4), 489–498.
- Carrell, S., R. Fullerton, and J. West (2009). Does your Cohort Matter? Measuring Peer Effects in College Achievement. *Journal of Labor Economics*, 27(3), 439–464.
- De Giorgi, Pellizzari, M. and W. Woolston (2012). Class Size and Class Heterogeneity. *Journal of the European Economic Association*, 10(4), 795–830.
- Figlio, D. (2007). Boys Named Sue: Disruptive Children and Their Peers. *Education Finance and Policy* 2(4), 376–394.
- Foster, G. (2006). Its not your Peers, and it's not your Friends: Some Progress Toward Understanding the Educational Peer Effect Mechanism. *Journal of Public Economics*, 90(8-9), 1455–1475.
- Gneezy, U., M. Niederle, and A. Rustichini (2003). Performance in competitive environments: gender differences. *Quarterly Journal of Economics*, 118, 1049–1074.
- Goulas, S. and R. Megalokonomou (2018). Which degrees do students prefer during recessions? *Empirical Economics*, 1–33.

- Halliday, T. and S. Kwak (2012). What is a Peer? The Role of Network Definitions in Estimation of Peer Effects. *Applied Economics*, 44(3), 289–302.
- Hill, A. J. (2015). The Girl Next Door: The Effect of Opposite Gender Friends on High School Achievement. *American Economic Journal: Applied Economics*, 7(3), 147–177.
- Hoxby, C. (2000). Peer Effects in the Classroom: Learning from Gender and Race Variation. *NBER Working Paper No. 7867*.
- Hu, F. (2015). Do Girl Peers Improve your Academic Performance? *Economic Letters*, 137, 54–58.
- Karpowitz, C., T. Mendelberg, and L. Shaker (2012). Gender Inequality in Deliberative Participation. *American Political Science Review*, 106(1), 41–65.
- Lalive, R. and M. A. Cattaneo (2009). Social Interactions and Schooling Decisions. *The Review of Economics and Statistics*, 91(3), 457–477.
- Lavy, V. and R. Megalokonomou (2017). Persistency in teachers’ grading biases and effect on longer term outcomes: University admission exams and choice of field of study. *Working Paper*.
- Lavy, V. and E. Sand (2015). On the origins of gender human capital gaps: Short and long term consequences of teachers? Stereotypical Biases. *NBER Working Paper (w20909)*.
- Lavy, V. and A. Schlosser (2011). Mechanisms and Impacts of Gender Peer Effects at School. *American Economic Journal: Applied Economics*, 3(2), 1–33.
- Lindenlaub, I. and A. Prummer (2013). More versus Closer Friends: How Gender Shapes Social Networks and their Effects on Performance. *Mimeo*.
- Lu, F. and M. Anderson (2015). Peer Effects in Microenvironments: The Benefits of Homogeneous Classroom Groups. *Journal of Labor Economics*, 33(1), 91–122.
- Manski, C. (1993). Identification of Endogenous Social Effects: The Reflection Problem. *The Review of Economic Studies*, 60(3), 531–542.
- Mota, N., E. Patacchini, and S. Rosenthal (2016). Neighborhood Effects, Peer Classification, and the Decision of Women to Work. *IZA Working Paper, No. 9985*.
- Murphy, R. and F. Weinhardt (2016). Top of the class: The Importance of Rank Position. *CEP Discussion Paper No. 1241*.
- Niedelre, M. and L. Vesterlund (2007). Do women shy away from competition? Do men compete too much? *Quarterly Journal of Economics*, 112(3), 1067–1101.

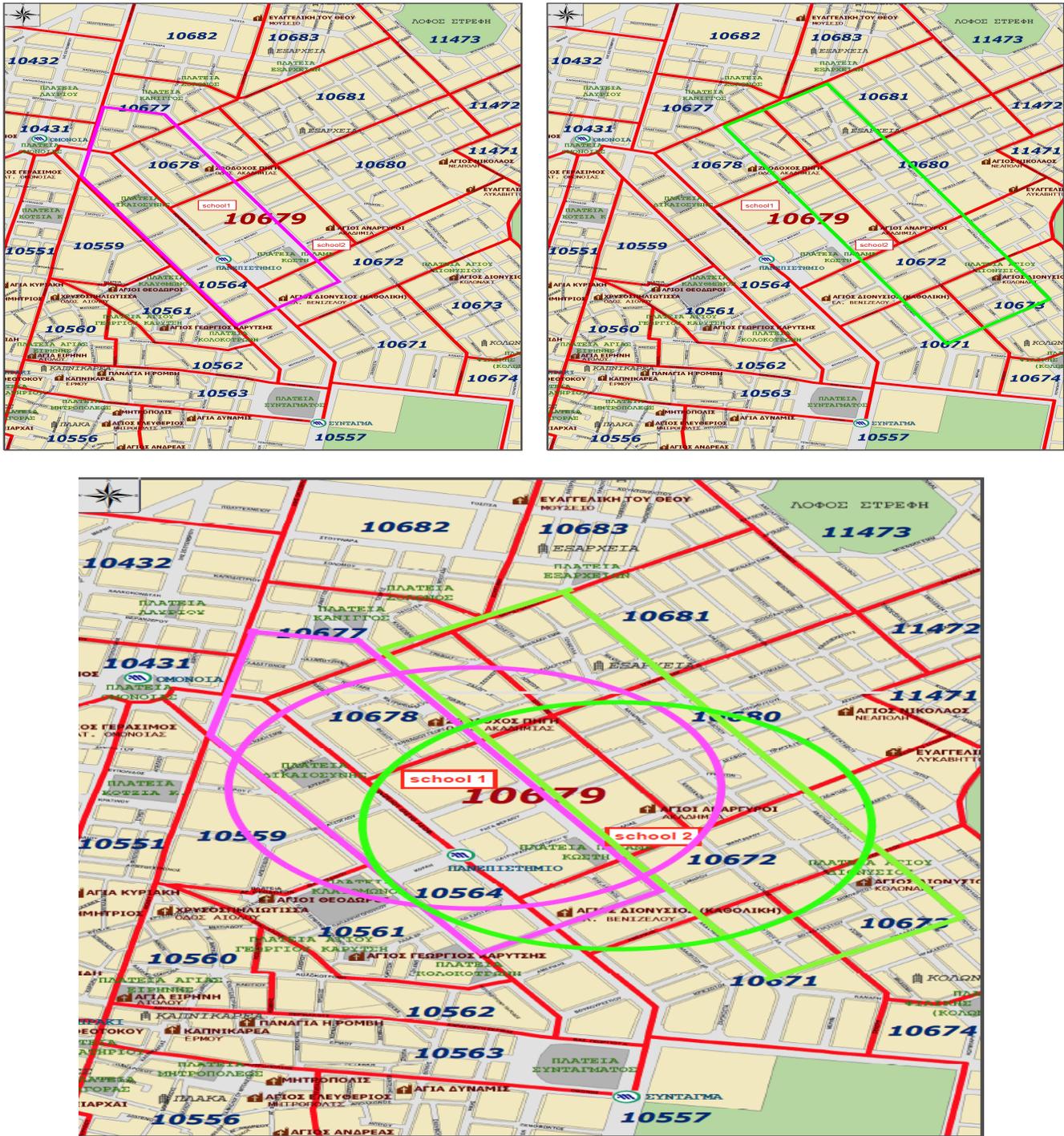
- OECD (2010). *Highlights from Education at a Glance 2010*.
- OECD (2013). PISA 2012 Results: What Makes Schools Successful: Resources, Policies and Practices.
- OECD (2017). Report on the Implementation of the OECD Gender Recommendations. *OECD Publishing, Paris*.
- Oreopoulos, P. and U. Petronijevic (2013). Making College Worth it: A Review of the Returns to Higher Education. *The Future of Children*, 23(1), 41–65.
- Sacerdote, B. L. (2001). Peer Effects with Random Assignment: Results for Dartmouth Roommates. *Quarterly Journal of Economics*, 116(2), 681–704.
- Schneeweis, N. and M. Zweimuller (2012). Girls, Girls, Girls: Gender Composition and Female School Choice. *Economics of Education Review*, 31(4), 482–500.
- Schone, P., K. von Simson, and M. Strom (2017). Girls Helping Girls: The Impact of Female Peers on Grades and Educational Choices. *IZA Working Paper, No. 10586*.
- Stinebrickner, R. and T. Stinebrickner (2006). What can be Learned about Peer Effects Using College Roommates? Evidence from New Survey Data and Students from Disadvantaged Backgrounds. *Journal of Public Economics*, 90(8-9), 1435–54.
- Webber, D. (2014). Is the Return to Education the Same for Everybody? *IZA World of Labor*, 92, 1–10.
- Xiong, H., D. Payne, and S. Kinsella (2016). Peer Effects in the Diffusion of Innovations: Theory and Simulation. *Journal of Behavioral and Experimental Economics*, 63, 1–13.
- Zimmerman, D. (2003). Peer Effects in Higher Education: Evidence from a Natural Experiment. *Review of Economics and Statistics*, 85(1), 9–23.
- Zolitz, U. and J. Feld (2017). The Effect of Peer Gender on Major Choice. *University of Zurich, Department of Economics, Working Paper No. 270*.

