

Anelli, Massimo; Peri, Giovanni

**Working Paper**

## The Effects of High School Peers' Gender on College Major, College Performance and Income

CESifo Working Paper, No. 6014

**Provided in Cooperation with:**

Ifo Institute – Leibniz Institute for Economic Research at the University of Munich

*Suggested Citation:* Anelli, Massimo; Peri, Giovanni (2016) : The Effects of High School Peers' Gender on College Major, College Performance and Income, CESifo Working Paper, No. 6014, Center for Economic Studies and ifo Institute (CESifo), Munich

This Version is available at:

<https://hdl.handle.net/10419/145049>

**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.*



# Working Papers

[www.cesifo.org/wp](http://www.cesifo.org/wp)

## The Effects of High School Peers' Gender on College Major, College Performance and Income

Massimo Anelli  
Giovanni Peri

CESIFO WORKING PAPER NO. 6014

CATEGORY 4: LABOUR MARKETS

JULY 2016

*An electronic version of the paper may be downloaded*

- from the SSRN website: [www.SSRN.com](http://www.SSRN.com)
- from the RePEc website: [www.RePEc.org](http://www.RePEc.org)
- from the CESifo website: [www.CESifo-group.org/wp](http://www.CESifo-group.org/wp)

ISSN 2364-1428

CESifo

Center for Economic Studies & Ifo Institute

# The Effects of High School Peers' Gender on College Major, College Performance and Income

## Abstract

Using an originally constructed dataset that follows 30,000 Italian individuals from high school to the labor market, we analyze whether the gender composition of peers in high school affected their choice of college major, their academic performance and their labor market income. We exploit the within-school, cohort-by-cohort variation in the gender composition of high school classmates (peers), after controlling for school and teachers fixed effects. We find that male students graduating from classes with a large majority of male peers were more likely to choose “prevalently male” (PM) college majors (Economics, Business and Engineering). However, this impact was partially undone during college through attrition, worse academic performance and change in major. And in the long run it did not produce any difference in income or labor market outcomes. We do not find significant effects of the high school class gender composition on women. Our results are consistent with the fact that individuals are affected by the choice/pressure of the network of friends and with the observation that network size responds to class gender composition more for men than for women.

JEL-Codes: I210, J160, J240, J310, Z130.

Keywords: peer effects, high school, gender, networks, choice of college major, academic performance, wages.

*Massimo Anelli*  
*Bocconi University*  
*Via Roentgen 1*  
*Italy – 20136 Milan*  
*massimo.anelli@unibocconi.it*

*Giovanni Peri*  
*University of California, Davis*  
*One Shields Avenue*  
*USA – Davis, CA 95616*  
*gperi@ucdavis.edu*

July 18, 2016

We are grateful to Tito Boeri, Scott Carrell, Marta De Philippis, Hilary Hoynes, Marianne Page, Michele Pellizzari, Paolo Pin, Paola Profeta and Chiara Pronzato for useful suggestions. We also thank participants of seminars at Tilburg University, University of California Davis, the EEA Conference, at EALE conference and CESifo Summer Institute for helpful comments.

# 1 Introduction

Two regularities motivate this paper. First, while women have surpassed men in high school (Almås et al. 2016) and college graduation rates (Goldin 2006; Turner and Bowen 1999), their choices of college major are persistently different from those of men. Second, the peer environment in school can have important effects on the academic performance and on the choices of students in the short run (e.g. Carrell et al. 2009, Carrell and Hoekstra 2010). Combining these two facts, we ask the following two questions: do men (women) who attended high school classes with a large share of males (females) peers show higher likelihood to choose “typically male” (“typically female”) college majors right after high school (short run)? Does this effect on the choice of major have consequences on college performance, graduation and on labor market outcomes (long-run), or does it disappear over time? Does this effect differ for men and women? We address these questions using a large and newly collected individual administrative dataset and the variation in gender composition across high school classes in the city of Milano, Italy.

In many developed countries, men represent a significant majority of enrollment in Engineering and Business/Economics majors, which we will call “Prevalently Male (PM) Majors”. Instead, men represent a minority in the Humanities and Education majors, which we will call “Prevalently Female (PF) Majors”.<sup>1</sup>

These differences in the choice of college major accounts for a significant part of the existing male-female wage gap (Altonji et al. 2012) because the jobs usually available after graduation from PM majors pay more than those typically taken after graduating in PF majors. However, college majors are chosen based on preferences, expectations and comparative skills, and it is not clear that exogenously shifting people into PM majors would improve their labor market outcomes. The composition of peers in high school could be one of the factors affecting the choice of college major. Previous studies have shown that peer environment in school matters in affecting the performances, outcomes and choices of men and women<sup>2</sup>. For instance, some studies have shown that men tend to become more confident in competitive environments with other men. Women, instead, tend to become less confident under competition pressure with men (Bengtsson et al. 2005, Niederle and Vesterlund 2011). In a related research, Buser et al. (2014) finds significant evidence that women choose math-intensive and prestigious high school tracks less frequently than men in large part because they tend to shy away from highly competitive environments. However, this

---

<sup>1</sup>The OECD Education at a Glance shows that in the UK in 2012 only 20% of graduates in Engineering, but 62% of graduates in the Humanities, were women. The respective statistics for Italy are 37% and 77% and for the U.S. 21% and 58%.

<sup>2</sup>For instance Gneezy, Niederle and Rustichini (2003) Niederle and Vesterlund (2007) show different degrees of competitive behavior of men and women in same-gender environments, through an experimental setting. Booth and Nolen (2012a, 2012b) and Booth et al. (2014) show difference in performance of men and women in class with different gender composition.

evidence does not appear to apply also to college choices: Reuben et al. (2016) indeed find an impact of self confidence on expected income but not on the choice of college major.

Men and women tend also to choose a different structure in their groups of friends (which we will call network). Women prefer smaller and tighter networks and men larger and looser ones (e.g. Lindenlaub and Prummer 2013). As people, especially in their teen years, tend to have a larger percentage of own-gender friends in their networks, the number of own-gender classmates could increase the size of men's network more than the size of women's one (Stehl  et al 2013). This may imply stronger information and imitation effects on men when exposed to environments (classes) with a large number of other men. Building on and contributing to this literature, we analyze the effect of own-gender peers in high school on short-run outcomes, such as high school performance and choice of college major, and on long-run outcomes such as income.

This paper is among the very few to address these questions in a rigorous quantitative setting using administrative data. Previous studies (Solnick 1995, Billger 2002, and Billger 2009) have used the variation between single-gender and coed colleges to analyze the effect on the individual choice of field of study. These papers, lacking a source of exogenous assignment of students to colleges, had potentially important selection issues. More recently Booth et al (2014) and Schneeweis and Zweim ller (2012) have used random assignment or idiosyncratic variation of gender composition across cohorts to identify the impact of gender composition of peer on performance and choice of field of study. They however have only analyzed short-run (contemporaneous) outcomes. Our identification strategy builds on Hoxby (2000) and Schneeweis and Zweim ller (2012) in that it uses variation in class characteristics due to idiosyncratic cohort differences. However we can focus on both short-run outcomes (at the end of high school) and long-run outcomes (in college and in the labor market) and we can control much better for unobserved school and teacher characteristics.

There are two important challenges when setting up an appropriate research design for these questions. First, one needs to observe the gender composition of peers during high school and then follow individuals in their college choice, career and into the labor market. To this end, we collected in 2011 a database of 30,000 individuals who graduated from college preparatory public high schools in the municipality of Milan, Italy, between 1985 and 2005, and linked this information to their college career and to labor market outcomes. Second, to draw causal inference, one needs an exogenous source of variation in the gender composition of peers. We define as "peers" an individual's high school classmates and we exploit the idiosyncratic variation of the high school class gender composition across cohorts, within school and within teacher-group. By using a fixed-effect model that controls for school-cohort effects and teacher-group effects, we eliminate any variation across classes that could be due to observable and unobservable school-cohort characteristics and teacher-characteristics. We are left only with variation in the peer composition for

which we can establish absence of correlation with individual and other class characteristics.

The most relevant findings of our empirical analysis are as follows. First, we find that male students attending a high school class with 80% or more male classmates have a probability to choose a Prevalently Male (PM) college major between 6 and 15 percentage points higher than the average, which in our sample equals 43%. Second, the increase in probability of choosing PM Majors is particularly strong for male students in the lower part of the academic quality distribution. That group increased the frequency of PM major choice by 20 to 46 percentage points, from an average baseline of 25.5 percent. Third, these effects on the choice of PM majors faded away by the time of college graduation, because of attrition and change of major. Males from high school classes with more than 80% male peers were only marginally more likely to *graduate* from a PM major. They also took more time to graduate, they received a lower final score and were less likely to graduate. Finally, consistently with the idea that choice of major was reversed by the time of graduation, the increased likelihood of choosing PM majors did not translate into any significant effects on income, employment or occupation of male individuals who attended a high school class with more than 80% (or even 90%) of male peers.

Interestingly, we do not find effects of classmates' gender on women's choice nor on their academic performance, neither in the short nor in the long run. This is broadly consistent with some recent studies (e.g. Park et al 2013) that found same-gender environments in schools mainly affecting male students (albeit in terms of academic performance) in the short run<sup>3</sup>. An interesting recent paper (Stevenson, 2015) shows a similar difference in peer effects between men and women, even in juvenile jails girls are not influenced by a higher share of high risk female peers, while men are). A possible explanation for this differential effect that we explore at the end of the paper is that males are more likely to form and be affected by a larger network when a large number of their peers are male relative to women who prefer smaller networks and are insensitive to a large number of women in their class (Stehlé et al. 2013).

What we can learn from our analysis in the city of Milan is informative for other contexts too. First, in most European countries the choice of college major is made by students during their last year of high school. Estimating the impact of the gender composition and in general of class peers at the time of this choice is an important component to understanding the choice of college majors in European countries. Second, in the US, where the choice of college major is usually made during the first two years of college, these results can inform the consequences of gender composition of classes during the first two years of college. Separating or mixing students in classes according to gender is an easily controlled mechanism (policy) at the high school or college level and this article speaks to its effects in terms of college major choice.

---

<sup>3</sup>Jackson (2011), however, found that same sex schools in Trinidad and Tobago have a positive effect only on academic performance of girl and only on those with the strongest preference for same sex schools.

The rest of the paper is organized as follows. In section 2 we present our dataset and we illustrate more carefully the correlation of major choice with income and with earning gender gap. In section 3 we present our identification strategy, potential challenges and proposed solutions and we describe the empirical specification we use. In section 4 we show the main estimation results on the choice of major, on college performance and labor market outcomes. In section 5 we discuss and test two mechanisms that can explain the estimated effects. We conclude our analysis discussing implications of our findings in the final section 6.

## 2 Data Description

### 2.1 Construction of the Data Set

Our sample comprises individuals who graduated from all but one<sup>4</sup> of the public college-preparatory high schools (Licei) in the city of Milan, Italy, between 1985 and 2005<sup>5</sup>. These individuals are currently between the ages of 30 and 50. We gathered information from administrative records in the high schools by sending research assistant to collect and digitize them. While some missing and destroyed records prevented full coverage, we were able to cover more than 90% of all records for students who graduated from the thirteen different schools between 1985 and 2005. The sample includes about 30,000 individuals distributed into 1,371 high school classes. We only have information relative to their last (fifth) year of high school. Six of the high schools included are of the “Classical” type and seven are of the “Scientific” type.<sup>6</sup> Both tracks of high schools granted access to any college major in the considered period, but the first type has a curriculum more intensive in Humanities while the second type is more focused on Science and Math. Many college educated individuals from Milan became professionals in business, finance, administration, education and academia. Our analysis, therefore, pertains to a group in the upper tail of the income and educational distribution in Italy.

For these individuals we have information on the year of high school graduation, the score in the high school exit exam, the school attended, place of residence and the identity of their parents. Most importantly, we know the identity of their peers (classmates) in the last year of high school and the set of professors they shared.

We have linked these high school data (using name and date of birth) with records from all five universities in Milan (two private universities, Università Cattolica and Bocconi, and three public universities, Politecnico, University of Milano, University of Milano-Bicocca). Several of the best

---

<sup>4</sup>Only the Liceo Classico Carducci did not authorize us to collect Data.

<sup>5</sup>There were also a few private Licei in Milan. They enrolled a much smaller number of students usually among the low performing ones from wealthy families.

<sup>6</sup>Every school offers one track only

University Departments in Italy are located in Milan. The information about their university career includes whether they graduated, the year, major, university of graduation and their overall exit score. It is overwhelmingly common for high school graduates in Italy, and in general for many European college students, to attend the university, if one exists, in their city of origin. Therefore, the case analyzed is representative of large metropolitan areas in Europe.

Further, we linked these records with personal income figures in year 2005. This is the total income of each individual as reported to the tax authority. The advantage of using these data is that the administrative file of reported income includes all individuals in the nation; it is mandatory to report any income. Hence, if a person does not appear, he/she has no income or he/she is not living within Italy.<sup>7</sup> Self-employed are included in the sample. The limitation of these data is that we do not have a measure of labor supply (hours or weeks worked). Hence we will focus on yearly income. Finally, we linked the address where students lived at the time of high school to the average house value of their specific neighborhood<sup>8</sup>. We use this measure as a proxy for family wealth during high school, as the house is the most important financial asset of families in Italy.