Figure 1:
MAP OF SCHOOLS

Prop. of Females per School



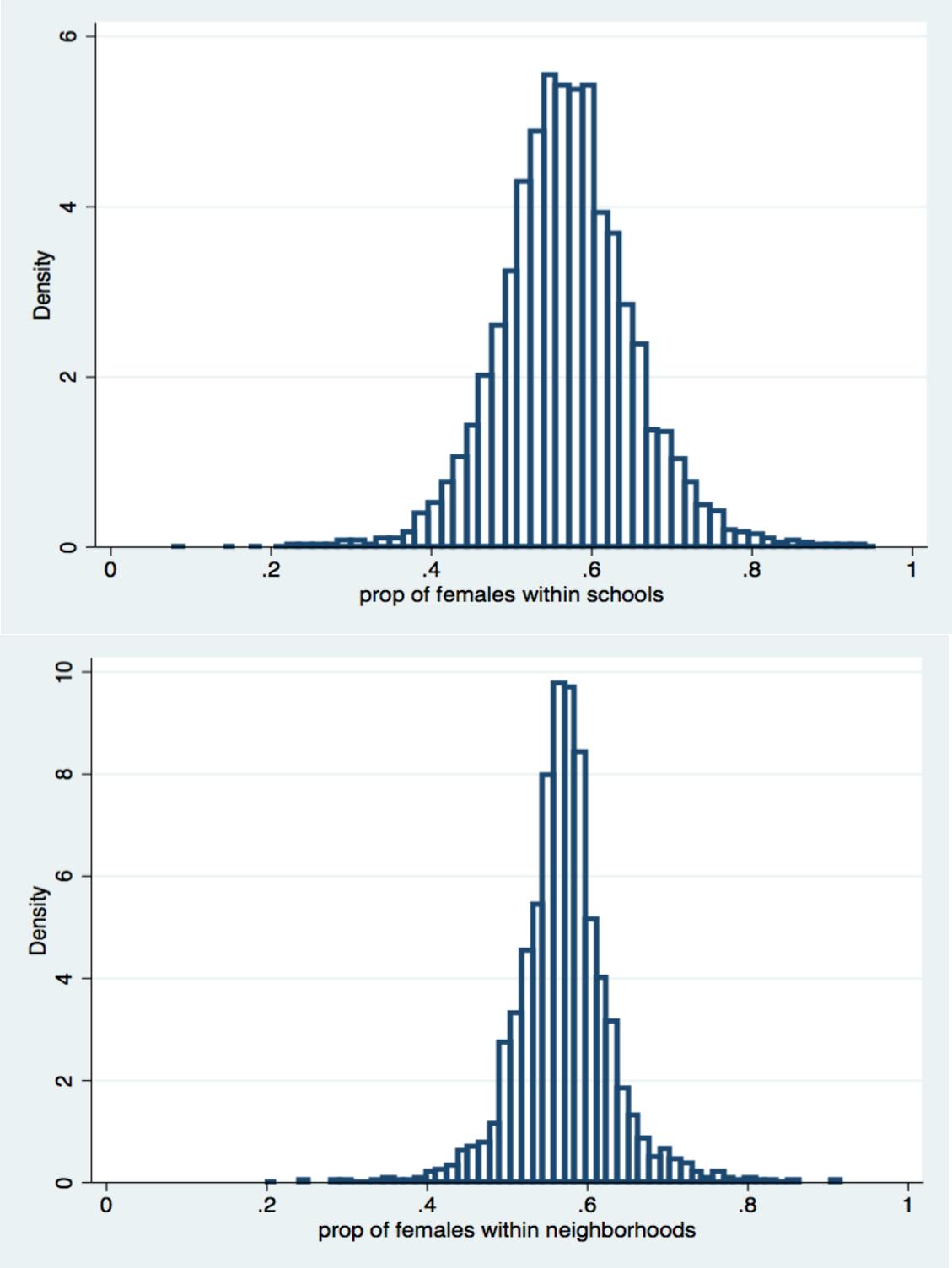
Figure 2:
CONSTRUCTION OF NEIGHBORHOODS



Note: Each panel shows an area in the center of Athens. Students are assigned to schools based on geographical proximity from their residential address. The top left panel shows those residential addresses -within the pink polygon- that are assigned to school 1. The top right panel shows all those residential addresses -within the green rectangle- that are assigned to school 2. To identify neighbors we draw circles of 1 mile radius around all schools. The neighbors for students who attend school 1 are those that attend school 2. The neighbors for students who attend school 2 are those that attend school 1.

Figure 3:

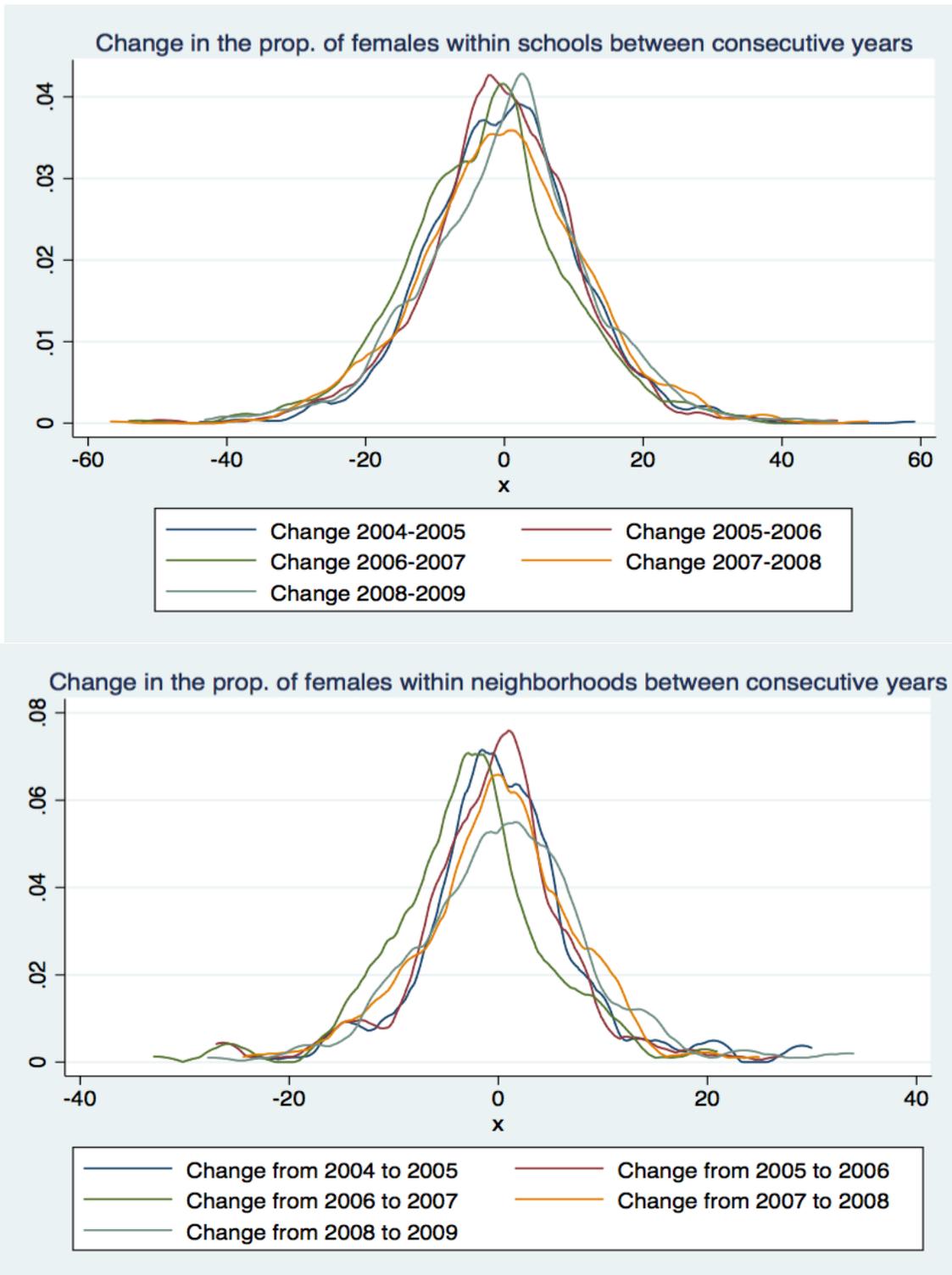
HISTOGRAM OF THE PROPORTION OF FEMALES WITHIN SCHOOLS AND NEIGHBORHOODS



Note: The first histogram presents the distribution of the proportion of female students within schools, while the second histogram presents the distribution of the proportion of female students within neighborhoods.

Figure 4:

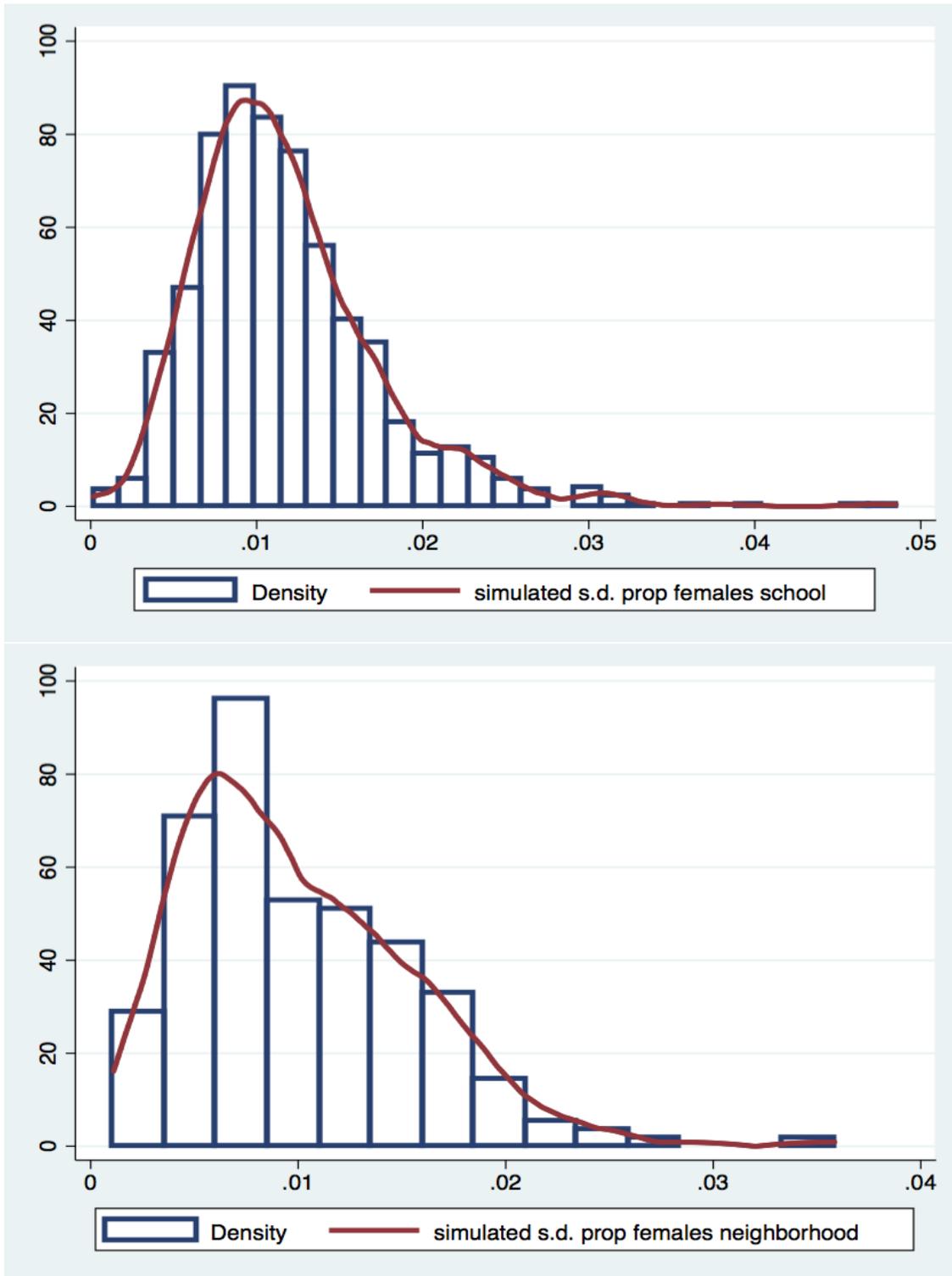
DENSITIES FOR CHANGES IN THE PROPORTION OF FEMALES BETWEEN CONSECUTIVE YEARS



Note: The top figure presents the density of the change in the proportion of females within schools for each pair of consecutive years. The bottom figure presents the density of the change in the proportion of females within neighborhoods for each pair of consecutive years.

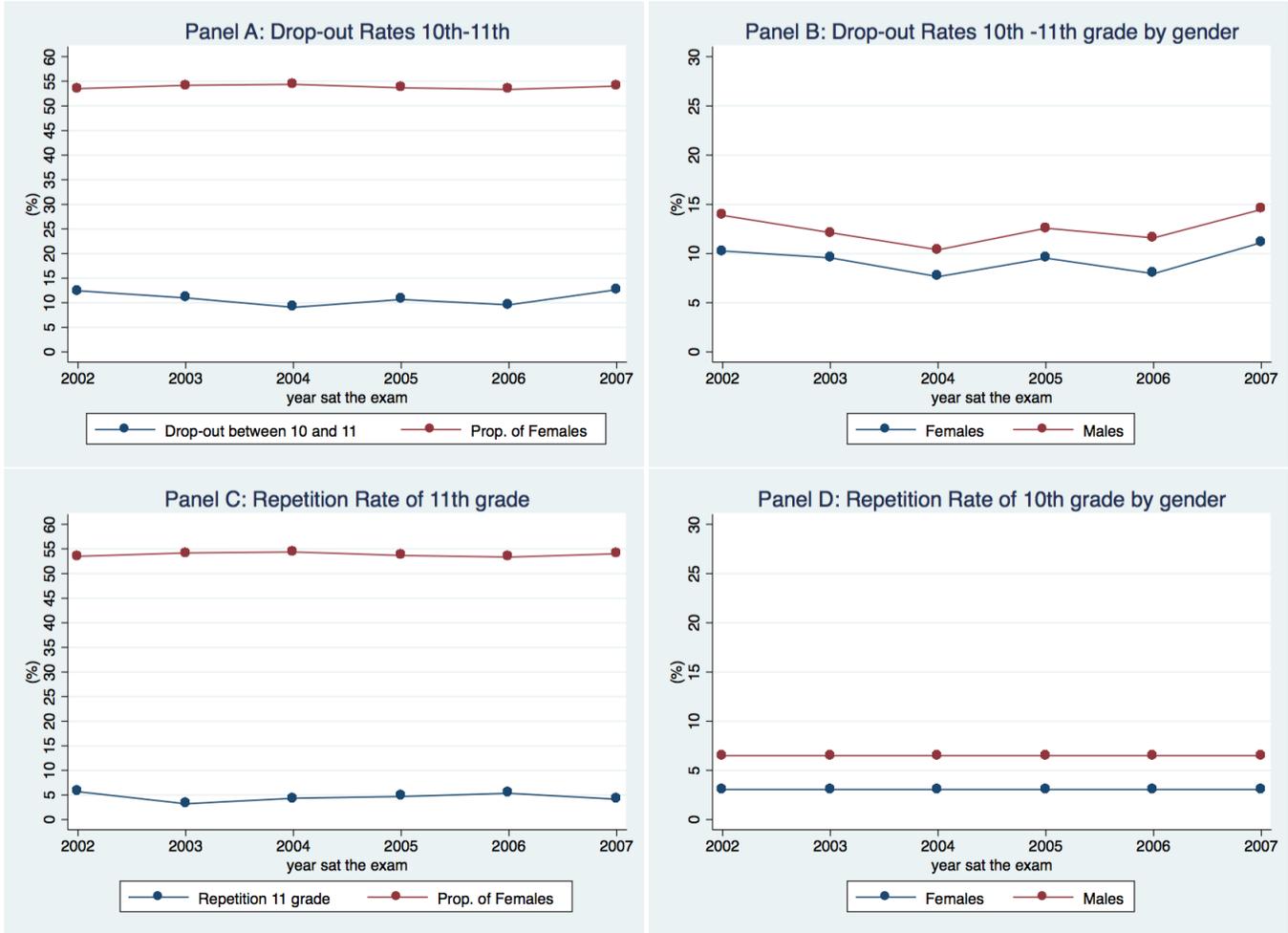
Figure 5:

SIMULATED MONTE CARLO S.D. WITHIN SCHOOLS AND NEIGHBORHOODS



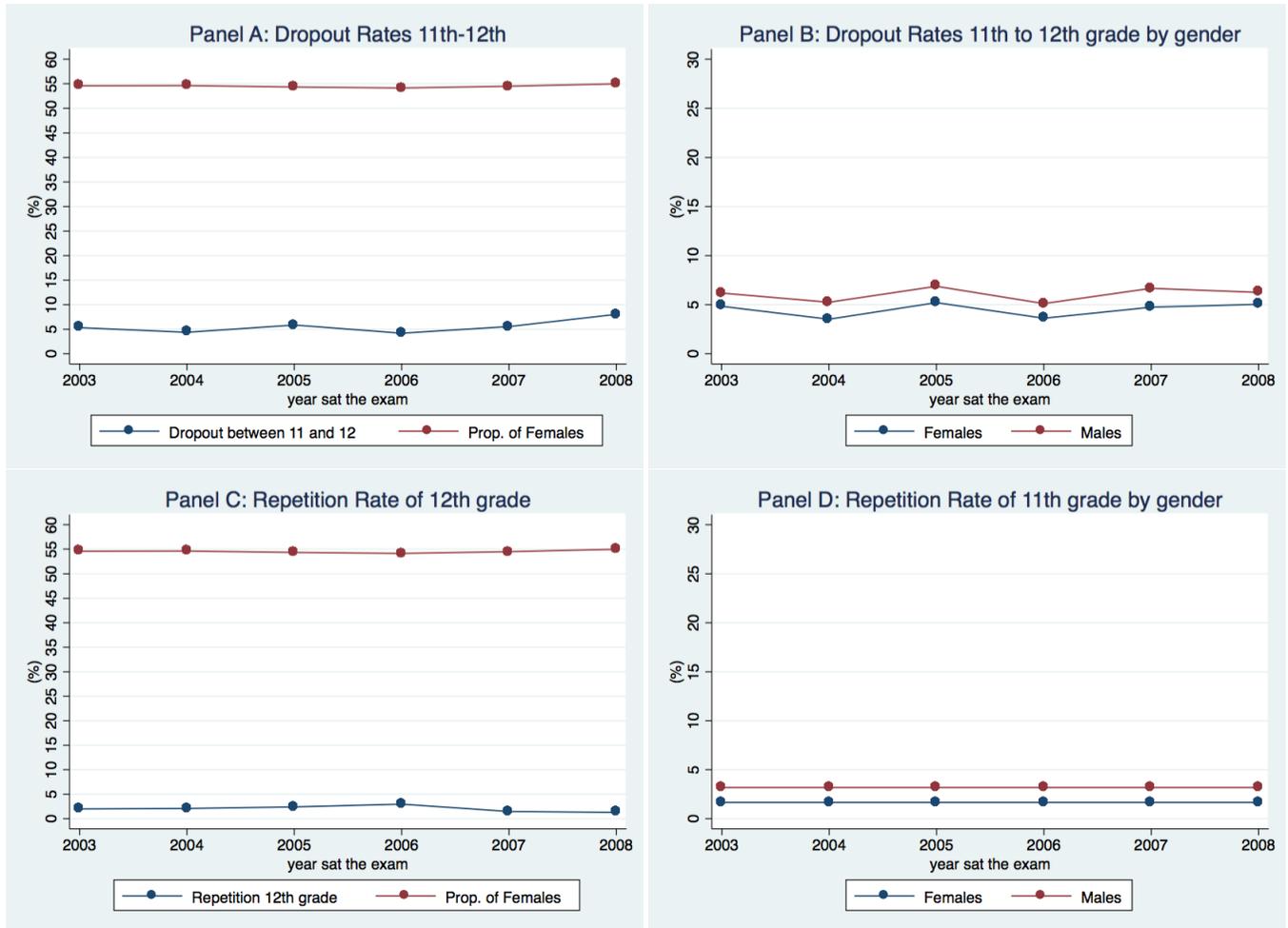
Note: For each school and neighborhood, we randomly generate the gender of the students in each cohort using a binomial distribution function with p equal to the average proportion of females in the school across all years. We then compute the within-school and -neighborhood standard deviation of the proportion of females. Details about the Monte Carlo simulation are provided in the text.

Figure 6:
 REPETITION AND DROP OUT RATES (10^{th} - 11^{th}) BY GENDER, SMALLER SAMPLE



Note: A smaller sample of 144 schools is used. Panel A presents the annual drop-out rate from 10^{th} to 11^{th} grade, along with the 11^{th} grade annual percentage of female students. Panel B presents the annual dropout rate from 10^{th} to 11^{th} grade for females and males, separately. Panel C presents the annual repetition rate in 11^{th} grade, along with the 11^{th} grade annual percentage of female students. Panel D presents the repetition rate in 10^{th} grade for males and females, separately.

Figure 7:
 REPETITION AND DROPOUT RATES (11th-12th) BY GENDER, SMALLER SAMPLE



Note: A smaller sample of 144 schools is used. Panel A presents the annual dropout rate from 11th to 12th grade, along with the 12th grade annual percentage of female students. Panel B presents the annual dropout rate from 11th to 12th grade for females and males, separately. Panel C presents the annual repetition rate in 12th grade, along with the 12th grade annual percentage of female students. Panel D presents the repetition rate in 11th grade for males and females, separately.

Table 1: DESCRIPTIVE STATISTICS

| | Mean | Std. Dev. | Min. | Max. | N |
|---|---------|-----------|--------|--------|---------|
| Panel A: Individual Level | | | | | |
| Age | 16.892 | 0.538 | 14 | 58 | 324,451 |
| Female | 0.571 | 0.495 | 0 | 1 | 324,451 |
| Born in First Quarter of Birth Year | 0.160 | 0.367 | 0 | 1 | 324,451 |
| National Exams Score | 64.920 | 20.26 | 2.6 | 99.75 | 324,451 |
| School Midterm Score | 87.057 | 10.406 | 0 | 100 | 324,451 |
| School Rank (based on School Midterm Score) | 37.651 | 30.061 | 1 | 244 | 324,451 |
| Neighborhood Rank (based on School Midterm Score) | 209.363 | 271.543 | 1 | 1,928 | 324,451 |
| Specialty in Classics | 0.366 | 0.482 | 0 | 1 | 324,451 |
| Specialty in Science | 0.148 | 0.355 | 0 | 1 | 324,451 |
| Specialty in Exact Science | 0.486 | 0.500 | 0 | 1 | 324,451 |
| Matriculation Status | 0.806 | 0.395 | 0 | 1 | 324,451 |
| Matriculation Score (std.) | 0.000 | 1.000 | -2.876 | 3.083 | 261,509 |
| Postcode Income (Euros, 2009) | 20,866 | 6,179 | 10,767 | 74,798 | 324,451 |
| Panel B: School Level | | | | | |
| Age | 16.898 | 0.096 | 16.462 | 18.103 | 1,097 |
| Prop. of Females | 0.578 | 0.060 | 0.143 | 0.882 | 1,097 |
| Prop. of Students Born in First Quarter of Birth Year | 0.165 | 0.048 | 0 | 0.538 | 1,097 |
| Public | 1 | 0 | 0 | 1 | 1,097 |
| Urban | 0.807 | 0.395 | 0 | 1 | 1,097 |
| Postcode Income (Euros, 2009) | 19,766 | 6,108 | 10,767 | 74,798 | 1,097 |
| No. of Students in each School | 53 | 32 | 12 | 190 | 1,097 |
| Panel C: Neighborhood-Cluster Level | | | | | |
| Age | 16.897 | 0.055 | 16.774 | 17.186 | 222 |
| Prop of Females | 0.574 | 0.041 | 0.382 | 0.749 | 222 |
| Prop. of Students Born in First Quarter of Birth Year | 0.164 | 0.029 | 0 | 0.302 | 222 |
| Neighborhood Income (Euro, 2009) | 19,961 | 6,153 | 6,260 | 52,591 | 222 |
| No. of Schools in each Cluster | 4.234 | 4.172 | 2 | 32 | 222 |
| No. of Students in each Cluster | 231.387 | 239 | 25 | 1,635 | 222 |

Note: Data span six cohorts, 2004-2009. Number of schools: 1,097. Number of neighborhoods: 222. The dummy variable “Born in First Quarter of Birth Year” equals one if a student is born in the first quarter of the birth year. The dummy variable “Matriculation Status” equals one if a student is enrolled in some postsecondary institution and zero otherwise. Students must enroll in a track at the beginning of the 12th grade, and that is their track or *specialty*. There are three track options: classics, science, and exact science. The sample contains only public school data. The university enrollment grade is observed only for students who enroll in some university department.