Longitudinally linked individual data on income, university career, high school performance and family background are rare in any country. For Italy, our originally collected database is, to our knowledge, the only one that contains such information for such a large sample. Hence, this data provides an interesting tool to analyze the long-run effect of schooling on income. We use the income data only for people who finished high school between 1985 and 2000. Considering an average college attendance of four to five years, people in our sample would have been in the labor market between 0 and 15 years as of 2005. Hence, this provides a good assessment of the consequences of the last year of high school in the long-run (up to 20 years after high school graduation). Of the 30,000 individuals for which we have information about high school, 14,000 whom graduated between 1985 and 2000 were matched to the information on income in year 2005. For a stratified 10% random sub-sample of the initial sample (equal to around 3,000 individuals) we also collected more detailed information from telephone interviews conducted in June 2011 by the professional company “Carlo Erminero & Co.”. The interviews contain additional information covering family background, parental income, and extra-curricular activities during high school. We will use some of this additional data in robustness checks and extensions.

---

<sup>7</sup>There is also a small category of employees with only standard salary income and no deductions that need not report it. Usually this is a very small percentage of the population.

<sup>8</sup>We have transformed each address into geographic coordinates using Google’s Geocoding Service and then matched each address to market value per square meter as provided by the government agency “Agenzia del territorio” for 55 different homogenous areas in Milan.

## 2.2 Summary Statistics for Individuals and Classes

In our data we can identify individuals belonging to each fifth-year (senior) class in each section of each high school included in the sample. Students determined to pursue a college education in Italy usually choose one of two types of high schools, either “Classical” (Liceo Classico) or “Scientific” (Liceo Scientifico). Often they attend the public school of the chosen type closest to their residence.<sup>9</sup> Each public school admitted all students that applied. During the period under analysis, classes were formed in the first year of high school by pooling all enrolled students and randomly drawing the students in each class. The entry cohort in a school, therefore, varied from year to year, mirroring the demographics of the relevant age group living in proximity to the school. Students in the same fifth-year class shared the same peers (classmates) and professors during the fifth year of high school and, likely, for most of the previous years. The group of high school classmates is, for the majority of people, a very important group of peers at the time of college choice. Frequent interactions and the shared school experience is likely to affect the information available to students, as well as their preferences. Among students attending the “Classical” high school track, the average share of men was 33%, while for the “Scientific” track the average share of men was 60% male. This implies that the average gender composition of classes was more male-dominated in the scientific high schools. However, while classes in those schools were more likely to have large shares of men, the variation in gender composition of classes was large.

In Table 1 we present descriptive statistics for our data. We divide variables between individual-level (top portion of the Table) and class-level (bottom portion). For individual-level variables we present statistics relative to the whole sample as well as the separate means for men and women (in columns 6 and 7). We also show the t-statistic of the difference in averages (men-women). The table summarizes the information about individuals’ academic career in high school and in college.

We have re-scaled the high school exit test score<sup>10</sup> to be between 0 and 1, with 0 being the minimum passing score (60 out of 100 on the high school exit test) and 1 the maximum score achievable (100 out of 100). Summary statistics show the distribution of scores is left-skewed (mean score is 0.416) and the mean score for women is substantially higher than for men (the t-statistics of the difference in means is 11.4). Even when we rank students according to exit-scores within school and cohort, females perform better, being ranked on average at 0.516 (ranked from 0 to 1) versus an average rank of 0.471 for men with a t-statistic for the difference of 12.4.

Of these 29,370 high school graduates, 23,118 students (80%) enrolled in one of the universities in Milan. Of those 23,118, 27.5% enrolled in PM majors (Engineering, Economics & Business)

---

<sup>9</sup>In our data, 42% of all students attended the high school closest to their residence, independently of the high school track offered by the closest school. Conditional on choice of track, the percentage of students who attended the high school closest to their residence is 69%.

<sup>10</sup>At the end of high school, in the fifth grade, all students take an exit test, called “Maturità”. This test is prepared by the ministry of education and it is the same for all schools in the country and determines graduation.

with a substantial gender difference: only 13.9% of women chose one of the PM majors compared to 42.3% of men. The t-statistic for the difference in means is 50.3. Of the 23,118 students enrolling in a university, 17,140 actually graduated (by 2011, when we collected the data), implying a 26% attrition rate. Out of all students enrolling in a university, 11.8% of women earned a degree in a PM major compared to 32% of men. Interestingly, conditional on enrolling in a PM major, the drop out rate among women was lower than among men: 17.5% versus 25% with a t-statistic of the difference equal to 6.

To complete the list of college outcomes in our dataset, women took an average of 3 months less than men (t-statistic is 7) to graduate from college when the average time to completion was 6 years and 8 months<sup>11</sup>. At the end of college, every Italian student receives a final test-score out of 110 points, computed on the basis of G.P.A. and a final thesis<sup>12</sup>. We have re-scaled this final score to between 0 and 1 (with 0 being the minimum passing score). Consistent with other outcomes, women perform better on average than men (0.86 versus 0.78 with a t-statistic of 30.25). The averages for men and women suggest substantial differences in academic performance and academic choice.

The most important pre-treatment socioeconomic characteristic available to us (besides age, gender, and school) is the value of the home each student lived in during their last year of high school. This serves as a proxy for the family's income as a house was the primary asset held by families, and the value of real estate varies significantly within the city of Milan. In order to confirm that the house value is a reasonable proxy of the income of a family we consider the 10% sub-sample of individuals selected for the phone-survey, for which we know more detailed family background characteristics<sup>13</sup>. Figure 1 shows the correlation between the logarithm of the house price measure and the logarithm of average parental wages, constructed from the survey, by neighborhood. We observe a significantly positive correlation (OLS coefficient of 0.12 with standard error of 0.036 and  $R^2$  of 0.23) between household income and home value. In the remaining of the analysis, we will consider house price as a reasonable proxy for family wealth. Table 1 shows that the log of house value does not show any statistically significant difference between men and women. This demonstrates the average family background of female and male students included in our sample was similar.

For labor market outcomes, we observe the income for 17,004 students attending fifth year in high school by year 2000: since we have only 2005 income data available, we exclude individuals

---

<sup>11</sup>Italian students, especially during this period (1985-2005), had very long spells between college enrollment and graduation.

<sup>12</sup>The exception is engineering students who got a score out of 100. We re-scaled their scores accordingly.

<sup>13</sup>In particular we know from the survey occupation and sector of employment of both parents of the students at the time of high school. Using that information and ISTAT (2007) data we calculated average occupational-based parental wages by neighborhood.

that are still attending college by 2005, as their income might not be representative of their potential income (i.e. part-time job while in college). The matching rate in the relevant sample is 76%, which we consider very high. It means that 76% of the individuals graduated from high school by 2000 had a tax record in Italy (including a record of 0 income) in 2005. A large part of individuals without a matched tax returns record likely lived abroad in 2005 as early international mobility for work or study reasons of this sample was quite high.<sup>14</sup> The average log of income is 9.68 and the statistically significant difference (t-statistics is 22) across gender is 0.45 log points in favor of men. From the randomly selected 10% sub-sample that was interviewed, we also know that women had 30.8% probability to reach a top occupation (defined as manager, professional or self-employed) versus 43.2% of men with a t-statistic of 7.1. When comparing men and women earnings for 2005, there might be concerns that women might be more likely out of the labor market (e.g. for interrupted labor market history or family reasons). We thus regressed the probability of not having a matched tax returns record for 2005 on the dummy female and we found no correlation (the coefficient equals 0.005 with a standard error of 0.005).

As for class-level data we observe a total of 1,371 graduating classes with an average size of 21.4 students and a standard deviation of 3.8. The average share of female per class is 52% with a large standard deviation of 0.18. The average students' exit score by class ranges from 0.12 to 0.78, showing high variance in the ability composition of classes. Finally, the average socioeconomic status of classes is captured by the percentage of students in the class that live in houses valued in the bottom 10% of the house value distribution. The variance of this statistic is large; for instance, there were classes composed of only students coming from families in the bottom decile of the wealth distribution (they were in schools located in poor neighborhoods). The highest percentage of students in a class living in houses at the top 10% of the value distribution was 57%.

## **2.3 Prevalently Male Majors, Prevalently Female Majors and Income**

The main outcomes analyzed in this study are the choice of college major and the subsequent performance in college. In Italy college major is chosen at the end of high school and is costly to change later, as one must re-enroll and start the new major from the beginning. Hence, the major at enrollment is highly correlated with major at graduation.<sup>15</sup> It is also correlated with labor market outcomes. In our analysis we consider 11 college-major categories spanning all college degrees awarded in our sample. We organize those 11 majors into three groups associated with different shares of females among enrolled students. Figure 2 shows the female share among total enrolled students for the 11 college majors from our sample (measured in the right scale and reported as the

---

<sup>14</sup>Using the phone survey information available for the 10% interviewed random sub-sample, we calculated that almost 20% of all individuals without a tax return record responded to live abroad.

<sup>15</sup>56% of the students in our data graduated in the same major enrolled in right after high school graduation.

darker grey histogram). We rank majors from left to right according to this share and define the two majors with the smallest share of women (Engineering and Business/Economics) as constituting the Prevalently Male (PM) group. Those are the only two majors in which females constitute less than 40% of enrolled students. At the opposite end of the spectrum, Humanities and Education are the two majors showing the largest share of women among enrolled students; we call them the Prevalently Female (PF) group. These two are the only majors where women make up more than 70% of enrolled students. The remaining seven Majors with a more balanced gender composition are called the Gender Balanced (GB) group.<sup>16</sup>

Figure 2 also shows, as light grey bars, the average income in year 2005 by major of graduation (left scale), as measured in our sample (not corrected for any characteristics). Notice the PM majors are associated with the highest income while the PF majors are associated with lowest income and GB majors are in between. This suggests the different choice of major across gender may account for a significant part of the earning gap between men and women. To make the above correlations more formal, and to quantify them, in Table 2 we regress the logarithm of income in 2005 on a series of individual characteristics and on the Prevalently Male (PM) and Prevalently Female (PF) dummies (leaving Gender Balanced majors -GB- as the omitted category). The dummies equal one for individuals who graduated in the corresponding majors and zero otherwise. The controls include the high school exit score, the final college exit score, a dummy for living in a house in the top 10% of price distribution at the time of high school, one for living in a house in the bottom 10% and school-by-cohort fixed effects to control for cohort-specific and high school-specific factors.

The regression results, first shown separately for men and women, in Columns 1 and 2 of Table 2, reveal substantial positive income gaps associated with graduating from a PM major and significant negative income gaps associated with graduating from a PF majors for both men and women. Women with similar observable characteristics earn 0.65 logarithmic points more (92% more) if they graduated from a PM major relative to those graduating from a GB major (the omitted category). On the other hand, they earn 23% less if they graduated from a PF major (relative to GB). The estimate for males is equal to a 73% income premium (0.55 log points) from PM majors relative to GB and to a 42% penalty for PF majors graduates. The regression also reveals that college exit score has a strong correlation with high school exit score, while the family house value is much less correlated with it. Clearly no causal interpretation can be attached to these coefficients; they are simply a measure of the partial correlation between major and income for our sample.

Columns (3) and (4) of Table 2 show another interesting fact. They identify the role played by

---

<sup>16</sup>In robustness checks (available upon request) we alternatively considered the top and bottom three majors in terms of share of female, denoting them as PM and PF majors and perform similar analysis. The main findings presented in the paper remain unchanged.

college major in accounting for the income gap between men and women. In column (3) men and women are pooled and the coefficient on the female dummy is an estimate of the average income gap (in logarithmic points) between women and men (controlling for all the individual and family characteristics described above). Remarkably, this difference equals -0.4 logarithmic points, about 33%<sup>17</sup>, in favor of men, which is a very large average difference.<sup>18</sup>

In column (4) we simply add to the previous controls the PM and PF majors dummies. Besides being very significant, as expected, the introduction of these two dummies, by accounting for the major of graduation, reduces the men-women income gap from 0.4 to 0.25 logarithmic points (from 33% to 22%). Hence, one third of the male-female income gap is accounted for by their lower graduation rates from PM Majors and higher graduation rates from PF Majors. This is interesting, but the correlation shown above is far from establishing causality. One may be tempted to infer that pushing people into PM majors would increase their wage by the estimated amount. We analyze whether an increase in enrollment into PM majors, driven by exogenously distributed peer characteristics has any effect on the university performance and labor market income of individuals. In doing this we will discover the OLS coefficients of Table 2 is a misleading assessment of the effects a policy pushing individuals to enroll in PM majors would have on their income.

### 3 Identification Strategy and Empirical Specification

#### 3.1 Variation of Class Gender Composition

The vast majority of Italian students, from rich and poor families, attended public high schools. In Milan, the very good reputation of the public college-preparatory high schools that we consider implies that most youth who intended to go to college attended one of them. Hence, our sample is representative of individuals who intend to go to college. As we consider public schools, there was virtually no difference in the monetary cost of attendance.<sup>19</sup>

Within each high school, the entry cohort of students was randomly assigned to one “section” (coded with a letter: A, B, C...) in the first year. This assignment corresponded to a set of teachers that remained unchanged year after year for each section. Students remained in this letter-section until their fifth year, except for attrition and a few transfers (which are discouraged and bureaucratically cumbersome). While the official procedure to determine the composition of sections was

---

<sup>17</sup>The conversion from logarithmic points into percentage is always calculated as the exponential of the logarithmic points minus one.