Table 2: DESCRIPTIVE STATISTICS BY YEAR IN THE SAMPLE

| Year | No. of students | No. of schools | No. of neigh. | Proportion | Proportion | National exam score | |
|-----------|-----------------|----------------|---------------|--------------------------|-------------------------|---------------------|---------|
| | | | | of girls in school (sd.) | of girls in neigh.(sd.) | Males | Females |
| (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| 2003-2004 | 62,104 | 1,006 | 213 | 0.558 (0.076) | 0.558 (0.051) | 58.7 | 60.6 |
| 2004-2005 | 63,617 | 1,026 | 213 | 0.560 (0.074) | 0.558 (0.052) | 57.0 | 59.7 |
| 2005-2006 | 61,325 | 1,007 | 210 | 0.564 (0.076) | 0.563 (0.050) | 59.1 | 62.5 |
| 2006-2007 | 47,334 | 945 | 201 | 0.586 (0.083) | 0.588 (0.055) | 69.5 | 71.0 |
| 2007-2008 | 46,173 | 925 | 198 | 0.587 (0.082) | 0.590 (0.057) | 70.1 | 72.9 |
| 2008-2009 | 44,190 | 926 | 198 | 0.581 (0.087) | 0.582 (0.062) | 72.6 | 74.7 |
| All | 322,507 | 1,097 | 222 | 0.571 (0.080) | 0.571 (0.056) | 63.4 | 66.1 |

| | Matriculation Status | | Matriculation Score | | Enrolled in University vs Technical School | | Expected Occupation Wage (std.) | |
|-----------|----------------------|---------|---------------------|---------|--|---------|---------------------------------|---------|
| | Males | Females | Males | Females | Males | Females | Males | Females |
| | (11) | (12) | (13) | (14) | (15) | (16) | (17) | (18) |
| 2003-2004 | 0.800 | 0.814 | -0.069 | 0.054 | 0.497 | 0.565 | 0.082 | -0.108 |
| 2004-2005 | 0.798 | 0.821 | -0.091 | 0.069 | 0.494 | 0.551 | 0.057 | -0.113 |
| 2005-2006 | 0.629 | 0.677 | -0.071 | 0.051 | 0.598 | 0.666 | 0.203 | -0.062 |
| 2006-2007 | 0.841 | 0.822 | -0.089 | 0.064 | 0.590 | 0.647 | 0.201 | -0.077 |
| 2007-2008 | 0.869 | 0.875 | -0.139 | 0.097 | 0.647 | 0.698 | 0.173 | -0.116 |
| 2008-2009 | 0.911 | 0.909 | -0.106 | 0.076 | 0.696 | 0.712 | 0.147 | -0.123 |
| All | 0.797 | 0.813 | -0.093 | 0.068 | 0.578 | 0.634 | 0.136 | -0.101 |

Note: This table presents the evolution of descriptive statistics from year to year in the sample. Data span six graduating cohorts from 2004 to 2009.

Table 3: VARIANCE DECOMPOSITION FOR THE PROPORTION OF FEMALES

| | School | | | Neighborhood | | |
|-----------------------------|-----------------------|-----------------------|-----------|-----------------------|-----------------------|-----------|
| | Sum of squares (1) | Share of total (2) | DF (3) | Sum of squares (4) | Share of total (5) | DF (6) |
| Within school/Neighborhood | 35.97 | 69% | 4,731 | 2.99 | 63% | 1,011 |
| Between school/Neighborhood | 15.97 | 31% | 1,096 | 1.74 | 37% | 221 |
| Total | 51.94 | | 5,827 | 6.31 | | 1,397 |

Note: The variance decomposition is done separately at the school and neighborhood level. Columns (1)-(3) present the sum of squares, share of total, and degrees of freedom for the within- and between-schools variation. Columns (4)-(6) present the sum of squares, share of total, and degrees of freedom for within- and between-neighborhoods variation.

Table 4: BALANCING TESTS

| | School | | Neighborhood | |
|---|---------------------|-------------------|--------------------|-------------------|
| | (1) | (2) | (3) | (4) |
| Age | 0.005 (0.018) | 0.013 (0.020) | -0.021 (0.025) | -0.006 (0.026) |
| Born in First Quarter of Birth Year | -0.002 (0.010) | -0.005 (0.012) | -0.008 (0.015) | -0.025 (0.015) |
| Enrollment/Size | -5.349 (2.484)** | -3.246 (2.394) | 15.214 (16.075) | 0.885 (10.690) |
| Year FE | ✓ | ✓ | ✓ | ✓ |
| School/neighborhood FE | ✓ | ✓ | ✓ | ✓ |
| School/neighborhood specific linear trend | | ✓ | | ✓ |
| No of students | 324,451 | | 283,733 | |
| No of schools/neighborhoods | 1,097 | | 222 | |

Note: The table presents school (columns 1, 2) and neighborhood (columns 3, 4) fixed-effects estimates. The dependent variable is students' predetermined characteristics (age or a dummy for being born in first quarter of birth year) or school/neighborhood enrollment/size. Our variable of interest is the proportion of females in each unit (school or neighborhood). For each unit, we report unit (school or neighborhood) fixed-effect estimates. Columns 1-2 report the estimates of the proportion of females at the school level, while columns 3-4 report the estimates of the proportion of females at the neighborhood level. In columns (1) and (3) we include school and neighborhood fixed effects, respectively. In columns (2) and (4) we also add unit (school or neighborhood)-specific linear trends. Year dummies are included in all regressions. Standard errors are clustered at the school level. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively.

Table 5: ESTIMATES OF THE EFFECT OF PROPORTION FEMALE IN THE SCHOOL AND NEIGHBORHOOD ON HIGH SCHOOL SCHOLASTIC OUTCOMES

| | School | | | | | | Neighborhood | | | | | |
|--|--------|------------|------------|--------|------------|------------|--------------|------------|------------|--------|------------|------------|
| | Mean | BOYS | | Mean | GIRLS | | Mean | BOYS | | Mean | GIRLS | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) |
| National Exam Score | 63.351 | 3.557 | 2.873 | 66.101 | 4.434 | 3.433 | 63.556 | 3.281 | 2.909 | 66.268 | 3.912 | 3.346 |
| | | (1.132)*** | (0.910)*** | | (1.084)*** | (0.777)*** | | (1.565)** | (1.555)* | | (1.432)*** | (1.311)** |
| <i>N</i> | | 139,345 | 139,345 | | 185,106 | 185,106 | | 122,194 | 122,194 | | 161,539 | 161,539 |
| Matriculation Status | 0.797 | 0.055 | 0.046 | 0.813 | 0.047 | 0.033 | 0.800 | -0.021 | -0.028 | 0.816 | 0.050 | 0.042 |
| | | (0.023)** | (0.021)** | | (0.020)** | (0.017)** | | (0.031) | (0.030) | | (0.026)* | (0.025)* |
| <i>N</i> | | 139,345 | 139,345 | | 185,106 | 185,106 | | 122,194 | 122,194 | | 161,539 | 161,539 |
| Matriculation Score (std.) | -0.093 | 0.162 | 0.143 | 0.068 | 0.252 | 0.211 | -0.083 | 0.337 | 0.342 | 0.076 | 0.235 | 0.217 |
| | | (0.067)** | (0.053)*** | | (0.066)*** | (0.048)*** | | (0.087)*** | (0.085)*** | | (0.080)*** | (0.073)*** |
| <i>N</i> | | 111,023 | 111,023 | | 150,475 | 150,475 | | 97,737 | 97,737 | | 131,741 | 131,741 |
| Enrolled in an Academic University | 0.578 | 0.062 | 0.056 | 0.634 | 0.097 | 0.084 | 0.582 | 0.154 | 0.156 | 0.637 | 0.109 | 0.104 |
| | | (0.030)** | (0.026)** | | (0.027)*** | (0.022)*** | | (0.038)*** | (0.038)*** | | (0.035)*** | (0.033)*** |
| <i>N</i> | | 111,023 | 111,023 | | 150,475 | 150,475 | | 97,737 | 97,737 | | 131,741 | 131,741 |
| Quality of Enrolled Postsecondary Degree | 47.259 | 4.830 | 4.315 | 51.736 | 8.091 | 6.930 | 47.541 | 10.740 | 10.872 | 51.923 | 7.837 | 7.395 |
| | | (1.889)** | (1.493)*** | | (1.885)*** | (1.373)*** | | (2.530)*** | (2.478)*** | | (2.298)** | (2.083)** |
| <i>N</i> | | 111,023 | 111,023 | | 150,475 | 150,475 | | 97,737 | 97,737 | | 131,741 | 131,741 |
| Year FE. | | ✓ | ✓ | | ✓ | ✓ | | ✓ | ✓ | | ✓ | ✓ |
| School/Neighborhood FE. | | ✓ | ✓ | | ✓ | ✓ | | ✓ | ✓ | | ✓ | ✓ |
| School/Neighborhood-specific linear trends | | ✓ | ✓ | | ✓ | ✓ | | ✓ | ✓ | | ✓ | ✓ |
| Student-level controls | | ✓ | ✓ | | ✓ | ✓ | | ✓ | ✓ | | ✓ | ✓ |
| Student characteristics | | | ✓ | | ✓ | ✓ | | | ✓ | | ✓ | ✓ |
| School/Neighborhood -by-year characteristics | | | ✓ | | | ✓ | | | ✓ | | | ✓ |
| No. of Schools/ Neighborhoods | | | | 1,097 | | | | | | 222 | | |

Note: Columns (1), (4), (7) and (10) report the mean of each outcome variable for boys and girls in the related unit (school/neighborhood). Columns (2), (3), (5) and (6) present estimates from school fixed-effects regressions, for boys (columns 2, 3) and girls (columns 5, 6), separately. Columns (8), (9), (11) and (12) present estimates from neighborhood fixed-effects regressions, for boys (columns 8, 9) and girls (columns 11, 12), separately. Regressions control for year fixed effects; school/neighborhood fixed effects; school/neighborhood-specific linear time trends; student-level controls (prior performance and dummies for special interest in mathematics, science, or exact science); student predetermined characteristics (age, dummy for the quarter of birth); and school/neighborhood-by-year characteristics, which include cohort mean controls (students' characteristics averaged by school/neighborhood and year) and each year's school/neighborhood enrollment. Each estimate is generated from a different regression. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively. Robust standard errors clustered at the school level are reported in parentheses.

Table 6: ESTIMATES OF THE EFFECT OF PROPORTION FEMALES IN THE SCHOOL AND NEIGHBORHOOD ON THE CHOICE OF UNIVERSITY STUDY

| | School | | | | | | Neighborhood | | | | | |
|---|--------|------------|------------|--------|------------|------------|--------------|---------|---------|--------|------------|------------|
| | Mean | BOYS | | Mean | GIRLS | | Mean | BOYS | | Mean | GIRLS | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) |
| Enrol in Mathematics & Science | 0.082 | 0.048 | 0.047 | 0.075 | 0.021 | 0.019 | 0.067 | 0.029 | 0.027 | 0.062 | 0.040 | 0.040 |
| | | (0.014)*** | (0.014)*** | | (0.011)** | (0.014)* | | (0.019) | (0.019) | | (0.014)*** | (0.014)*** |
| <i>N</i> | | 111,023 | 111,023 | | 150,505 | 150,505 | | 97,737 | 97,737 | | 131,742 | 131,742 |
| Enrol in STEM | 0.408 | 0.023 | 0.022 | 0.217 | 0.040 | 0.038 | 0.326 | 0.054 | 0.047 | 0.178 | 0.051 | 0.048 |
| | | (0.026) | (0.027) | | (0.018)** | (0.018)** | | (0.037) | (0.037) | | (0.025)** | (0.025)* |
| <i>N</i> | | 111,023 | 111,023 | | 150,505 | 150,505 | | 97,737 | 97,737 | | 131,742 | 131,742 |
| Expected Occupation Wage (std.) | 0.136 | 0.007 | 0.002 | -0.101 | 0.135 | 0.125 | 0.136 | 0.090 | 0.089 | -0.100 | 0.210 | 0.207 |
| | | (0.053) | (0.051) | | (0.047)*** | (0.045)*** | | (0.069) | (0.069) | | (0.065)*** | (0.065)*** |
| <i>N</i> | | 111,023 | 111,023 | | 150,475 | 150,475 | | 97,737 | 97,737 | | 131,741 | 131,741 |
| Year FE. | | ✓ | ✓ | | ✓ | ✓ | | ✓ | ✓ | | ✓ | ✓ |
| School/Neighborhood FE. | | ✓ | ✓ | | ✓ | ✓ | | ✓ | ✓ | | ✓ | ✓ |
| School/Neighborhood-specific linear trends | | ✓ | ✓ | | ✓ | ✓ | | ✓ | ✓ | | ✓ | ✓ |
| Student-level controls | | ✓ | ✓ | | ✓ | ✓ | | ✓ | ✓ | | ✓ | ✓ |
| Student characteristics | | | ✓ | | ✓ | ✓ | | | ✓ | | ✓ | ✓ |
| School/Neighborhood-by-year characteristics | | | ✓ | | | ✓ | | | ✓ | | | ✓ |
| No. of Schools/ Neighborhoods | | | | | 1,097 | | | | | 222 | | |

Note: Columns (1), (4), (7) and (10) report the mean of each outcome variable for boys and girls in the related unit (school/neighborhood). Regressions control for year fixed effects; school/neighborhood fixed effects; school/neighborhood-specific linear time trends; student-level controls (prior performance and dummies for special interest in mathematics, science, or exact science); student predetermined characteristics (age, dummy for the quarter of birth); and school/neighborhood-by-year characteristics, which include cohort mean controls (students' characteristics averaged by school/neighborhood and year) and each year's 12th-grade school/neighborhood enrollment. STEM university degrees include degrees in mathematics, science, computer science and engineering. Science includes departments in physical and earth sciences, biology, veterinary science, medicine and pharmacy. Each estimate is generated from a different regression. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively. Robust standard errors clustered at the school level are reported in parentheses.

Table 7: NONLINEAR ESTIMATES OF THE EFFECT OF PROPORTION FEMALES IN THE SCHOOL ON ACADEMIC OUTCOMES IN 12th GRADE AND CHOICE OF UNIVERSITY STUDY

| | BOYS | | | | GIRLS | | | |
|---|-------------------|---------------------|--------------------|---------------------|-------------------|---------------------|---------------------|---------------------|
| | Quintile 2 | Quintile 3 | Quintile 4 | Quintile 5 | Quintile 2 | Quintile 3 | Quintile 4 | Quintile 5 |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| National Exam Score | 0.149 (0.242) | 0.490 (0.226)** | 0.562 (0.220)** | 0.678 (0.226)*** | 0.232 (0.184) | 0.530 (0.183)*** | 0.995 (0.186)*** | 0.688 (0.195)*** |
| Matriculation Status | 0.0002 (0.005) | 0.004 (0.005) | 0.009 (0.005)* | 0.008 (0.005) | -0.002 (0.004) | 0.004 (0.004) | 0.012 (0.004)*** | 0.005 (0.004) |
| Matriculation Score (std.) | 0.007 (0.013) | 0.024 (0.012)** | 0.030 (0.012)** | 0.037 (0.013)*** | 0.019 (0.011)* | 0.035 (0.011)*** | 0.053 (0.011)*** | 0.040 (0.011)*** |
| Enroll in an Academic University | 0.007 (0.006) | 0.016 (0.006)** | 0.014 (0.006)** | 0.017 (0.007)*** | 0.003 (0.005) | 0.009 (0.006) | 0.015 (0.005)*** | 0.016 (0.006)*** |
| Quality of Enrolled Postsecondary Degree | 0.257 (0.368) | 0.914 (0.347)** | 1.090 (0.347)** | 1.297 (0.386)*** | 0.501 (0.331) | 0.941 (0.327)*** | 1.668 (0.327)*** | 1.414 (0.351)*** |
| University Field Enrollment | | | | | | | | |
| Enrol in Science & Mathematics | 0.004 (0.003) | 0.008 (0.003)*** | 0.008 (0.003)** | 0.010 (0.003)*** | -0.002 (0.002) | 0.002 (0.002) | 0.004 (0.002)* | 0.004 (0.002)* |
| Enrol in STEM | 0.002 (0.005) | -0.001 (0.005) | 0.003 (0.005) | 0.005 (0.005) | 0.0001 (0.003) | 0.004 (0.003) | 0.006 (0.003)* | 0.008 (0.004)** |
| Expected Occupation Wages (std.) | -0.012 (0.011) | -0.012 (0.012) | -0.002 (0.012) | -0.002 (0.013) | 0.008 (0.011) | 0.018 (0.011)* | 0.016 (0.011) | 0.028 (0.011)** |
| <i>Sample Size</i> | 28,584 | 28,344 | 27,385 | 25,209 | 36,243 | 36,621 | 37,318 | 39,693 |
| Year FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| School FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| School-specific linear trends | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Student-level controls | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Student characteristics | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| School/Neighborhood-by-year characteristics | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |

Notes: The table reports nonlinear effects of the proportion of female students in school on boys' and girls' academic outcomes and choices. The model replaces the single treatment variable with a set of quintile indicators for the different quintiles for the proportion of female students in schools. The omitted category is quintile 1 (0.076<proportion of females<0.509). STEM university degrees include degrees in the departments of science, technology, engineering and mathematics. Science includes departments in physical and earth sciences, biology, veterinary science, medicine and pharmacy. Estimates in each row by gender are generated from the same regression. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively. Robust standard errors clustered at the school level are reported in parentheses.

Table 8: NONLINEAR ESTIMATES OF THE EFFECT OF PROPORTION FEMALES IN THE NEIGHBORHOOD ON ACADEMIC OUTCOMES IN 12th GRADE AND CHOICE OF UNIVERSITY STUDY

| | BOYS | | | | GIRLS | | | |
|---|-------------------|---------------------|---------------------|---------------------|--------------------|---------------------|---------------------|---------------------|
| | Quintile 2 | Quintile 3 | Quintile 4 | Quintile 5 | Quintile 2 | Quintile 3 | Quintile 4 | Quintile 5 |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| National Exam Score | -0.069 (0.242) | 0.487 (0.297)* | 0.311 (0.322) | 0.354 (0.275) | 0.049 (0.224) | 0.594 (0.247)** | 0.581 (0.275)** | 0.420 (0.234)* |
| Matriculation Status | -0.002 (0.005) | 0.001 (0.005) | -0.002 (0.005) | -0.007 (0.005) | 0.001 (0.004) | 0.010 (0.004)* | 0.009 (0.005)* | 0.006 (0.004) |
| Matriculation Score (std.) | -0.003 (0.012) | 0.042 (0.013)*** | 0.046 (0.014)** | 0.050 (0.012)*** | -0.001 (0.011) | 0.034 (0.012)*** | 0.042 (0.013)*** | 0.026 (0.012)** |
| Enroll in an Academic University | 0.002 (0.006) | 0.019 (0.006)*** | 0.026 (0.007)*** | 0.026 (0.007)*** | 0.003 (0.005) | 0.014 (0.006)*** | 0.020 (0.006)*** | 0.016 (0.006)*** |
| Quality of Enrolled Postsecondary Degree | -0.260 (0.354) | 1.152 (0.392)*** | 1.329 (0.434)*** | 1.707 (0.419)*** | -0.120 (0.343) | 0.863 (0.355)** | 1.297 (0.397)*** | 1.092 (0.368)*** |
| University Field Enrollment | | | | | | | | |
| Enrol in Science & Mathematics | -0.001 (0.003) | 0.003 (0.003) | 0.003 (0.003) | 0.004 (0.003) | 0.006 (0.002)** | 0.008 (0.002)*** | 0.004 (0.003)* | 0.006 (0.002)** |
| Enrol in STEM | -0.001 (0.005) | 0.002 (0.005) | 0.005 (0.006) | 0.012 (0.006)*** | 0.001 (0.003) | 0.005 (0.003) | 0.006 (0.004)* | 0.006 (0.004)* |
| Expected Occupation Wages (std.) | -0.005 (0.002) | 0.011 (0.002) | -0.003 (0.002) | 0.0132 (0.002) | 0.019 (0.011)* | 0.027 (0.010)*** | 0.023 (0.011)** | 0.039 (0.010)*** |
| <i>Sample Size</i> | 28,584 | 28,344 | 27,385 | 25,209 | 36,243 | 36,621 | 37,318 | 39,693 |
| Year FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| School FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| School-specific linear trends | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Student-level controls | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Student characteristics | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| School/Neighborhood-by-year characteristics | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |

Notes: Notes: The table reports nonlinear effects of the proportion of female students in neighborhood on boys' and girls' academic outcomes and choices. The model replaces the single treatment variable with a set of quintile indicators for the different quintiles for the proportion of female students in neighborhoods. The omitted category is quintile 1 (0.200 < proportion of females < 0.532). STEM university degrees include degrees in the departments of science, technology, engineering and mathematics. Science includes departments in physical and earth sciences, biology, veterinary science, medicine and pharmacy. Estimates in each row by gender are generated from the same regression. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively. Robust standard errors clustered at the school level are reported in parentheses.