<sup>18</sup>This measure of earning gap is not far from what is estimated for Italy in recent years. The gender gap in yearly wages estimated from the EU-SILC data (a representative household sample), and limited to college educated over 25 was 34% in 2009, while it was 40% for all workers.

<sup>19</sup>All public schools charge minimal fee per year. Currently it is \$150 per year. It was far less during the years 1985-2005.

random assignment of students, we cannot document directly its implementation (no records of the procedures were kept) and we do not observe the class-composition in the first year; only its composition in the fifth year. Ideally, the variation of female share in a class is only driven by the randomness in the student's section assignment in the first year of high school. We will discuss, however, potential threats to identification in the data available to us due to non randomness and selective attrition. We also test the independence of the female share from other class and students' characteristics.

There are three sources of variation in the gender composition of fifth-year high school classes in our data. The first is the gender composition of the school-cohort entering their first year. An entry cohort with a large share of female students would make the average first-year class in that school-cohort more likely to have a large share of women. Such shares were determined by demographics in the area of the school and by the choice of students between that college-preparatory and other high schools in the area. While such variation was outside the control of the school and its teachers, the choice of type of high school may differ between men and women. If correlated with some time-varying characteristics of the schools, which also affect the choice of college major, this could create an omitted variable bias. To avoid any possibility that unobserved school-cohort characteristics can cause omitted variable bias, we include school-cohort fixed effects in our regressions.

The second source of variation is the random assignment of students to "sections" within a school that partitions the entering cohort into several (usually 7-8) first-year "sections" (one class per section) with different female shares. This partition was random and assigned a student to section "X" in each school. The specific section had the same group of teacher every year. We want our identification strategy to capture this source of variation: the random variation year after year of women (men) in a section within a school. The third source of variation is generated by attrition, transfers and grade retention of students over the five years. Transfers were kept at a very minimum and rarely allowed. Attrition and retention mainly depended on average student quality and on teachers assigned to each section. Consider, for instance, a specific group of teachers (assigned to one section) with particular abilities, teaching methods and attitudes that may better accommodate students of one gender, at the same time inspire their choices or affect their outcomes. This, over time, could affect attrition differentially and could produce a correlation, for example, of male-friendly professors with higher shares of male students that pressures them towards a PM major. More in general, a key concern of our analysis is separating the effect of peers from that of teachers. Our data allow us to control for teacher-group fixed effects as we know that within a section they remained the same year after year. Teachers in the considered schools were usually tenured public servants with very high job stability, so changes in the body of teachers in a section over time were infrequent. Our inclusion of a "section by five-year intervals" (quinquennium) fixed

effect, therefore, is a conservative way of controlling for teacher fixed effects. Differential attrition due to teachers will be controlled by this effect. Therefore, the relevant source of variation for our identification strategy is the within-section-quinquennium (teacher-group), cohort-to-cohort variation in gender share of classes relative to the school-cohort average (controlled for by the set of fixed effects). This is likely due to different realizations of the initial random assignment of other idiosyncratic factors associated to attrition and retention. In any case we will also check the independence of class share of women (men) with student and class characteristics, conditional on the fixed effects.

This identification strategy is implemented empirically by estimating – separately for males and females – the following equation:

$$y_{i,l,s,t} = \lambda_{s,t} + \phi_{l,s,T} + \beta(CGC)_{i,l,s,t} + \delta X_{i,l,s,t} + \gamma Z_{l,s,t} + \varepsilon_{i,l,s,t} \quad (1)$$

The outcome  $y_{i,l,s,t}$  is relative to individual  $i$ , in “letter-section”  $l$ , in school  $s$ , in the fifth-grade class graduating in year  $t$  (which corresponds to cohort  $t$ ). In our main specifications, the outcome variable equals one if the individual chooses a Prevalently Male (PM) college major (or a PF major, or a GB major) and 0 otherwise. The term  $\lambda_{s,t}$  captures all the school by cohort fixed effects. The term  $\phi_{l,s,T}$  captures school-section-quinquennium effects ( $T = 5$ -year interval) which, as explained above, controls for the teacher-group. The term  $CGC_{i,l,s,t}$  represents the main explanatory variable of our analysis and hence  $\beta$  is the coefficient of interest. It captures a measure of the “Class Gender Composition” of individual  $i$  graduating in year (cohort)  $t$  from section  $l$  of school  $s$ . Most frequently in our analysis, that variable will be a dummy equal to one for classes with more than 80% (90%) classmates of the same sex as student  $i$  and 0 otherwise.

The variables  $X_{i,l,s,t}$  control for the pre-determined characteristics of individual  $i$  that we can observe, namely the measure of the student’s family house value as proxy for family wealth. We also control for observable class-level characteristics  $Z_{l,s,t}$ . They include a dummy for class size in the bottom 25% of the observed class size distribution and one for class size in top 25%, class geographical concentration measured as Herfindhal index of concentration of the students’ homes across city-blocks, the share of students in the class in the bottom 10% of house value and share of students in class in top 10% of house value. We first estimate specification (1) separately for men and women. We then consider it separately for students in different parts of the male-specific or female-specific ability distribution and in different parts of the wealth distribution, to see if peer gender has a different effect on the choices made by some specific subgroups.

We will also consider other outcomes  $y_{i,l,s,t}$  besides the choice of college major, exploiting our ability to follow students in their university career and in the labor market thanks to the longitudinal nature of the dataset. We consider several measures of performance in college, such as the graduation rate, time to graduation and the exit test score in college. Other outcomes occur decades

after the end of high school, such as realization of income on the labor market (log of income). In order to explore potential mechanisms taking place inside the high school classroom, we also consider proxies for network formation and other high school outcomes. In all the estimates, we cluster the standard errors at the school/cohort level<sup>20</sup>.

## 3.2 Independence of Class Gender Composition

Our identification strategy uses the residual variation of class gender shares, after we control for school-cohort and teacher group (“school-section-quinquennium”) fixed effects. It exploits the cohort-by-cohort variation of the gender share in classes assigned to the same group of teachers in a school, relative to the average gender share in the school-cohort. This identification strategy allows us to isolate the impact of gender share in the class, separating it from the potential effects of observable and unobservable characteristics of teachers and school-cohort. This identification strategy is more demanding than the identification used in other education contexts where there was no certainty of random assignment, such as Hoxby (2000), Carrell and Hoekstra (2010), Lavy and Schlosser (2011) and Schneeweis and Zweimüller (2012). Those studies rely solely on year-to-year variation in some demographic characteristics of entry cohorts.

To ensure that the residual variation of the gender share is as good as random and specifically independent from class and individual characteristics, we performed a series of tests. First, we show the residual female share used for identification is normally distributed. Moreover, we check that classes with extreme gender composition – i.e. with more than 80% or 90% male or female – that turn out to have an important role in the empirical analysis, occur with probabilities consistent with randomness. Figure 3, plots the histogram of the female class-share residuals from our data (solid bars) and a histogram from a simulated normal distribution with the same standard deviation as our sample (equal to 0.10) and the same number of observations (1,371) and shows this as empty bars with dashed borders. As one can see, there is very little departure from normality in the observed distribution of residuals. The tails of the simulated normal distribution (corresponding to the classes with extreme gender composition), in particular, has a similar number of observations compared to the actual distribution. Specifically, the observed distribution of residuals has 66 observations (classes) outside the interval of  $\pm 2$  standard deviations from 0, while the simulated distribution has 64 such observations. Several of these observations in the tails of the distribution are the “extreme composition classes” (having shares of males or female  $>80\%$ ), occurring at a frequency consistent with simple random variation<sup>21</sup>. In fact, their deviation from the mean (0) is larger than 2 standard deviations, but smaller than three. Only eight of the “extreme composition

<sup>20</sup>We have also clustered at the more conservative school level. Results are essentially identical.

<sup>21</sup>On average schools in our sample had 7.6 classes with more than 80% of students of the same gender and a standard deviation of 5.5 classes

classes” are outside the reasonable distribution range of the residuals; namely, outside the  $\pm 3$  standard deviations interval. Seven of those correspond to all-male and all-female classes (6 all-male and one all-female) in the same school (the Liceo Einstein) between 1986 and 1988<sup>22</sup>. As this may imply unusual gender composition of its classes, related to other features of the school in those years, we omit the whole school as a robustness check. Results are highly robust to this alternative specification, showing the effects are not driven by these outliers.

In Table 3, we show the correlation of different measures of the gender composition of a class with the observable predetermined characteristics of students in the class, controlling for the two sets of fixed effects (“school-cohort” and “teacher group”). The units of observations are classes and the dependent variables are described at the top of the Table for each specification. Specification (1) uses the share of women in the class as a dependent variable. Specifications (2) and (3) use a dummy for classes with a share of females (males) larger than 80% in the class (respectively). Columns (4) and (5) use a dummy for classes with a share larger than 90% for females and males in the class, respectively. Specifications (2)-(5) are ways of conducting a balancing test between the extreme gender composition classes and the remaining classes within a regression context. We check the correlations with several predetermined characteristics of the class as a way of testing the orthogonality of gender composition. We include as explanatory variables the average log price of the house where a student lived to measure family wealth), as well as indicators for the presence of significantly rich and poor families in the class. Specifically, for the share of students from families living in houses valued in the top 10% of all houses, and the share of students living in houses valued in the bottom 10%. We also include the average size of the class (and two dummies for whether it was in the top 25% or the bottom 25% of the size distribution) to see whether gender composition was associated with class size. Finally, we include measures of the average distance of students from school, as well as an Herfindhal index of concentration of students in the class in city-blocks. These measures check whether classes with a certain gender composition are also made of students living close to each other, which could suggest the clustering of friends in the same class, possibly as result of transfers after section assignment. The coefficients show that none of these variables – family location, wealth, class size or neighborhood ties between students – has a significant correlation with any of the gender composition measures (share of female, or prevalence of males/females in the class). Each individual variable has an extremely low and non-significant coefficient. The F-test of significance of all variables together rejects their joint significance at the 1% confidence level for all dependent variables and at the 5% level for all but the  $>90\%$  female classes. Overall, no predetermined characteristic of the class seems to be significantly correlated with the residuals of its gender composition.

Finally, we test the randomness of peer characteristics more stringently, conditional on school-

---

<sup>22</sup>Results are robust to replicating the estimation without these outlier classes in terms of gender composition.

cohort and teacher-group effects. In Table 4 we show the results of tests performed to verify that there is no correlation between *any* predetermined individual characteristics and the average characteristic of peers in the class (not just the gender ratio), after controlling for school-cohort and teacher-group effects. Following Guryan et al. (2009), in Table 4 we analyze the correlation between predetermined individual characteristics and the average of that predetermined characteristic among peers, once we control for the average characteristic of the cohort. We consider, as predetermined characteristics, the log value of the house, the distance from school, the probability of being in the top 10% or the bottom 10% of the house value distribution. We do not find any significant correlation between individual and peer characteristics in any case.<sup>23</sup> All our tests are consistent with the identifying assumption that individual student characteristics across classes within school-cohort and teacher-groups are not correlated with any observable class and peer characteristics, including the class gender composition. Reassured by these tests, we proceed under the assumption that, conditional on school-cohort and teacher-group, the distribution across classes of female (and male) shares is as good as random. We now move to estimate its causal impact on individual outcomes.

## 4 Main Results

We present our main results on different individual outcomes beginning right after high school. In the first subsection 4.1 we show how gender composition of the class affects the choice of college major at the end of high school. In subsection 4.2 we look at the effect of class gender composition on academic outcomes during college. In subsection 4.3 we look at the long-run effect of high school class gender composition on income after college.

### 4.1 Class Gender Composition and Choice of Major

Before analyzing the potential effects on college choice, we analyze whether the gender share in a class affected the academic performance of male or female students during high school. In this test we use a regression framework exactly as in (1) with the ranking of an individual’s exit test score relative to his/her school-cohort as the dependent variable  $y_{i,l,s,t}$ .<sup>24</sup> Table 5 reports the coefficients of different measures of class gender composition (*CGC*). Columns (1) to (4) report coefficients on female students’ school-cohort ranking position, while columns (5) to (8) report the coefficients

---

<sup>23</sup>The test by Guryan et al. (2009) is aimed at correcting for the fact that an individual cannot be assigned to himself: sampling of peers is done without replacement because the individual himself is removed from the “urn” from which his peers are chosen. This result in a positive bias when the average value of the “urn” is not included in the regression as control for testing the randomness. For the same argument the coefficients for the average value of the “urn” are expected to be negative and significant when the corrected test is performed.