Table 9: FALSIFICATION EXERCISE, FALSE PEERS, SCHOOL AND NEIGHBORHOOD

| | False Peers in | | | |
|---|----------------|---------|--------------|----------|
| | School | | Neighborhood | |
| | BOYS | GIRLS | BOYS | GIRLS |
| | (1) | (2) | (3) | (4) |
| National Exam Score | -1.178 | -0.200 | 5.823 | 11.561 |
| | (1.370) | (1.278) | (8.179) | (7.282) |
| <i>N</i> | 97,442 | 127,221 | 121,773 | 160,823 |
| Matriculation Status | -0.024 | -0.009 | 0.055 | -0.062 |
| | (0.029) | (0.026) | (0.173) | (0.141) |
| <i>N</i> | 97,442 | 127,221 | 121,773 | 160,823 |
| Matriculation Score (std.) | 0.017 | -0.020 | -0.030 | 0.541 |
| | (0.082) | (0.080) | (0.417) | (0.387) |
| <i>N</i> | 78,598 | 104,675 | 95,531 | 128,490 |
| Enroll in an Academic University | 0.055 | 0.006 | 0.066 | 0.086 |
| | (0.039) | (0.035) | (0.196) | (0.174) |
| <i>N</i> | 78,598 | 104,675 | 97,380 | 131,156 |
| Quality of Enrolled Postsecondary Degree | 1.026 | -0.873 | 1.158 | 11.120 |
| | (2.400) | (2.274) | (11.799) | (10.934) |
| <i>N</i> | 78,598 | 104,675 | 97,380 | 131,156 |
| University Field Enrollment | | | | |
| Enroll in Science & Mathematics | 0.003 | 0.002 | -0.035 | 0.115 |
| | (0.023) | (0.016) | (0.111) | (0.096) |
| <i>N</i> | 78,598 | 104,675 | 97,380 | 131,156 |
| Enroll in STEM | 0.013 | 0.023 | -0.374 | -0.081 |
| | (0.039) | (0.028) | (0.216)* | (0.155) |
| <i>N</i> | 78,598 | 104,675 | 97,380 | 131,156 |
| Expected Occupation Wages (std.) | 0.071 | 0.052 | 0.468 | -0.151 |
| | (0.082) | (0.071) | (0.436) | (0.353) |
| <i>N</i> | 78,598 | 104,675 | 97,380 | 131,156 |
| Year FE. | ✓ | ✓ | ✓ | ✓ |
| School/Neighborhood FE. | ✓ | ✓ | ✓ | ✓ |
| School/Neighborhood-specific linear trends | ✓ | ✓ | ✓ | ✓ |
| Student-level controls | ✓ | ✓ | ✓ | ✓ |
| Student characteristics | ✓ | ✓ | ✓ | ✓ |
| School/Neighborhood-by-year characteristics | ✓ | ✓ | ✓ | ✓ |

Note: The outcome variable in columns (1) and (2) is the proportion of females in all other schools in a student's postcode, except their own school. The outcome variable in columns (3) and (4) is the proportion of females in all other clusters in a student's periphery, except their own cluster. Regressions control for year fixed effects, school/neighborhood fixed effects, school/neighborhood-specific linear time trends, student-level controls (prior performance and dummies for special interest in mathematics, science, or exact science), student predetermined characteristics (age, dummy for the quarter of birth), and school/neighborhood-by-year characteristics, which include cohort mean controls (students' characteristics averaged by school/neighborhood and year) and controls for each year's 12th grade school/neighborhood enrollment. Each estimate is generated from a different regression. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively. Robust standard errors clustered at the school level are reported in parentheses.

Table 10: HETEROGENEOUS PEER EFFECTS ON ACADEMIC OUTCOMES AND DECISIONS BY SCHOOL/NEIGHBORHOOD SIZE

| | School | | | | Neighborhood | | | |
|---|---------------------|---------------------|---------------------|---------------------|--------------------|----------------------|--------------------|----------------------|
| | BOYS | | GIRLS | | BOYS | | GIRLS | |
| | Below | Above | Below | Above | Below | Above | Below | Above |
| | School Average | | School Average | | Neigh/hood Average | | Neigh/hood Average | |
| | <=53 | >53 | <=53 | >53 | <=232 | >232 | <=232 | >232 |
| (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | |
| National Exam Score | 3.159 (1.311)** | 3.900 (1.450)** | 3.643 (1.097)*** | 4.769 (1.332)*** | 1.397 (1.369) | 3.977 (6.097) | 0.418 (1.221) | 8.348 (4.862)* |
| Matriculation Status | 0.061 (0.033)* | 0.034 (0.031) | 0.061 (0.025)** | 0.031 (0.026) | -0.040 (0.031) | -0.046 (0.091) | 0.002 (0.025) | 0.101 (0.087) |
| Matriculation Score (std.) | 0.177 (0.084)** | 0.235 (0.084)*** | 0.208 (0.072)*** | 0.275 (0.082)*** | 0.158 (0.082)* | 0.962 (0.229)*** | 0.053 (0.076) | 0.855 (0.215)*** |
| Enroll in an Academic University | 0.048 (0.042) | 0.078 (0.040)** | 0.125 (0.034)*** | 0.084 (0.037)** | 0.073 (0.039)* | 0.569 (0.106)*** | 0.060 (0.034)* | 0.216 (0.091)** |
| Quality of Enrolled Postsecondary Degree | 5.225 (2.281)*** | 6.686 (2.400)*** | 7.027 (2.065)*** | 9.111 (2.338)*** | 5.791 (2.383)** | 28.574 (6.523)*** | 3.029 (2.134) | 23.964 (5.890)*** |
| University Field Enrollment | | | | | | | | |
| Enroll in Science & Mathematics | 0.038 (0.018)*** | 0.040 (0.022)* | 0.003 (0.014) | 0.024 (0.016) | 0.037 (0.020)* | -0.046 (0.057) | 0.031 (0.014)** | 0.089 (0.050)* |
| Enroll in STEM | 0.026 (0.039) | 0.043 (0.055) | 0.030 (0.030) | 0.025 (0.027) | 0.072 (0.038)* | -0.141 (0.103) | 0.039 (0.027) | 0.106 (0.076) |
| Expected Occupation Wages (std.) | 0.045 (0.078) | 0.022 (0.079) | 0.083 (0.068) | 0.191 (0.068)*** | 0.080 (0.074) | 0.008 (0.205) | 0.159 (0.070)** | 0.377 (0.180)** |
| Year FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| School/Neighborhood FE | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| School/Neighborhood-specific linear trends | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Student-level controls | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Student characteristics | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| School/Neighborhood-by-year characteristics | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |

Notes: The table reports estimates of the proportion of females by school and neighborhood size on boys' and girls' cognitive outcomes. The results are reported separately for schools/neighborhoods that have an enrollment size in a given year above and below the school/neighborhood average. The school average enrollment size is 53 students, while the neighborhood average enrollment size is 232 students. Regressions control for student-level controls (prior performance and dummies for special interest in mathematics, science, or exact science), student predetermined characteristics (age, dummy for the quarter of birth), school/neighborhood-by-year characteristics (students' characteristics averaged by school/neighborhood and year, school/neighborhood annual enrollment) year dummies, school/neighborhood fixed effects and school/neighborhood-specific linear time trends. Each estimate is generated from a different regression. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively. Robust standard errors clustered at the school level are reported in parentheses.

Table 11: PREVIOUS AND FOLLOWING COHORT PEER EFFECTS ESTIMATES

| Variable | School | | | | Neighborhood | | | |
|---|-----------------------------------|--------------------|-------------------|--------------------|---------------------------------------|--------------------|-------------------|-------------------|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| | Prop. of females in school | | | | Prop. of females in neigh/hood | | | |
| | in cohort t-1 | | in cohort t+1 | | in cohort t-1 | | in cohort t+1 | |
| | Males | Females | Males | Females | Males | Females | Males | Females |
| Variable | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| National Exams Score | -1.157 (0.989) | -0.202 (0.840) | -0.849 (0.905) | -1.520 (0.781)* | -0.882 (2.453) | -0.971 (1.986) | 3.048 (2.137) | 1.787 (1.728) |
| Matriculation Status | -0.017 (0.022) | 0.006 (0.018) | -0.008 (0.022) | -0.023 (0.017) | -0.042 (0.055) | 0.010 (0.042) | 0.039 (0.046) | -0.010 (0.037) |
| Matriculation Score (std.) | -0.075 (0.050) | -0.047 (0.055) | -0.084 (0.055) | -0.060 (0.048) | -0.084 (10.126) | -0.103 (0.099) | 0.114 (0.115) | 0.066 (0.093) |
| Enroll in an Academic University | -0.006 (0.028) | -0.043 (0.024)* | -0.036 (0.026) | -0.019 (0.023) | -0.091 (0.064) | -0.081 (0.049) | -0.014 (0.056) | 0.055 (0.048) |
| Quality of Enrolled Postsecondary Degree | -1.579 (1.591) | -1.396 (1.419) | -3.219 (1.699) | -2.132 (1.385) | -2.466 (3.662) | -2.677 (2.852) | 1.052 (3.326) | 1.478 (2.207) |
| University Field Enrollment | | | | | | | | |
| Enroll in Science & Mathematics | -0.015 (0.011) | -0.007 (0.009) | -0.013 (0.010) | -0.011 (0.008) | 0.014 (0.25) | -0.030 (0.020) | 0.000 (0.024) | -0.011 (0.018) |
| Enroll in STEM | -0.028 (0.023) | -0.028 (0.014)* | 0.013 (0.021) | 0.004 (0.014) | 0.032 (0.55) | -0.050 (0.033) | 0.045 (0.049) | 0.050 (0.032) |
| Expected Occupation Wages (std.) | -0.048 (0.053) | -0.048 (0.046) | -0.010 (0.051) | 0.039 (0.045) | 0.064 (0.119) | -0.181 (0.103)* | 0.051 (0.114) | 0.047 (0.101) |
| Year FE. | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| School/Neighborhood FE. | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| School/Neighborhood-specific linear trends | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Student-level controls | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Student characteristics | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| School/Neighborhood-by-year characteristics | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| No. of Schools/Neighborhoods | 1,097 | | | | 222 | | | |

Note: In this table, we replace the actual variable of interest (proportion of female students in year t) with the proportion of female students in the younger (t-1) or older (t+1) cohort within the same unit (school or neighborhood). Regressions control for student-level controls (prior performance and dummies for special interest in mathematics, science, or exact science), student predetermined characteristics (age, dummy for the quarter of birth), school/neighborhood-by-year characteristics (students' characteristics averaged by school/neighborhood and year, school/neighborhood annual enrollment) year dummies, school/neighborhood fixed effects and school/neighborhood-specific linear time trends. Each estimate is generated from a different regression. *, **, and *** denotes significance at the 10%, 5%, and 1% level, respectively. Robust standard errors clustered at the school level are reported in parentheses.