<sup>24</sup>Rank was rescaled to be between 0 (worst score in the school-cohort group) and 1 (best score).

on male students' school-cohort ranking position. Within gender, the columns of the table differ in the specific explanatory variable used to capture the class gender composition. In specification (1) we include linearly the female share among classmates and in specification (5) we include the share of males among classmates. Then we analyze whether a variable capturing strongly unbalanced gender composition of the class, instead of the linear measure of shares, affects outcomes. As most studies in the literature consider same-gender groups vis-a-vis mixed-gender, in the following specifications we focus on the potential effects of "extreme" gender-composition of classes. In specification 2 (specification 6) the explanatory variable is a dummy equal to one for classes in which 80% or more of classmates are females (males). In specification 3 (specification 7) we consider classes in which 90% or more of classmates are females (males)<sup>25</sup>. Then in specification 3 (specification 8) we consider classes in which 10% or less of classmates are females (males). The estimated coefficients in Table 5 show the impact of gender composition on the individual position in the school-cohort ranking of exit test scores. The only coefficient significant at the 10 percent level is the one on the  $> 90\%$ -male classes for male test scores, and no coefficient is significant at 5%. It shows a small positive effect of  $> 90\%$ -male classes on male performance in the exit test score. We do not detect any significant effect of the classes that are  $> 80\%$ -males (or female) on men's (or women's) performance. This weak result, only in the  $> 90\%$ -male classes, suggests feeble evidence of any impact of peer gender on school performance.

Table 6 shows the main results when we use the choice of college major dummies as dependent variables and the same measures of class gender composition (*CGC*) as in the specifications shown in Table 5. In Panel A of the table the outcome variable is a dummy for enrolling in a Prevalently Male (PM) college major, in Panel B the outcome is enrolling in a Gender Balanced (GB) major, while in Panel C the outcome is enrolling in Prevalently Female (PF) college major. The full set of individual level controls, class level controls and fixed effects is included in all regressions, as noted in the lower part of the table.

First, our data show no evidence that the share of own-gender peers affects – in a linear way – the probability of choosing PM or PF majors for either males or females. The point estimates are small and non-significant (they are only positive for men). This emphasizes that in the presence of "small variations" in gender composition of classes around the average (as is the identifying variation available to some previous studies such as Schneeweis and Zweimüller, 2012), one may not identify any effect of peer gender composition on the choice of college major. In largely mixed-gender classes it may not matter much if own gender peers are 40 or 50% of the total. The second interesting result is that the major choice of women is not significantly affected even by the most extreme gender composition of the classes. Girls in  $> 80\%$  or  $> 90\%$ -female classes (specifica-

---

<sup>25</sup>For brevity in the rest of the paper we will call these classes as " $> 80\%$ -male" (-female) and " $> 90\%$ -male" (-female), respectively.

tions 2 and 3) do not exhibit any different propensity of enrolling in PF (or in PM) majors. The point estimates of those coefficients are small (usually close to 0.01) and never significant. Similarly, girls in classes with a large majority of males do not show any stronger propensity to enroll in PM majors (specification 4). Male students, however, show a significantly larger probability of enrolling in Prevalently Male majors if they have attended a  $> 80\%$ -male or, even more so, a  $> 90\%$ -male class (see specifications 6 and 7). The estimated effects are larger than for the rest of the table and significant at the 5 and 1% level. By attending a  $>80\%$  ( $>90\%$ ) -male class, a male student increases his probability of enrolling in PM majors by 6.3 (15.4) percentage points. These results are robust to using a broader definition of Prevalently Male Majors that includes Mathematics, Statistics, Computer Science and Physics (not reported in the Table). In that case the coefficient on the  $> 80\%$ -male dummy is 0.047 (s.e. 0.027) and the one on the  $> 90\%$ -male equals 0.128 (s.e 0.032).

In Table A1 in the appendix we also show a more systematic non-parametric approach to estimating the effect of classes with different shares of own-gender peers. We include nine different dummies capturing the effect of a share of same-sex peers (separately for men and women) between 10 and 20%, between 20 and 30%, between 30 and 40% and so on, omitting the reference category:  $<10\%$ . The table shows that for women, the point estimates on each dummy are small and insignificant. For men the point estimates of the effect on PM major enrollment are small, but progressively increase (with a few exceptions) from the lowest value for the dummy 10-20% males (3.3 percentage points) to the largest value for the 80%-90% males (10.8) and for the  $> 90\%$ -male (23.8) classes. Only the  $> 90\%$ -males dummy is significant at the 1% level in this non parametric approach of partitioning the effects, but the progression of coefficients clearly show the (non linear) increasing effect of the share of male peers on the probability that a men chooses a PM major.

The average probability of enrolling in PM majors is 42.7% for males; hence the effect described above is large and significant, increasing the probability of PM enrollment by between 1/7th and 1/3rd of its average. Interestingly, and somewhat symmetrically, males have a smaller probability of enrolling in PM majors if they attended a class that was  $>90\%$  female. However, this effect (equal to -7.3 percentage points and reported in column 8 of Table 6) is not statistically significant. Panels B and C of Table 6 reveal the shift into PM majors for males takes place by diverting students out of Gender Balanced (GB) majors. The probability of males enrolling in GB majors decreases by about 4.5 percentage points (not statistically significant) if they attended a  $> 80\%$ -male class and by 12.7 percentage points (statistically significant) if they attended a  $> 90\%$ -male class. This suggests that some males with “marginal” preferences for GB majors might have been pushed into PM majors when attending the last year of high school with a group of peers that were overwhelmingly male. The negative effect on the probability of enrolling in PF majors for males who attended  $> 80\%$ -male or  $> 90\%$ -male classes was weaker, but significant

for the  $> 80\%$ –male group.

Given the small effect on academic performance of males (zero for  $> 80\%$  male classes and significant but small for  $> 90\%$ -male classes) shown in Table 5, we do not think that the effect on major choice is due to improved academic ability of males. Rather, it may be indicative of the fact that peers and their gender can affect preferences and attitudes, especially of male youths<sup>26</sup>. As classes with  $> 80\%$  male and  $> 90\%$  males are rare, we also check that our results do not depend on any specific high school or university. We check that eliminating the high school “Liceo Einstein” (which included the only single-gender classes in the sample, and corresponded to excessively large deviations from randomness) estimates do not change much, and in fact the point estimate of the  $> 80\%$  male dummy increases to 0.27 (standard error 0.075). We then separately exclude the two universities that are specialized in the Prevalently Male majors, “Politecnico” for Engineering and “Bocconi” for Business/Economic and are also known for their high quality and selectivity.<sup>27</sup> Results (available upon request) show that is neither of these two universities that is driving our main results.

Table 7 analyzes the effects of class gender composition on the choice of major, partitioned by student academic performance. If the share of males in a class affects the size of network of males, students who are more uncertain on their academic career (possibly those not so strong academically or from less informed families) may be more affected by the information retrieved through a larger network.<sup>28</sup> We focus on the  $> 80\%$  and  $> 90\%$  own-sex share dummy as explanatory variables. We show, again, the main estimated effect of being in a class prevalently ( $> 80\%$  or  $> 90\%$ ) of the own gender (in columns 1 and 4). We then separate the effects by splitting the sample between individuals whose measure of academic performance, as revealed by their relative gender-specific ranking in the high school text score, was in the bottom quartile of the gender/school/cohort distribution (“bottom quality” in the table) and those whose performance was in the top quartile (“top quality”). Let us emphasize that while this measure of academic quality is not fully predetermined with respect to the class gender composition, by taking the gender-specific ranking, we avoid mechanical correlation with gender shares. We also showed in Table 5

---

<sup>26</sup>In a very demanding robustness check, we include only siblings and we include family fixed effects, school effects and the usual individual and class controls in the regression. In this case, the result is identified only by the different choices made by male siblings within a family. This reduces the sample from almost 10,000 to 1,300 and adds several hundreds fixed effects. The point estimate of the effect of the  $> 90\%$ -male classes is 14.3 percentage points, very close to the main estimate reported in Column (7) of Table 6. The standard error, however, is more than three times larger (0.163). The almost-unchanged point estimate for the  $> 90\%$  male classes and the extremely demanding nature of the check, convey the idea that classes with  $> 90\%$ -male shares can be associated with a shift of 13 – 14 points in the percent probability of attending a PM major by men.

<sup>27</sup>These specifications have to be interpreted carefully since we are selecting the sample on variables that can be considered as outcomes of our analysis.

<sup>28</sup>Lindenlaub and Prummer (2013) found that men acquire more information thanks to a larger network on average and as a consequence they perform better in more uncertain contexts.

that there is an extremely low correlation between individual test score ranking and class gender composition.

We show the effect of the  $> 80\%$ -own gender dummy (first row of each Panel) and the  $> 90\%$ -own gender dummy (second row of each Panel), individually, on the bottom and top quartile for females (column 2 and 3 respectively) and for males (columns 5 and 6). The effects for women are mainly non-significant. Only the effect of  $> 90\%$  female classes is significant for bottom-quality women and shows an increase in their probability of PM major choice. While the effect is not present for the  $> 80\%$  female classes, nor does it affect the probability of other major choices, the point estimates suggest a mild potential effect of heavily-female classes in encouraging PM major choice for low-performing women, too.

The positive average effect on the probability of enrolling in PM majors for men is much larger for the academically weakest group. For this group of male students, being in a  $> 80\%$  ( $> 90\%$ ) male class increases the probability of choosing a PM college major by 20 (46) percentage points. For high quality students, however, the effect is much smaller and statistically non-significant. The effect on low-academic-quality students is particularly remarkable because the average probability of choosing a PM major for bottom quartile male students is only 25.5 percent. Hence, graduating from a prevalently male class makes that probability two to three times larger. Panels B and C reveal that between half and three-quarters of the shift to PM majors for bottom-quality students is from GB majors, and between one-quarter and half is from PF majors. These estimates suggest that students in a prevalently male class who are not too strong academically, are subject to pressure/imitation from the network of male peers pushing them towards more demanding prevalently-male majors. Higher achieving students instead, who were already choosing PM majors with high probability – 57% on average) do not seem to be influenced by such class environments<sup>29</sup>.

Students from higher income families might have parents that are better informed and are more involved and influential in their children's decision of college major. This can act to reduce the impact of peer influence on College Major choice. Table 8 separates the effects on students from families in the bottom from those in the top quartile of the house value distribution, which could be considered a proxy for economic status. This variable is fully predetermined with respect to the class gender composition. Similar to what is done in Table 7, we show again the total effects for males and females (in Columns 1 and 4) from being in a high school class with  $> 80\%$ -own gender (first row of each panel) or  $> 90\%$ -own gender (second row of each panel) classmates, and for the sub-group of students in bottom and top quartiles of house value (column 2 and 3 for women and 5 and 6 for men). The effects on women's PM major choice are non-significant, even

---

<sup>29</sup>In Table A2 in the Appendix we split the male and female sample into above-below the median (rather than top-bottom quartile) of the academic quality distribution. Effects for the below-the-median subgroup are half of those for the bottom quartile in Table 7, confirming that most of the action is happening for the least proficient male students in classes with  $>80\%$  and  $>90\%$  male peers.

in the stratified specifications. For male students, those from families in the bottom-part of the house value distribution drive the overall results. For students with low economic status, a class that was  $>80\%$  ( $>90\%$ ) males was associated to a 14 (24) percentage points higher probability of choosing PM majors. Both estimates are significant at the 1% confidence level. Correspondingly, those males had a lower probability of enrolling in GB majors (between -7 and -21 percentage points) and lower probability of enrolling in PF majors (by -6.6 or -2.7 percentage points). To the contrary, male students from families in the top quartile of the house value distribution only show a weak increase in the probability of choosing a PM major when graduating from a  $> 80\%$  or  $> 90\%$  male class, (equal to 7 to 10 percentage points). Male students from less wealthy families may have had less parental involvement in the process of major choice, which resulted in stronger peer effects<sup>30</sup>.

## 4.2 Performance during College

In the universities included in our sample, PM majors (Engineering and Economics-Business) were more math intensive, more academically demanding and more selective than GB and PF majors. Using as a metric the high school exit test score (re-scaled to be between 0 and 1), which is common to all schools, students enrolled in the PM majors had an average of 0.51, those enrolled in PF majors averaged 0.42, and GB majors averaged 0.38. The differences are large and significant. Hence, if “prevalently male” high school classes pressured male students – particularly those with low academic quality – towards enrolling in those majors, this could imply the average college performance of males graduating from  $> 80\%$  and  $> 90\%$ -male classes was worse than those of other males. If some marginal students were pushed into majors that were “too hard” for their abilities this could lead to negative effects on their performance, and possibly to higher probability of dropping out or transferring to less demanding majors during college. De Giorgi, Pellizzari and Redaelli (2010) found that peer-pressure, or peer-imitation, in college increased mismatch in the choice of classes and majors, with negative impacts on their performance. We want to test whether such an effect is present in our data. More importantly, we test whether gender composition in the high school class affect college performance and whether its effect is still detectable by the time of college graduation. The effects of graduating from a  $> 80\%$  and  $> 90\%$  own-gender class on different college outcomes are shown in Table 9. In Panel A the dependent variable is the probability of having graduated from college by 2011. In Panel B we analyze the probability

---

<sup>30</sup>In the 10% subsample for which we surveyed families, we also stratified the effects by education of the mother (college or less than college). Even in this case we find stronger peer-effects on the probability of choosing PM majors among children of less-educated mothers. The coefficient estimated for the effect of  $> 90\%$  male peers is 0.73 for children of non-college educated mothers and it is significant. The standard error, however, is very large and does not allow the rejection of the hypothesis that this coefficient is the same as for children of non-college-educated mothers (point estimate 0.58).

of graduating in PM majors (which is different from the probability of enrolling that was the dependent variable in Tables 6-8). In Panel C the dependent variable is a dummy for dropping out of PM majors (conditional on ever enrolling in a PM major). In Panel D we analyze time-to-graduation (in months, controlling for major and university fixed effects to take into account average differences in time to degree across majors) for those who graduated, and in Panel E the outcome is the final graduation score in college (standardized to be between 0 and 1), controlling for time-to-graduation, major and university fixed effects to take into account the different grading styles across degrees and the fact that by postponing exams and graduation can lead to a higher final score. The regressions include the individual and class controls as the previous tables.

Looking first at the impact on female outcomes, out of 30 specifications only two have an effect significant at the 5% level. Recall that graduating from a prevalently female class did not have any significant impact on the choice of major, or on performance in high school for women, hence it is reasonable to expect it did not affect performance in college either. Considering males, whose choice of major at enrollment was significantly affected by their high school class gender composition, we notice graduating from a  $> 90\%$  male class is still associated with somewhat higher probability of graduating from PM majors (Panel B) and the effect is stronger for males in the bottom part of the academic quality distribution. These effects, however, are half the size of the effects estimated on enrollment (shown in Table 6) and they are only significant for males who graduated in a  $> 90\%$  male class, and for males in the bottom quality distribution from  $> 80\%$  male classes. These estimates suggest that male peers might have pressured marginal male students to enroll into PM majors, but many of them did not have the commitment or ability to graduate from them. Hence the effect on graduation was much smaller than the effect on enrollment. This possible mechanism is confirmed by the estimates of Panel A, in which males from  $> 90\%$  male high school class have a (marginally) lower probability of graduating from college altogether (at the 10% confidence level). Some of them may have dropped out, others may have abandoned the major and enrolled in a non-PM major. The signs of the coefficients of Panel C are consistent with this story, even though they are not precisely estimated: males from  $> 90\%$  male high school class enrolling in PM majors appear to be more likely to drop out (especially those in the bottom part of the academic quality distribution).<sup>31</sup> The graduation rate effect suggests a potentially higher mismatch into PM majors for the group of students graduating from  $>90\%$  male classes. Two other outcomes also appear to be negatively affected by having been in  $> 90\%$  male class: time-to-graduation (which was longer, as seen in Panel D) and final graduation score (which was lower, as seen in Panel E), but only for students in the lower quartile of the academic quality distribution. As for those from the  $> 80\%$ -male classes, the positive and significant effect on PM major enroll-

---

<sup>31</sup> Among all students, 28% initially enroll in a PM major (42% of men and 14% of women). Out of those 25% drop out of PM majors and do not graduate in those (27% of men and 19% of women).

ment (which was weaker than in the  $> 90\%$  male classes) completely faded away by the time of graduation and the other effects were not significant, implying the effect on major choice was only temporary and was reversed through transfers during the college years.

Having shown that high school peer gender was an exogenous event that affected male students' choice of college major we see here that such peer effects were not strong enough to change significantly the long-run match between students and majors. In fact, high school peer composition may have generated a higher probability of short-run mismatch driven by imitation/pressure in males. Those were especially strong for males in the lowest portion of the academic performance distribution. In the next section we analyze whether we can detect any long-run impact on income of individuals 1 to 15 years into the labor market. We focus only on individuals who graduated from  $> 90\%$  male classes for which we identified the strongest effects, still present in some outcomes at the end of college<sup>32</sup>.

### 4.3 Long-run Effects

By affecting their exposure to PM majors, on one hand, but decreasing the probability of graduating and their academic performance, on the other, the  $> 90\%$ -male classes could have affected the long-run individual performance of men. Those two intermediate outcomes affect income in opposite directions. The net effect could be positive by increasing the probability of exposure, at least for a while, to majors such as Engineering and Business that are associated with high-paying jobs. However, it may also be negative: by reducing college performance and probability of graduation as individuals are driven into mis-matched choices. In this section we tackle this question. In Table 10 we first analyze whether there was an effect of  $>90\%$  male classes on the probability of college enrollment (Panel A), and then on the probability of having nonzero income in 2005 (Panel B).

The data on income of the individuals are obtained from the Italian internal revenue service for fiscal year 2005 and matched to individuals based on their name, date and place of birth. We consider all individuals who graduated from high school between 1985 and 2000 and hence a group of people that was on the labor market between 1 and 15 years as of 2005. As long as they remained in Italy and earned any income, their record was collected by the internal revenue service and is present in our data. We analyze whether there was any impact on total (log) income by considering all individuals (Panel C), and then only those who enrolled in college (Panel D). Panels C and D capture the long-run effect on labor market outcomes. In Panel E we consider only those individuals who enrolled in college (comparable to the sample used in D) and we associate to the "major of enrollment" (not of graduation) the average income for that major in 2005 (estimated

---

<sup>32</sup>We performed the analysis also on individuals graduated from  $>80\%$ -own gender classes and all effects on post-college outcomes for males are non-significant.

using our sample). Such a variable allows us to isolate, in terms of income, the effect of higher enrollment in PM majors, zeroing out all other potential effects of the  $> 90\%$  male classes, which we have seen are present in the college performance. In this exercise, we assume that individuals in those classes graduated from the PM major of enrollment at the average rate (which we know is not true from Table 9) and went on to earn the average wage for that major. The dependent variable is the log of income in Panels C, D and E, therefore the coefficients represent a percentage change in income.

The results reported in Panels A-D do not show any significant coefficient on the  $>90\%$  male variable. The point estimates on the college enrollment and nonzero income are very small and non-significant. The point estimates on income (for all workers and for college educated only) are positive on average, but are very noisy and non-significant. Being in a  $>90\%$ -male class increased exposure of students to PM majors (that generate high paying expertise), but then worsened the outcomes of their college career (hurting earning potential). These opposing effects seem to cancel out. For women we also find no effect. Certainly observing only total income and only one point in time after graduation (2005) limits the depth of an analysis of labor market dynamics and outcomes and one should not overstate these results. However, we find no sign of a significant persistent effect of high school class gender composition.

Isolating, instead, the wage-effect of major of enrollment alone, reported in Panel E, male students in  $> 90\%$ -male classes had about 10% higher expected potential income. If we only observed enrollment in those majors for the male students (and not their performance, completion rates and graduation rates) we would be tempted to infer beneficial effects of  $> 90\%$  male high school classes on male labor market potentials. However, the analysis we are able to perform – because of our long-run longitudinal data – shows that the consequences of peer effects fade away in the long run. They are much less significant by the end of college and completely undetectable on the labor market. The peer effect on the choice of major of enrollment is significant the year after high school for males, but then the skills and abilities of individuals seem to be the more fundamental determinants of their major of graduation and academic performance.

## **5 Channels: Gender Differences in Network Formation**

A possible explanation for the finding of a strong and significant effect of gender composition on the choice of major for men but not for women is rooted in the difference in how they form their friendship networks. There is evidence in the literature that women form smaller and closer networks while men form larger, inclusive and activity-based ones (e.g. Lindenlaub and Prummer, 2013). As teenagers tend to form networks of own-gender friends, the number of own-gender high school classmates could increase the size of men's network more than the size of women's one

(Stehl  et al. 2013). Hence, it is likely that teenage boys in a class with a larger share of males end up with a higher number of friends, while teenage girls do not similarly expand their number of friends in classes with a larger share of females. Moreover, Lindenlaub and Prummer (2013) show how larger networks for males lead to more information-sharing and larger commonality of choice under uncertainty. This hypothesis is consistent with our results. Boys in high school classes with a large share of male classmates might end up with a larger network of peers with preferences for PM majors and be more influenced in their choice by this enlarged information set.

Although we do not have direct information on the size of friends' network in the classroom for the students under analysis, we have constructed an outcome variable that should be strongly related to it, allowing us to test this channel. This variable is the number of high school classmates of individual  $i$  who enrolled exactly in the same university major. This outcome reveals how many of the other students in a class made the same major choice as an individual. Its size can clearly be an indicator of how large is the reference group of individuals in the same friendship network who influence each-other's choice.

In Table 11-panel A we run our baseline specification 1 using this measure as dependent variable. First, notice that the average size of the network is larger for males than for females (last row of panel A), consistently with the findings of the network literature. While women have on average 2.6 classmates choosing the same major, men have about 2.8 of them. Also the effect of the class gender composition is fully consistent with our expectations. While a higher share of own-gender peers has no effect on the size of women's network (negative and not significant coefficient estimated in column 1), being in  $> 80\%$ -male classes significantly increases the size of men's network by 0.75 individuals. In  $> 90\%$ -male classes the effect is even larger (an increase of 2.4 peers). Moreover, consistently with an important role played by information-sharing in the network (as argued by Lindenlaub and Prummer, 2013) column 7 to 10 show that the proxy for network size increases more significantly for boys with low academic performance and from families of low socioeconomic status who attend male-dominated classes. Students from less wealthy families and with lower academic motivation are likely to be less informed regarding the choice of college and may thus experience stronger influence of peers in their own choice. In panel B, we estimate the effect of class gender composition on a more precise proxy of friendship network, represented by the number of individual  $i$ 's classmates who enrolled in the same major *and* university. Since most of the majors are offered by multiple universities in Milano, enrolling in the same major and in the same university is even more likely to be the outcome of belonging to a common friendship network.<sup>33</sup> Results for this second proxy of network formation are in panel B of Table 11 and confirm closely the effects found in panel A.

An alternative potential mechanism consistent with our results, relates to gender differences

---

<sup>33</sup>Each student in our sample has 26 possible university-major combinations to choose upon.

in confidence and competitive behavior (Almås et al. 2012). Some studies have shown that men and women are affected differently by an environment made of same-sex peers: while men tend to become more confident (sometimes overconfident) when competing with other men (Bengtsson et al. 2005) women do not. In order to test, albeit imperfectly, this alternative channel, we exploit information (collected in our phone survey) about non-academic student activity during the last year of high school, allowing us to introduce information on other aspects of student life. If the large majority of male peers boosted the confidence and competitiveness of boys, we should potentially observe effects on other outcomes outside the academic environment. A reasonable proxy for confidence and competitiveness is whether the student participated in competitive sport activities during high school. We do not find any significant effect on the probability of participating in sports, typically pursued outside the high school environment in Italy, and more common for men than for women. Following the same reasoning, we also analyzed whether volunteering for charities (an outcome that might be negatively correlated with competitiveness) was affected by the prevalently male environment of the class and again find no significant effect.

Overall, these exercises provide support for gender differences in network formation in a direction which is consistent with our results. The idea that class gender composition boosts the confidence and competitive behavior of men but not of women students, instead, does not find support in the data.

## 6 Conclusions

This paper has found a significant tendency of males in prevalently male high school classes to choose typically male (“PM”) college majors. This confirms the relevance of peers in affecting student choices and attitudes. However, thanks to the possibility of observing long-run outcomes, we also found that the increased pressure/influence to choose PM majors is largely reversed during college and has no measurable long-run effects on labor market outcomes. The results of our analysis constitute a cautionary tale on two important points that are very relevant when thinking of policies to increase the enrollment in math-intensive college majors (such as STEM majors) on the basis of their positive impact on jobs and careers. First, we show that while PM majors are associated with higher wages based on correlations alone, shifting people from non-PM to PM majors (as it happens for males in  $> 80\%$  male high school classes) does not produce the same result. Peer composition exogenously influences the choice of PM majors for males, but in the long run the mismatch produces a negative effect on college performance and no effect on wages. This confirms previous analysis that peer-pressure may result in mismatched choices (De Georgi et al. 2010). Second, the debate on the effects of peers in school has so far focused on academic outcomes and performances. We emphasize the importance of looking at long-run effects and

post-educational outcomes, as short-run effects on educational outcomes might quickly fade away in the long run. Third, while males appear to be influenced by the gender of their classmates, the choice of major of females remains unaffected by the gender composition of the class. To explore the mechanisms that could explain these differences we test two mechanisms that are known to be sources of gender differences in the social networks and behavioral literature. First, we explore whether male networks are more likely to expand in classes with a large majority of males relative to female networks. This would result in a larger friends' network for men in male-dominated classes, and stronger potential pressure for "typically male" behavior on boys. Second, we discuss whether attending a high school class with a large majority of male classmates boost confidence in boys that would otherwise shy away from competitive PM majors. While we find supporting evidence for gender differences in network size as determined by gender of classmates, we do not find supporting evidence for confidence effects of peer gender on male students. We hope this paper encourages similar data collections and research analysis in other countries and environments to validate and extend our analysis in different environments.