Table 12: ROBUSTNESS CHECK: CONTROLS FOR THE EXISTENCE OF SINGLE-SEX AND PRIVATE SCHOOLS

| | School | | | | Neighborhood | | | |
|---|---------------------|---------------------|---------------------|---------------------|----------------------|----------------------|---------------------|---------------------|
| | BOYS | | GIRLS | | BOYS | | GIRLS | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| National Exam Score | 2.873 (0.910)*** | 2.867 (0.912)*** | 3.433 (0.777)*** | 3.431 (0.777)*** | 2.909 (1.555)* | 2.897 (1.555)* | 3.346 (1.311)** | 3.346 (1.310)** |
| <i>N</i> | 139,345 | 139,345 | 185,106 | 185,106 | 122,194 | 122,194 | 161,539 | 161,539 |
| Matriculation Status | 0.046 (0.021)** | 0.046 (0.021)** | 0.033 (0.020)* | 0.033 (0.017)* | -0.021 (0.031) | -0.028 (0.030) | 0.050 (0.026)* | 0.041 (0.025)* |
| <i>N</i> | 139,345 | 139,345 | 185,106 | 185,106 | 122,194 | 122,194 | 161,539 | 161,539 |
| Matriculation Score (std.) | 0.143 (0.053)*** | 0.143 (0.053)** | 0.211 (0.048)*** | 0.212 (0.048)*** | 0.337 (0.087)*** | 0.339 (0.085)*** | 0.235 (0.080)*** | 0.215 (0.072)*** |
| <i>N</i> | 111,023 | 111,023 | 150,475 | 150,475 | 97,737 | 97,737 | 131,741 | 131,741 |
| Enroll in an Academic University | 0.056 (0.026)** | 0.055 (0.026)** | 0.084 (0.022)*** | 0.084 (0.022)*** | 0.154 (0.038)*** | 0.154 (0.038)*** | 0.109 (0.035)*** | 0.105 (0.033)*** |
| <i>N</i> | 111,023 | 111,023 | 150,475 | 150,475 | 97,737 | 97,737 | 131,741 | 131,741 |
| Quality of Enrolled Postsecondary Degree | 4.315 (1.493)*** | 4.285 (1.496)*** | 6.930 (1.373)*** | 6.920 (1.373)*** | 10.872 (2.478)*** | 10.763 (2.480)*** | 7.395 (2.083)*** | 7.312 (2.077)*** |
| <i>N</i> | 111,023 | 111,023 | 150,475 | 150,475 | 97,737 | 97,737 | 131,741 | 131,741 |
| University Field Enrollment | | | | | | | | |
| Enroll in Science & Mathematics | 0.047 (0.014)*** | 0.046 (0.014)*** | 0.019 (0.014) | 0.021 (0.011)* | 0.027 (0.019) | 0.028 (0.019) | 0.040 (0.014)*** | 0.038 (0.014)** |
| <i>N</i> | 111,023 | 111,023 | 150,475 | 150,475 | 97,737 | 97,737 | 131,741 | 131,741 |
| Enroll in STEM | 0.022 (0.027) | 0.021 (0.026) | 0.038 (0.018)** | 0.037 (0.018)** | 0.047 (0.037) | 0.049 (0.036) | 0.048 (0.025)* | 0.046 (0.025)* |
| <i>N</i> | 111,023 | 111,023 | 150,475 | 150,475 | 97,737 | 97,737 | 131,741 | 131,741 |
| Expected Occupation Wages (std.) | 0.002 (0.051) | 0.001 (0.051) | 0.125 (0.045)*** | 0.124 (0.045)** | 0.090 (0.069) | 0.087 (0.069) | 0.210 (0.065)*** | 0.197 (0.065)*** |
| <i>N</i> | 111,023 | 111,023 | 150,475 | 150,475 | 97,737 | 97,737 | 131,741 | 131,741 |
| Year FE. | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| School/Neighborhood FE. | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| School/Neighborhood-specific linear trends | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Student-level controls | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Student characteristics | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| School/Neighborhood-by-year characteristics | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| No. of boys and girls attending private schools | | ✓ | | ✓ | | ✓ | | ✓ |
| No. of Schools/ Neighborhoods | | | 1,097 | | | | 222 | |

Note: In columns (2), (4), (6) and (8) we control for the number of boys and girls that attend single-sex or private schools in the locality. We also report the estimates from the main analysis (columns 1, 3, 5 and 7) for comparison purposes. Regressions control for student-level controls (prior performance and dummies for special interest in mathematics, science, or exact science), student predetermined characteristics (age, dummy for the quarter of birth), school/neighborhood-by-year characteristics (students' characteristics averaged by school/neighborhood and year, school/neighborhood annual enrollment) year dummies, school/neighborhood fixed effects and school/neighborhood-specific linear time trends. Each estimate is generated from a different regression. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively. Robust standard errors clustered at the school level are reported in parentheses.

Table 13: ROBUSTNESS EXERCISE: CONTROLS FOR PROPORTION OF FEMALES IN PREVIOUS AND FOLLOWING COHORTS

| | School | | | | | | Neighborhood | | | | | |
|---|-------------------|---------------------|-------------------|-------------------|---------------------|-------------------|-------------------|---------------------|--------------------|-------------------|---------------------|-------------------|
| | BOYS | | | GIRLS | | | BOYS | | | GIRLS | | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) |
| | t-1 | t | t+1 | t-1 | t | t+1 | t-1 | t | t+1 | t-1 | t | t+1 |
| National Exam Score | -0.401 (1.076) | 2.265 (1.035)** | -0.179 (0.982) | 0.966 (0.923) | 3.743 (0.880)*** | -0.156 (0.871) | 1.474 (2.583) | 4.196 (1.759)*** | 4.904 (2.447)* | 1.137 (2.066) | 4.331 (1.442)*** | 3.668 (1.880)* |
| Matriculation Status | -0.002 (0.023) | 0.048 (0.024)** | 0.006 (0.024) | 0.014 (0.020) | 0.037 (0.020)* | -0.009 (0.019) | -0.047 (0.056) | -0.025 (0.034) | 0.018 (0.050) | 0.029 (0.045) | 0.053 (0.027)* | 0.014 (0.040) |
| Matriculation Score (std.) | -0.056 (0.062) | 0.095 (0.062) | -0.067 (0.062) | 0.037 (0.0454) | 0.230 (0.054)*** | 0.021 (0.052) | 0.112 (0.134) | 0.416 (0.092)*** | 0.297 (0.127)** | 0.012 (0.107) | 0.257 (0.080)*** | 0.152 (0.100) |
| Enrolled in an Academic University | 0.007 (0.032) | 0.045 (0.031) | -0.018 (0.030) | -0.014 (0.027) | 0.081 (0.026)*** | 0.006 (0.026) | -0.027 (0.069) | 0.164 (0.040)*** | 0.048 (0.062) | -0.025 (0.053) | 0.115 (0.036)*** | 0.097 (0.051)* |
| Quality of Enrolled Postsecondary Degree | -1.288 (2.734) | 1.530 (1.986)*** | -2.113 (2.680) | 1.512 (2.426) | 5.390 (1.678)*** | 0.592 (2.248) | 4.215 (0.064)* | 7.085 (0.040)** | 4.313 (0.063) | 1.844 (0.043) | 4.640 (0.026)** | 2.111 (0.052) |
| University Field Enrollment | | | | | | | | | | | | |
| Enrol in Science & Mathematics | -0.000 (0.016) | 0.045 (0.017)*** | -0.005 (0.015) | -0.007 (0.012) | 0.013 (0.012) | 0.009 (0.011) | 0.034 (0.034) | 0.036 (0.020) | 0.021 (0.035) | -0.022 (0.027) | 0.034 (0.015)** | -0.007 (0.023) |
| Enrol in STEM | -0.004 (0.031) | 0.032 (0.031) | 0.023 (0.030) | -0.019 (0.020) | 0.036 (0.021)* | 0.019 (0.019) | 0.118 (0.064)* | 0.083 (0.040)** | 0.089 (0.063) | -0.014 (0.043) | 0.054 (0.026)** | 0.083 (0.052) |
| Expected Occupation Wages (std.) | -0.065 (0.059) | -0.032 (0.058) | -0.033 (0.058) | 0.020 (0.051) | 0.162 (0.052)*** | 0.093 (0.051)* | 0.134 (0.126) | 0.140 (0.074)*** | 0.127 (0.121) | -0.085 (0.107) | 0.217 (0.069)*** | 0.100 (0.108) |
| Year FE. | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| School/Neighborhood FE. | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| School/Neighborhood-specific linear trends | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Student-level controls | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Student characteristics | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| School/Neighborhood-by-year characteristics | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| No. of Schools/ Neighbourhoods | 1,097 | | | | | | 222 | | | | | |

Note: Columns (1)-(6) present estimates from the school fixed-effects regressions, while columns (7)-(12) presents estimates from the neighborhood fixed-effects regressions. In each regression we control for past and future proportion of females in each unit. All regressions control for student-level controls (prior performance and dummies for special interest in mathematics, science, or computer science), student predetermined characteristics (age, dummy for the quarter of birth), school/neighborhood-by-year characteristics (students' characteristics averaged by school/neighborhood and year, school/neighborhood annual enrollment) year dummies, school/neighborhood fixed effects and school/neighborhood-specific linear time trends. *, **, and *** denotes significance at the 10%, 5%, and 1% level, respectively. Robust standard errors clustered at the school level are reported in parentheses.

Table 14: ROBUSTNESS CHECK: MOBILITY OF STUDENTS ACROSS SCHOOLS

| Dependent Variable: School's Enrollment in year t+1 | | | | |
|---|-------------------|-------------------|--------------------|--------------------|
| | (1) | (2) | (3) | (4) |
| School Rank in Year t | -1.187 (1.162) | -0.443 (1.068) | | |
| Proportion of Females in School s in Year t | | | 64.082 (90.967) | 34.643 (78.027) |
| School FE | X | ✓ | X | ✓ |
| Year FE | X | ✓ | X | ✓ |
| School enrollment in year t | ✓ | ✓ | ✓ | ✓ |
| School enrollment in year $t-1$ | ✓ | ✓ | ✓ | ✓ |
| Neighborhood characteristics | ✓ | ✓ | ✓ | ✓ |
| School-specific linear trends | ✓ | ✓ | ✓ | ✓ |
| Observations | 3,543 | 3,543 | 3,543 | 3,543 |

Note: *School Rank in year t* is calculated based on the average school national exam performance across all subjects. *School Rank in year t* is increasing in performance and takes values from 0 to 1, while the mean rank equals 0.5. *Neighborhood characteristics* include neighborhood income and neighborhood unemployment. Robust standard errors clustered at the school level are reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively.

Appendices

Table A.1: CORRELATIONS BETWEEN PROPORTION OF FEMALES WITHIN SCHOOLS IN DIFFERENT YEARS

| | girls_sch_t | girls_sch_t+1 | girls_sch_t+2 | girls_sch_t+3 | girls_sch_t+4 | girls_sch_t+5 |
|---------------|-------------|---------------|---------------|---------------|---------------|---------------|
| girls_sch_t | 1 | | | | | |
| girls_sch_t+1 | 0.156 | 1 | | | | |
| girls_sch_t+2 | 0.131 | 0.151 | 1 | | | |
| girls_sch_t+3 | 0.160 | 0.110 | 0.143 | 1 | | |
| girls_sch_t+4 | 0.112 | 0.127 | 0.085 | 0.171 | 1 | |
| girls_sch_t+5 | 0.158 | 0.094 | 0.105 | 0.066 | 0.200 | 1 |

Table A.2: CORRELATIONS BETWEEN PROPORTION OF FEMALES WITHIN NEIGHBORHOODS IN DIFFERENT YEARS

| | girls_neig_t | girls_neig_t+1 | girls_neig_t+2 | girls_neig_t+3 | girls_neig_t+4 | girls_neig_t+5 |
|----------------|--------------|----------------|----------------|----------------|----------------|----------------|
| girls_neig_t | 1 | | | | | |
| girls_neig_t+1 | 0.081 | 1 | | | | |
| girls_neig_t+2 | 0.075 | 0.058 | 1 | | | |
| girls_neig_t+3 | 0.087 | 0.118 | 0.023 | 1 | | |
| girls_neig_t+4 | 0.098 | 0.080 | 0.133 | 0.028 | 1 | |
| girls_neig_t+5 | 0.275 | 0.007 | -0.097 | 0.097 | 0.122 | 1 |