## References

- Almås, Ingvild and Cappelen, Alexander W. and Salvanes, Kjell G. and Sorensen, Erik and Tungodden, Bertil (2012) Willingness to Compete in a Gender Equal Society (December 17, 2012). NHH Dept. of Economics Discussion Paper No. 24/2012
- Almås, Ingvild, Alexander W. Cappelen, Kjell G. Salvanes, Erik Ø. Sørensen and Bertil Tungodden. 2016. "What Explains the Gender Gap in College Track Dropout? Experimental and Administrative Evidence." *American Economic Review*, 106(5): 296-302.
- Altonji, J.G., Blom, E. and Meghir, C. (2012), "Heterogeneity in Human Capital Investments: High School Curriculum, College Major, and Careers", *Annual Review of Economics*, vol. 4, no. 1, pp. 185-223.
- Bengtsson, C., Persson, M. and Willenhag, P. (2005), "Gender and Overconfidence", *Economics Letters*, vol. 86, no. 2, pp. 199-203.
- Billger, Sherrilyn M. (2002). "Admitting men into a women's college: A natural experiment." *Applied Economics Letters*, 9:7, 479-483
- Billger, Sherrilyn M. (2009). "On reconstructing school segregation: The efficacy and equity of single-sex schooling." *Economics of Education Review* 38(3), 393-402.
- Booth, A. L. and Nolen, P. (2012a), Gender differences in risk behaviour: does nurture matter?. *The Economic Journal*, 122: F56F78
- Booth, A. & Nolen, P. (2012b), "Choosing to Compete: How Different Are Girls and Boys?", *Journal of Economic Behavior and Organization*, vol. 81, no. 2, pp. 542-555.
- Booth, Alison L, Cardona Sosa, Lina & Nolen, Patrick, 2014. "Do Single-Sex Classes Affect Achievement? An Experiment in a Coeducational University", CEPR Discussion Papers 10221, C.E.P.R. Discussion Papers.
- Buser Thomas, Muriel Niederle and Hessel Oosterbeek (2014). "Gender, Competitiveness and Career Choices", *Quarterly Journal of Economics*, 129(3), 1409-1447.
- Carrell, S. E., Fullerton, R. L., and West, J. E. (2009). "Does your cohort matter? measuring peer effects in college achievement." *Journal of Labor Economics*, 27(3), 439-464.
- Carrell, S. E., and Hoekstra, M. L. (2010). "Externalities in the classroom: How children exposed to domestic violence affect everyone's kids." *American Economic Journal: Applied Economics*, 2(1), 211-228.

Science: How Professor Gender Perpetuates the Gender Gap”. *Quarterly Journal of Economics*, Volume 125, Issue 3, August 2010.

De Giorgi, Giacomo, Michele Pellizzari, and Silvia Redaelli (2010). “Identification of Social Interactions through Partially Overlapping Peer Groups.” *American Economic Journal: Applied Economics*, 2(2): 241-75.

Gneezy, Uri, Niederle, Muriel, Rustichini, Aldo (2003). “Performance in competitive environments: gender differences.” *Quarterly Journal of Economics*. 118, 1049-1074.

Goldin, Claudia. (2006) “The ‘Quiet Revolution’ that Transformed Women’s Employment, Education and Family.” *American Economic Review*, 96 (2): 1-21.

Guryan, Jonathan, Kory Kroft, and Matthew J. Notowidigdo (2009). “Peer Effects in the Workplace: Evidence from Random Groupings in Professional Golf Tournaments.” *American Economic Journal: Applied Economics*, 1(4): 34-68.

Hoxby, C. M. (2000). “The effects of class size and composition on student achievement: New evidence from natural population variation.” *Quarterly Journal of Economics*, 115(4), 1239-1285.

ISTAT (2007) ”Indagine sulle Forze di Lavoro”, Roma, Istituto Nazionale di Statistica, 2007.

Jackson, C. K. 2011. Single-sex schools, student achievement, and course selection: Evidence from rule-based student assignments in Trinidad and Tobago. *Journal of Public Economics* 96(1-2): 173-87

Lavy, Victor, and Analia Schlosser (2011). “Mechanisms and Impacts of Gender Peer Effects at School.” *American Economic Journal: Applied Economics*, 3(2): 1-33.

Lindenlaub, Ilse and Prummer, Anja (2013) “More versus Closer Friends: How Gender Shapes Social Networks and their Effects on Performance.” Mimeo

Niederle, Muriel, Vesterlund, Lise (2007). “Do women shy away from competition? Do men compete too much?” *Quarterly Journal of Economics* 122 (3), 1067-1101.

Niederle, M. and Vesterlund, L. (2011), “Gender and Competition”, *Annual Review of Economics*, vol. 3, no. 1, pp. 601-630.

Park, H., Behrman, J. R., & Choi, J. (2013). “Causal effects of single-sex schools on college entrance exams and college attendance: Random assignment in seoul high schools.” *Demography*, 50(2), 447-469.

Reuben E., M. Wiswall, B. Zafar (forthcoming) "Preferences and Biases in educational Choice: Shrinking the Black Box of gender", *The Economic Journal*

Schneeweis, N. & Zweimüller, M. (2012). "Girls, girls, girls: Gender composition and female school choice", *Economics of Education Review*, Volume 31, Issue 4, August 2012, Pages 482-500

Solnick, Sarah J. (1995). "Changes in women's majors from entrance to graduation at women's and coeducational colleges." *Industrial and Labor Relations Review* 48(3), 505-514.

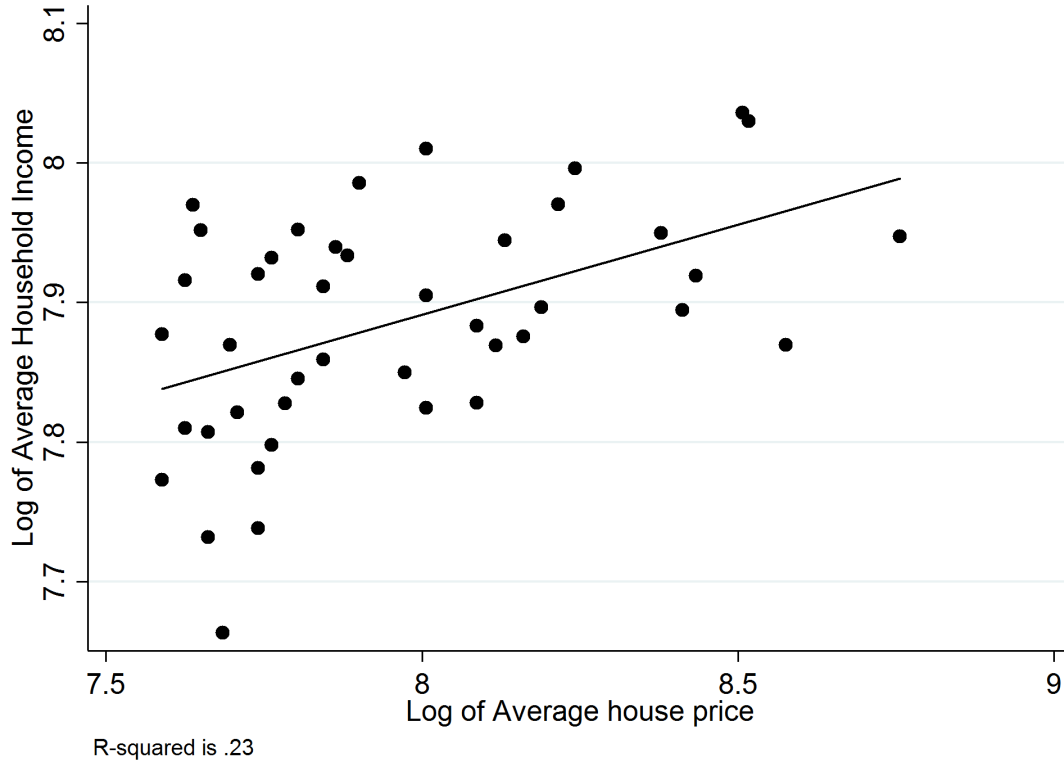
Stevenson Megan (2015) "Breaking Bad: Social Influence and the Path to Criminality in Juvenile Jails". Manuscript, UC Berkeley, April 2015.

Stehl, Juliette, Francois Charbonnier, Tristan Picard, Ciro Cattuto, Alain Barrat, "Gender homophily from spatial behavior in a primary school: A sociometric study" (2013), *Social Networks*, Volume 35, Issue 4, October 2013, Pages 604-613

Turner, Sarah E. and William G. Bowen (1999). "Choice of major: the changing (unchanging) gender gap." *Industrial and Labor Relations Review* 52(2), 289-313.

## Figures

Figure 1: Log of average household income vs Log of average house price by neighborhood



Note: Household Income is measured for high school students randomly selected for the phone survey as the sum of the expected parental income based on the occupation and sector of parents. We have computed the average parental income of parents for each neighborhood based on the address at which students in our data used to live at the time of high school. In this figure we plot the log average household income of each neighborhood vs the relative log average house price. Results of a simple regression of household income on house price by neighborhood shows a positive and significant relationship between these two variables and an  $R^2$  of 0.23

Figure 2: Annual Income and female shares by major

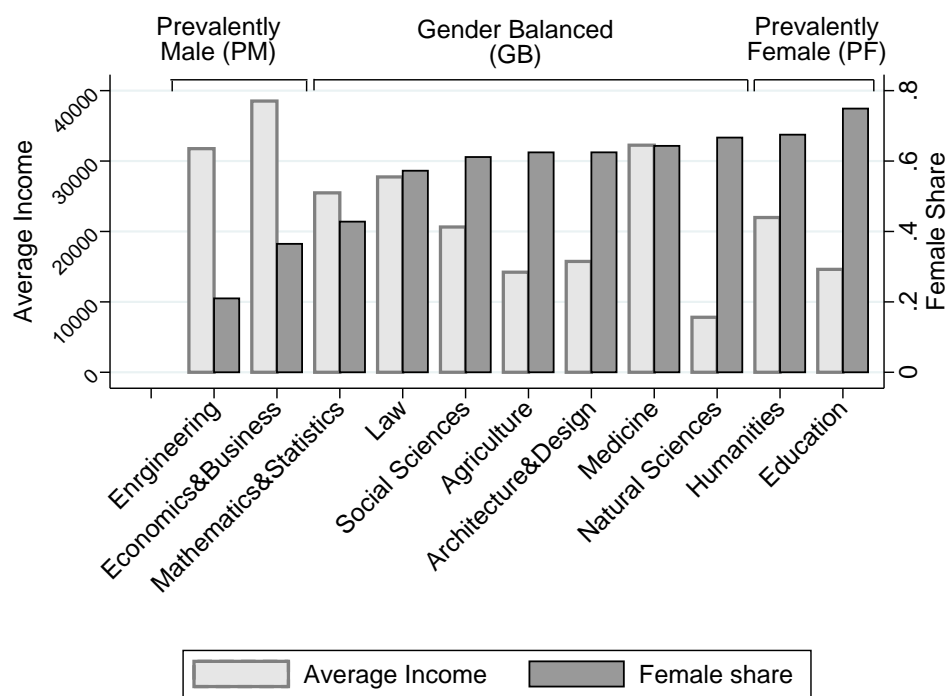
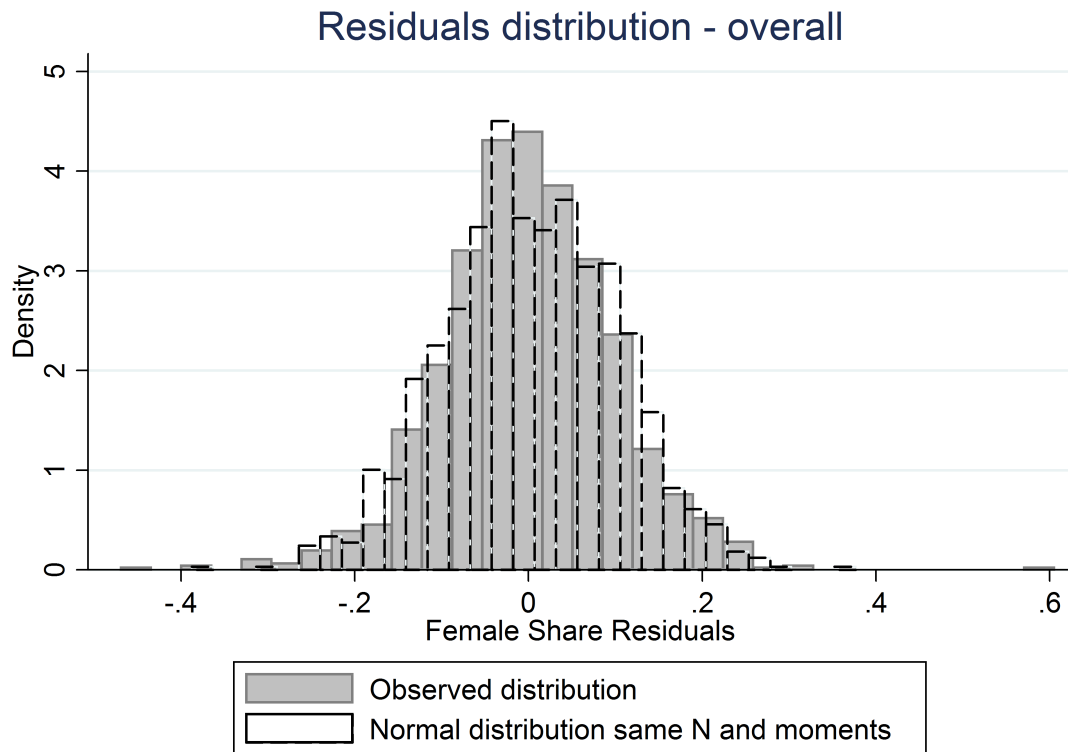


Figure 3: Residuals of female share after controlling for unobserved teachers and school/cohort effects



# Tables

Table 1: Summary Statistics

Variable	Obs	Mean	All Std.Dev.	Min	Max	Women Mean	Men Mean	Difference   T-stat
<b>Individual Variables:</b>								
Female	29370	0.523	0.499					
<b>High School Variables:</b>								
High school exit standardized score	29370	0.416	0.299	0	1	0.435	0.395	11.4
Within school/cohort rank	29370	0.495	0.313	0	1	0.516	0.471	12.4
<b>College Variables:</b>								
Enrolled in Prevalently Male majors	23118	0.275	0.447	0	1	0.139	0.423	50.3
Graduated in Prevalently Male majors	23118	0.215	0.411	0	1	0.118	0.320	39.0
Prevalently Male majors dropout	6362	0.229	.420	0	1	0.172	0.250	6.92
College time to graduation	16193	79.669	26.156	0	1	78.531	80.994	6.0
College exit score	16144	0.822	0.173	0	1	0.860	0.778	30.25
<b>Pre-treatment Variables:</b>								
Log(house value)	22365	7.992	0.298	7.409	9.143	7.986	7.997	2.2
<b>Outcome Variables:</b>								
Log(Income)	17004	9.679	1.279	0.693	13.746	9.461	9.906	22.0
Top Occupation	2957	0.367	0.482	0	1	0.308	0.432	7.10
<b>High School Class Variables:</b>								
Class size	1371	21.432	3.801	10	35			
Female share	1371	0.522	0.181	0	1			
Average high school exit score	1371	0.419	0.104	0.118	0.780			
Average log(House Value)	1219	7.982	0.161	7.606	8.477			
% students in bottom decile of house value distribution	1219	0.094	0.105	0	1			
% students in top decile of house value distribution	1219	0.082	0.107	0	0.57			

Table 2: Major Choice and Income

VARIABLES	(1) Females Log(income)	(2) Males Log(income)	(3) Log(income)	(4) Log(income)
(Female=1)			-0.401*** (0.031)	-0.251*** (0.031)
(Graduated in PM majors=1)	0.647*** (0.043)	0.549*** (0.043)		0.578*** (0.031)
(Graduated in PF majors=1)	-0.260*** (0.043)	-0.541*** (0.068)		-0.335*** (0.037)
H.S. exit score	0.161** (0.070)	-0.078 (0.063)	0.391*** (0.042)	0.036 (0.044)
College exit score	0.633*** (0.125)	0.890*** (0.111)	0.039 (0.073)	0.781*** (0.084)
(House value in top 10%=1)	0.095 (0.061)	0.026 (0.065)	0.089* (0.048)	0.068 (0.047)
(House value in bottom 10%=1)	-0.047 (0.058)	-0.062 (0.051)	-0.090** (0.043)	-0.052 (0.040)
(Commuting into city=1)	0.009 (0.060)	0.086 (0.065)	0.044 (0.046)	0.041 (0.044)
Constant	9.411*** (0.103)	9.812*** (0.091)	10.304*** (0.060)	9.703*** (0.069)
Observations	4,768	4,438	9,206	9,206
R-squared	0.233	0.311	0.224	0.269
School X Cohort FE	X	X	X	X

**Method:** OLS, Standard errors clustered at school/cohort level in parenthesis, \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

**Sample:** Students graduated from high school between 1985 and 2000 and completing college.

**Specifications:** (1) Female students only. (2) Male students only. (3),(4) both male and female students.

**Dependent variable:** logarithm of personal income, as revealed to the internal revenue service in year 2005.

**Independent variables:** For all specifications School/Cohort fixed effects, high school exit score re-scaled between 0 and 1, college exit score (composite of G.P.A. and a score for dissertation) re-scaled between 0 and 1, dummy=1 if student used to live in a house in top (and another dummy for bottom) 10% of house value distribution, dummy=1 if student used to commute from outside the city. For specifications (3) and (4) a dummy=1 if student is female.

**Definitions:** Following statistics computed on our sample we define Prevalently Female (PF) Majors to be Humanities and Education, Prevalently Male (PM) Majors to be Engineering, Economics & Business. The omitted variable is Gender Balanced (GB) Majors which includes all residual fields of study.

Table 3: Female share and Classroom characteristics

VARIABLES	(1) Female share	(2) Female >80%	(3) Male >80%	(4) Female >90%	(5) Male >90%
Class mean house value	-0.090 (0.124)	-0.201 (0.262)	0.104 (0.119)	-0.032 (0.053)	-0.012 (0.016)
Class share top 10% house value	0.046 (0.101)	-0.035 (0.107)	0.039 (0.097)	0.026 (0.039)	0.032 (0.033)
Class share bot 10% house value	0.056 (0.062)	0.216 (0.157)	-0.124 (0.123)	0.102 (0.074)	0.030 (0.036)
Class size	0.000 (0.004)	-0.009 (0.008)	-0.005 (0.006)	-0.001 (0.004)	-0.001 (0.001)
Class Size in Bottom 25%	0.006 (0.022)	-0.009 (0.041)	-0.013 (0.049)	-0.012 (0.023)	-0.023 (0.020)
Class Size in Top 25%	-0.015 (0.025)	0.036 (0.031)	0.014 (0.021)	-0.007 (0.013)	-0.001 (0.004)
Class mean distance from school	-0.003 (0.006)	-0.000 (0.016)	-0.003 (0.006)	-0.008 (0.006)	-0.002 (0.003)
Class geo. concentration	0.040 (0.063)	0.066 (0.054)	-0.181 (0.163)	0.007 (0.062)	-0.016 (0.058)
Observations	1,219	1,219	1,219	1,219	1,219
R-squared	0.721	0.424	0.389	0.417	0.383
School X Cohort FE	X	X	X	X	X
School X Teachers X 5yrs FE	X	X	X	X	X
Joint F-test p> F	0.19	0.50	0.27	0.02	0.92

**Method:** OLS, Standard errors clustered at school level in parenthesis, \*\*\* p<0.01, \*\* p<0.05, \* p<0.1.

**Sample:** Each observation corresponds to one class.

**Dependent variable:** Class female share for specification (1). In specifications (2) and (3) a dummy=1 if classes respectively have more than 80% females and more than 80% males. In specifications (4) and (5) a dummy=1 if classes respectively have more than 90% females and more than 90% males.

**Independent variables:** Class average logarithm of market value of the house where the students used to live at the time they attended high school, class share of students who used to live in a house valued in top decile of the house value distribution, class share of students who used to live in a house valued in top decile of the house value distribution, class size, a dummy for class size in the bottom 25% of observed class size distribution and one for class size in top 25%, class mean linear distance from school, class geographical concentration (measured as Herfindhal index of concentration in city-blocks). School/cohort fixed effects, group of teachers fixed effects every five years are also included (i.e. school-section fixed effects every five years).

Table 4: Class assignment and pre-determined characteristics

VARIABLES	(1) Log house value	(2) Distance from school	(3) Prob house top 10%==1	(4) Prob house bottom 10%==1
Peers' mean log house value	0.014 (0.017)			
Cohort mean house value	-101.873*** (10.657)			
Peers' mean distance from school		0.008 (0.016)		
Cohort mean dist from school		-119.195*** (8.171)		
Class share in top 10% of house values			0.002 (0.016)	
Cohort share in top 10% of house values			-127.612*** (15.516)	
Class share in bottom 10% of house values				-0.052 (0.045)
Cohort share in bottom 10% of house values				-110.004*** (10.554)
Constant	822.292*** (85.202)	315.823*** (21.461)	4.575*** (0.552)	12.282*** (1.169)
Observations	26,184	25,945	25,945	25,945
R-squared	0.830	0.839	0.849	0.786
School X Cohort FE	X	X	X	X
School X Group of Teachers X 5years FE	X	X	X	X

**Method:** OLS, Standard errors clustered at school/cohort level in parenthesis, \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

**Sample:** High school students graduating between 1985 and 2005.

**Dependent variable:** for specification (1) the logarithm of market value of the house where the students used to live at the time they attended high school, for specification (2) linear distance from school to the house where the students used to live at the time they attended high school in meters, for specification (3) a dummy = 1 if the student used to live in a house valued in top decile of the house value distribution, for specification (4) a dummy = 1 if the student used to live in a house valued in top decile of the house value distribution.

**Independent variables:** for specification (1) class peers' and school/cohort peers' average logarithm of market value of the house where the students used to live at the time they attended high school, for specification (2) class peers' and school/cohort peers' linear distance from school to the house where the students used to live at the time they attended high school in meters, for specification (3) shares of class peers and school/cohort peers who used to live in a house valued in top decile of the house value distribution, for specification (4) shares of class peers and school/cohort peers who used to live in a house valued in top decile of the house value distribution. In all specifications school/cohort fixed effects, Group of teachers fixed effects every five year.

Table 5: The effect of high school class gender composition on school/cohort ranking

	Females				Males			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treatment Variables →	Own Gender Share	Own Gender Share >80%	Own Gender Share >90%	Own Gender Share <10%	Own Gender Share	Own Gender Share >80%	Own Gender Share >90%	Own Gender Share <10%
Dependent Variable ↓								
School/Cohort Rank	-0.012 (0.031)	-0.013 (0.014)	-0.030 (0.027)	-0.057 (0.059)	0.021 (0.031)	0.003 (0.023)	0.048* (0.027)	0.000 (0.041)
Observations	13,412	13,412	13,412	13,412	12,533	12,533	12,533	12,533
Dep. var. mean	0.517	0.517	0.517	0.517	0.472	0.472	0.472	0.472
<i>Individual Controls:</i>								
Fam. wealth proxy	X	X	X	X	X	X	X	X
<i>Class-level controls:</i>								
Size	X	X	X	X	X	X	X	X
Geo. Concentration	X	X	X	X	X	X	X	X
Fam. wealth proxy	X	X	X	X	X	X	X	X
SchoolXCohort FE	X	X	X	X	X	X	X	X
SXTeachersX5yrs FE	X	X	X	X	X	X	X	X

**Method:** Each coefficient corresponds to one specification. OLS, Standard errors clustered at school/cohort level in parenthesis, \*\*\* p<0.01, \*\* p<0.05, \* p<0.1. Specifications (1)-(4) on women only. (5)-(8) on men only.

**Sample:** All students completing college-prep high schools in Milan between 1985 and 2005.

**Dependent variable:** Within School/Cohort rank in the high school exit score. Rank was re-scaled to be between 0 (worst score) and 1 (best score)

**Treatment:** In spec. (1) and (5) share of own-gender classmates (i.e. for spec. 1 female share and for 4 male share), in (2) and (6) a dummy=1 if class has more than 80% classmates of the own-gender, in (3) and (7) a dummy=1 if class has more than 90% classmates of the own-gender, in (4) and (8) a dummy=1 if class has less than 10% classmates of the own-gender.

**Individual Controls:** dummy=1 if student used to live in a house in top (and another dummy for bottom) 10% of house value distribution, dummy=1 if student used to commute from outside the city.

**Class-level controls:** dummy=1 if class size in the bottom 25% and one for class size in top 25%, class geographical concentration (measured as Herfindhal index of concentration in city-blocks), Class share in bottom 10% of house value, Class share in top 10% of house value, School/cohort fixed effects, Group of teachers fixed effects every five years (i.e. school-section fixed effects every five years).

Table 6: The effect of high-school class gender composition on students' major choice

		Females				Males			
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treatment	→	Own Gender Share	Own Gender Share >80%	Own Gender Share >90%	Own Gender Share <10%	Own Gender Share	Own Gender Share >80%	Own Gender Share >90%	Own Gender Share <10%
Dependent Variable ↓									
Panel A: Prevalently Male (PM) major choice at first college enrollment=1									
(PM Major=1)		-0.002 (0.032)	0.013 (0.011)	0.010 (0.037)	-0.056 (0.100)	0.030 (0.044)	0.063** (0.031)	0.154*** (0.043)	-0.073 (0.061)
Dep. var. mean		0.139	0.139	0.139	0.139	0.427	0.427	0.427	0.427
Panel B: Gender Balanced (GB) major choice at first college enrollment=1									
(GB Major=1)		0.062 (0.045)	0.005 (0.022)	-0.024 (0.039)	0.053 (0.164)	-0.023 (0.050)	-0.045 (0.035)	-0.127*** (0.038)	0.016 (0.072)
Dep. var. mean		0.586	0.586	0.586	0.586	0.460	0.460	0.460	0.460
Panel C: Prevalently Female (PF) major choice at first college enrollment=1									
(PF Major=1)		-0.060 (0.041)	-0.018 (0.020)	0.015 (0.053)	0.002 (0.112)	-0.007 (0.028)	-0.018* (0.011)	-0.027 (0.016)	0.057 (0.072)
Dep. var. mean		0.275	0.275	0.275	0.275	0.113	0.113	0.113	0.113
Observations		10,318	10,318	10,318	10,318	9,773	9,773	9,773	9,773
Individual Controls:									
Fam. wealth proxy		X	X	X	X	X	X	X	X
Class-level controls:									
Size		X	X	X	X	X	X	X	X
Geo. Concentration		X	X	X	X	X	X	X	X
Fam. wealth proxy		X	X	X	X	X	X	X	X
SchoolXCohort FE		X	X	X	X	X	X	X	X
TeachersX5yrs FE		X	X	X	X	X	X	X	X

**Method:** Each coefficient corresponds to one specification. OLS, Standard errors clustered at school/cohort level in parenthesis, \*\*\* p<0.01, \*\* p<0.05, \* p<0.1. Specifications (1)-(4) on women only. (5)-(8) on men only.