Table A.3: TRANSITION OF THE PROPORTION OF FEMALES ACROSS QUINTILES THROUGHOUT THE SAMPLE YEARS

| | Quintile 1 | Quintile 2 | Quintile 3 | Quintile 4 | Quintile 5 |
|--|-------------|-------------|-------------|-------------|-------------|
| A. Summary Statistics for Schools | | | | | |
| Range | 0.076-0.509 | 0.509-0.550 | 0.550-0.587 | 0.587-0.632 | 0.632-0.956 |
| Mean | 0.462 | 0.531 | 0.569 | 0.608 | 0.684 |
| B. School Transitions Across Quintiles | | | | | |
| Quintile 1 | 18 | 413 | 350 | 359 | 428 |
| Quintile 2 | | 10 | 328 | 329 | 358 |
| Quintile 3 | | | 5 | 367 | 375 |
| Quintile 4 | | | | 10 | 381 |
| Quintile 5 | | | | | 43 |
| C. Summary Statistics for Neighborhoods | | | | | |
| Range | 0.200-0.532 | 0.532-0.561 | 0.561-0.581 | 0.581-0.607 | 0.607-0.917 |
| Mean | 0.496 | 0.548 | 0.571 | 0.649 | 0.649 |
| D. Neighborhood Transitions Across Quintiles | | | | | |
| Quintile 1 | 4 | 108 | 73 | 68 | 100 |
| Quintile 2 | | 0 | 60 | 61 | 79 |
| Quintile 3 | | | 0 | 61 | 66 |
| Quintile 4 | | | | 0 | 74 |
| Quintile 5 | | | | | 3 |

Notes: Elements on the diagonal report the number of schools that appear in the same quintile during the whole period. Total static schools: 86/1,097. Total switches: 3,688. Total possible ideal switches: 5,485. Total static neighborhoods: 7/233. Total switches: 750. Total possible ideal switches: 1,165. Elements on the diagonal report the number of schools and neighborhoods that appear in the same quintile during the whole period of interest. Elements on the off diagonals report the number of schools observed in both quintile x and quintile y. The sum of observations across cells in panel B is larger than the total number of schools in the sample, since schools can be observed in multiple quintiles.

Table A.4: SAMPLE STUDY AND POPULATION

| | Sample | Population | |
|-------------------------------------|-------------------|---------------------|-------------------|
| | (144schools) | (Remaining schools) | Difference |
| | Mean/s.d. | Mean/s.d. | b/s.e. |
| | (1) | (2) | (3) |
| Female | 0.570 (0.570) | 0.580 (0.570) | -0.010 (0.006) |
| Age | 17.912 (0.070) | 17.906 (0.128) | 0.006 (0.011) |
| Born in First Quarter of Birth Year | 0.158 (0.035) | 0.162 (0.036) | -0.004 (0.003) |

Note: We use a sample of 144 schools to construct drop out rates, and in this table we present descriptive statistics to show that this sample is representative in terms of gender composition and other predetermined characteristics. Column 1 presents the descriptive statistics for the sample of 144 schools and column 2 presents the same descriptive statistics for the remaining schools. Column 3 shows the differences between columns 1 and 2 and the related standard errors.

Table A.5: ROBUSTNESS CHECK: SIMULTANEOUSLY CONTROLLING FOR THE PERCENTAGE OF FEMALES IN THE TWO GEOGRAPHICAL UNITS

| | School | | | | Neighborhood | | | |
|---|---------------------|---------------------|---------------------|---------------------|----------------------|---------------------|---------------------|---------------------|
| | BOYS | | GIRLS | | BOYS | | GIRLS | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| National Exam Score | 2.873 (0.910)*** | 2.949 (0.976)*** | 3.433 (0.777)*** | 3.353 (0.844)*** | 2.908 (1.554)* | 2.364 (1.540) | 3.347 (1.311)** | 2.485 (1.285)* |
| <i>N</i> | 139,345 | 123,146 | 185,106 | 162,991 | 122,193 | 122,193 | 161,540 | 161,540 |
| Matriculation Status | 0.046 (0.021)** | 0.042 (0.023)* | 0.033 (0.017)** | 0.034 (0.018)* | -0.028 (0.030) | -0.029 (0.031) | 0.042 (0.025)* | 0.031 (0.025) |
| <i>N</i> | 139,345 | 123,146 | 185,106 | 162,991 | 122,193 | 122,193 | 161,540 | 161,540 |
| Matriculation Score (std.) | 0.143 (0.053)*** | 0.154 (0.056)*** | 0.211 (0.048)*** | 0.190 (0.052)*** | 0.342 (0.085)*** | 0.279 (0.085)*** | 0.217 (0.073)*** | 0.161 (0.072)** |
| <i>N</i> | 111,023 | 98,463 | 150,475 | 132,889 | 97,736 | 97,736 | 131,742 | 131,742 |
| Enroll in an Academic University | 0.056 (0.026)** | 0.067 (0.028)** | 0.084 (0.022)*** | 0.080 (0.024)*** | 0.156 (0.038)*** | 0.144 (0.039)*** | 0.104 (0.033)*** | 0.086 (0.033)*** |
| <i>N</i> | 111,023 | 98,463 | 150,475 | 132,889 | 97,736 | 97,736 | 131,742 | 131,742 |
| Quality of Postsecondary Degree | 4.315 (1.493)*** | 4.861 (1.577)*** | 6.929 (1.373)*** | 6.681 (1.487)*** | 10.872 (2.478)*** | 8.967 (2.481)*** | 7.395 (2.083)*** | 6.146 (2.064)*** |
| <i>N</i> | 111,022 | 98,462 | 150,476 | 132,889 | 97,736 | 97,736 | 131,742 | 131,742 |
| University Field Enrollment | | | | | | | | |
| Enroll in Science & Mathematics | 0.046 (0.014)*** | 0.049 (0.015)*** | 0.019 (0.011)* | 0.017 (0.011) | 0.029 (0.019) | 0.035 (0.020)* | 0.039 (0.014)*** | 0.035 (0.014)** |
| <i>N</i> | 111,022 | 98,462 | 150,476 | 132,889 | 97,736 | 97,736 | 131,742 | 131,742 |
| Enroll in STEM | 0.022 (0.026) | 0.036 (0.028) | 0.038 (0.018)** | 0.034 (0.019)* | 0.052 (0.036) | 0.048 (0.038) | 0.048 (0.025)* | 0.049 (0.026)* |
| <i>N</i> | 111,022 | 98,462 | 150,476 | 132,889 | 97,736 | 97,736 | 131,742 | 131,742 |
| Expected Occupation Wages (std.) | 0.002 (0.051) | 0.011 (0.054) | 0.123 (0.045)*** | 0.089 (0.048)* | 0.089 (0.069) | 0.041 (0.070) | 0.200 (0.065)*** | 0.182 (0.066)*** |
| <i>N</i> | 111,022 | 98,462 | 150,476 | 132,889 | 97,736 | 97,736 | 131,742 | 131,742 |
| Year FE. | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| School/Neighborhood FE. | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| School/Neighborhood-specific linear trends | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Student-level controls | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Student characteristics | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| School/Neighborhood-by-year characteristics | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Proportion of females in neighborhood | | ✓ | | ✓ | | | | |
| Proportion of females in school | | | | | | ✓ | | ✓ |
| No. of Schools/ Neighborhoods | | | 1,097 | | | | 222 | |

Note: Columns (1)-(4) present estimates from the school fixed-effects regressions, while columns (5)-(8) presents estimates from the neighborhood fixed-effects regressions. In columns (2), (4), (6) and (8), we control for the proportion of females in the other unit, respectively, while in columns (1), (3), (5) and (7) we present the main estimates for comparison purposes. All regressions control for student-level controls (prior performance and dummies for special interest in mathematics, science, or computer science), student predetermined characteristics (age, dummy for the quarter of birth), school/neighborhood-by-year characteristics (students' characteristics averaged by school/neighborhood and year, school/neighborhood annual enrollment) year dummies, school/neighborhood fixed effects and school/neighborhood-specific linear time trends. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively. Robust standard errors clustered at the school level are reported in parentheses.

Figure A.1:

INTERACTIONS BETWEEN NEIGHBORS: THE CASE OF A SCHOOL COMPLEX



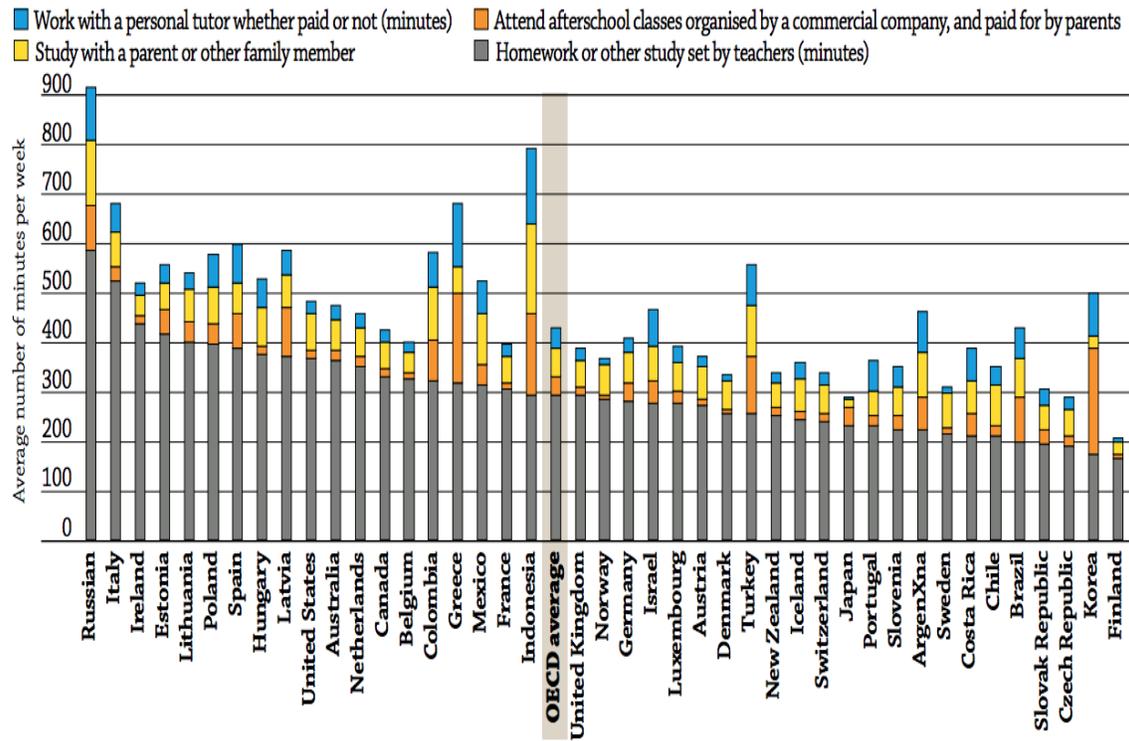
Figure A.2:

INTERACTIONS BETWEEN NEIGHBORS: THE CASE OF A SMALLER SCHOOL COMPLEX



Figure A.3:

AVERAGE TIME STUDENTS SPEND EACH WEEK ON AFTER-SCHOOL STUDY ACTIVITIES



Source: OECD (2013), PISA 2012 Results: What Makes Schools Successful (Volume IV): Resources, Policies and Practices, PISA, OECD Publishing, Paris.

Note: This figure shows the average time (in minutes per week) 15-year-old students spend each week on after-school study activities, with all school subjects combined, based on PISA data for 2012. Countries are ranked in descending order of the number of minutes per week that students spend doing homework or other study set by their teacher.