**Sample:** high school graduates enrolling in college.

**Dependent variables:** In panel A dummy=1 if student enrolled in Engineering, Economics and Business at first college enrollment. In panel B a dummy=1 if student enrolled in Natural Sciences, Mathematics, Statistics, Computer Sciences, Medicine, Agriculture, Architecture, Design, Social Sciences or Law. In panel C a dummy=1 if student enrolled in Humanities or Education.

**Treatment:** In spec. (1) and (5) share of own-gender classmates (i.e. for spec. 1 female share and for 4 male share), in (2) and (6) a dummy=1 if class has more than 80% classmates of own gender, in (3) and (7) a dummy=1 if class has more than 90% classmates of own gender, in (4) and (8) a dummy=1 if class has less than 10% classmates of own gender.

**Individual Controls:** dummy=1 if student used to live in a house in top (and another dummy for bottom) 10% of house value distribution, dummy=1 if student used to commute from outside the city.

**Class-level controls:** dummy=1 if class size in the bottom 25% and one for class size in top 25%, class geographical concentration (measured as Herfindhal index of concentration in city-blocks), Class share in bottom 10% of house value, Class share in top 10% of house value, School/cohort fixed effects, Group of teachers fixed effects every five years (i.e. school-section fixed effects every five years).

Table 7: The effect of high-school class gender composition on major choice by academic quality

Treatment↓	Sample→	Females			Males		
		(1)	(2)	(3)	(4)	(5)	(6)
		All	Bot Qual	Top Qual	All	Bot Qual	Top Qual
Panel A - Dep. Var.: (PM major=1)							
Own Gender Share>80%		0.013 (0.011)	0.021 (0.018)	-0.010 (0.031)	0.063** (0.031)	0.201** (0.095)	-0.051 (0.040)
Own Gender Share>90%		0.010 (0.037)	0.076** (0.037)	0.040 (0.056)	0.154*** (0.043)	0.461*** (0.142)	0.132 (0.081)
Dep. var. mean		0.139	0.057	0.225	0.427	0.255	0.571
Panel B - Dep. Var.: (GB major=1)							
Own Gender Share>80%		0.005 (0.022)	-0.034 (0.060)	0.035 (0.048)	-0.045 (0.035)	-0.095 (0.096)	0.022 (0.038)
Own Gender Share>90%		-0.024 (0.039)	-0.090 (0.077)	-0.048 (0.102)	-0.127*** (0.038)	-0.320** (0.157)	-0.161** (0.071)
Dep. var. mean		0.586	0.644	0.515	0.460	0.620	0.327
Panel C - Dep. Var.: (PF major=1)							
Own Gender Share>80%		-0.018 (0.020)	0.013 (0.057)	-0.026 (0.043)	-0.018* (0.011)	-0.106*** (0.035)	0.029 (0.023)
Own Gender Share>90%		0.015 (0.053)	0.014 (0.094)	0.008 (0.092)	-0.027 (0.016)	-0.141*** (0.053)	0.029 (0.036)
Dep. var. mean		0.275	0.299	0.260	0.113	0.125	0.102
Observations		10,318	2,100	3,075	9,773	1,973	2,864
Individual Controls:							
Fam. wealth proxy		X	X	X	X	X	X
Class-level controls:							
Size		X	X	X	X	X	X
Geo. Concentration		X	X	X	X	X	X
Fam. wealth proxy		X	X	X	X	X	X
SchoolX Cohort FE		X	X	X	X	X	X
SX TeachersX 5yrs FE		X	X	X	X	X	X

**Method:** Each coefficient corresponds to one specification. OLS, Standard errors clustered at school/cohort level in parenthesis, \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ . Specification (1) on women only. Specification (2) on women in the bottom quartile of the within school/cohort/gender rank in high school exit score, (3) women in the top quartile of the same rank. Specification (4) on men only. Specification (5) on men in the bottom quartile of the within school/cohort/gender rank in high school exit score, (6) men in the top quartile of the same rank.

**Sample:** High school graduates enrolling in college.

**Dependent variables:** In panel A dummy=1 if student enrolled in Engineering, Economics and Business at first college enrollment. In panel B a dummy=1 if student enrolled in Natural Sciences, Mathematics, Statistics, Computer Sciences, Medicine, Agriculture, Architecture, Design, Social Sciences or Law. In panel C a dummy=1 if student enrolled in Humanities or Education.

**Treatment:** In first row specifications of each panel dummy for classes with more than 80% of own-gender classmates. In second row of each panel dummy for classes with more than 90% of own-gender classmates.

**Individual Controls:** dummy=1 if student used to live in a house in top (and another dummy for bottom) 10% of house value distribution, dummy=1 if student used to commute from outside the city.

**Class-level controls:** dummy=1 if class size in the bottom 25% and one for class size in top 25%, class geographical concentration (measured as Herfindhal index of concentration in city-blocks), Class share in bottom 10% of house value, Class share in top 10% of house value, School/cohort fixed effects, Group of teachers fixed effects every five years (i.e. school-section fixed effects every five years).

Table 8: The effect of high-school class gender composition on major choice by parents' house value

Sample→  Treatment↓	Females			Males		
	(1)	(2)	(3)	(4)	(5)	(6)
	All	Low House Value	High House Value	All	Low House Value	High House Value
Panel A - Dep. Var.: (PM major=1)						
Own Gender Share>80%	0.013 (0.011)	0.007 (0.027)	0.041 (0.030)	0.063** (0.031)	0.139*** (0.053)	0.073 (0.073)
Own Gender Share>90%	0.010 (0.037)	0.044 (0.103)	0.090 (0.071)	0.154*** (0.043)	0.240*** (0.069)	0.103* (0.057)
Dep. var. mean	0.139	0.145	0.137	0.427	0.443	0.426
Panel B - Dep. Var.: (GB major=1)						
Own Gender Share>80%	0.005 (0.022)	0.091** (0.044)	-0.040 (0.045)	-0.045 (0.035)	-0.073 (0.056)	-0.060 (0.086)
Own Gender Share>90%	-0.024 (0.039)	-0.113 (0.162)	0.032 (0.092)	-0.127*** (0.038)	-0.213*** (0.071)	0.045 (0.077)
Dep. var. mean	0.586	0.588	0.587	0.460	0.447	0.460
Panel C - Dep. Var.: (PF major=1)						
Own Gender Share>80%	-0.018 (0.020)	-0.098** (0.041)	-0.001 (0.052)	-0.018* (0.011)	-0.066** (0.027)	-0.012 (0.037)
Own Gender Share>90%	0.015 (0.053)	0.068 (0.154)	-0.121 (0.088)	-0.027 (0.016)	-0.027 (0.026)	-0.148*** (0.054)
Dep. var. mean	0.275	0.267	0.276	0.113	0.110	0.114
Observations	10,318	2,442	2,142	9,773	2,215	2,277
<i>Individual Controls:</i>						
Fam. wealth proxy	X	X	X	X	X	X
<i>Class-level controls:</i>						
Size	X	X	X	X	X	X
Geo. Concentration	X	X	X	X	X	X
Fam. wealth proxy	X	X	X	X	X	X
SchoolX Cohort FE	X	X	X	X	X	X
SX TeachersX 5yrs FE	X	X	X	X	X	X

**Method:** Each coefficient corresponds to one specification. OLS, Standard errors clustered at school/cohort level in parenthesis, \*\*\* p<0.01, \*\* p<0.05, \* p<0.1.

**Specifications:** Specification (1) on women only. Specification (2) on women whose families used to live in a house in the bottom quartile of house price distribution of Milan at the time of high school attendance, (3) women in the top quartile of the same distribution. Specification (4) on men only. Specification (5) on men whose families used to live in a house in the bottom quartile of house price distribution of Milan at the time of high school attendance, (6) men in the top quartile of the same distribution.

**Sample:** High school graduates enrolling in college.

**Dependent variables:** In panel A dummy=1 if student enrolled in Engineering, Economics and Business at first college enrollment. In panel B a dummy=1 if student enrolled in Natural Sciences, Mathematics, Statistics, Computer Sciences, Medicine, Agriculture, Architecture, Design, Social Sciences or Law. In panel C a dummy=1 if student enrolled in Humanities or Education.

**Treatment:** In first row specifications of each panel dummy for classes with more than 80% of own-gender classmates. In second row of each panel dummy for classes with more than 90% of own-gender classmates.

**Individual Controls:** dummy=1 if student used to live in a house in top (and another dummy for bottom) 10% of house value distribution, dummy=1 if student used to commute from outside the city.

**Class-level controls:** dummy=1 if class size in the bottom 25% and one for class size in top 25%, class geographical concentration (measured as Herfindhal index of concentration in city-blocks), Class share in bottom 10% of house value, Class share in top 10% of house value, School/cohort fixed effects, Group of teachers fixed effects every five years (i.e. school-section fixed effects every five years).

Table 9: The effect of high-school class gender composition on college outcomes

Treatment↓	Sample→	Females			Males		
		(1)	(2)	(3)	(4)	(5)	(6)
		All	Bot Qual	Top Qual	All	Bot Qual	Top Qual
Panel A - Dep. Var.: (College Graduate=1)							
Own Gender Share>80%		0.004 (0.021)	-0.039 (0.049)	0.027 (0.028)	0.021 (0.021)	-0.011 (0.093)	0.020 (0.034)
Own Gender Share>90%		-0.022 (0.052)	-0.114 (0.094)	0.024 (0.071)	-0.045* (0.025)	-0.264 (0.173)	-0.067* (0.039)
Dep. var. mean		0.759	0.594	0.871	0.712	0.522	0.864
Panel B - Dep. Var.: (Ever graduated in PM Major=1)							
Own Gender Share>80%		0.008 (0.011)	0.006 (0.020)	-0.009 (0.028)	0.041 (0.026)	0.140** (0.068)	-0.024 (0.048)
Own Gender Share>90%		0.016 (0.033)	0.025 (0.036)	0.050 (0.049)	0.079*** (0.019)	0.170** (0.068)	0.108 (0.085)
Dep. var. mean		0.119	0.038	0.203	0.321	0.141	0.499
Observations		10,318	2,100	3,075	9,771	1,972	2,863
Panel C - Dep. Var.: Drop Out of PM Major=1 conditional on enrollment in PM Majors							
Own Gender Share>80%		0.024 (0.069)	0.250 (0.864)	0.195** (0.097)	0.016 (0.031)	0.138 (0.175)	-0.032 (0.049)
Own Gender Share>90%		-0.063 (0.173)	0.833 (2.180)	0.005 (0.198)	0.043 (0.046)	0.242 (0.233)	0.035 (0.024)
Dep. var. mean		0.169	0.378	0.107	0.253	0.462	0.130
Observations		1,433	119	692	4,171	504	1,635
Panel D - Dep. Var.: Time to Graduation - first spell							
Own Gender Share>80%		1.037 (0.987)	-6.736* (3.889)	1.429 (1.640)	0.542 (1.514)	4.802 (6.814)	1.298 (2.494)
Own Gender Share>90%		2.544 (2.882)	-2.627 (9.591)	2.319 (3.043)	-1.623 (3.657)	9.763* (5.032)	8.076 (5.960)
Dep. var. mean		79.301	87.028	73.602	82.508	89.515	77.204
Observations		7,438	1,160	2,588	6,537	945	2,363
Panel E - Dep. Var.: College graduation score conditional on time to graduation							
Own Gender Share>80%		-0.001 (0.007)	-0.040 (0.024)	-0.002 (0.008)	0.007 (0.010)	-0.064 (0.045)	-0.008 (0.016)
Own Gender Share>90%		-0.035** (0.014)	-0.093* (0.049)	-0.024 (0.025)	-0.012 (0.013)	-0.137*** (0.022)	-0.022 (0.030)
Dep. var. mean		0.860	0.764	0.926	0.780	0.689	0.853
Observations		7,416	1,149	2,584	6,511	941	2,359
Individual Controls:							
Fam. wealth proxy		X	X	X	X	X	X
Class-level controls:							
Size		X	X	X	X	X	X
Geo. Concentration		X	X	X	X	X	X
Fam. wealth proxy		X	X	X	X	X	X
SchoolXCohort FE		X	X	X	X	X	X
SXTeachersX5yrs FE		X	X	X	X	X	X

**Method:** Each coefficient corresponds to one specification. OLS, Standard errors clustered at school/cohort level in parenthesis, \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ . Spec. (1) on women only, (2) on women in the bottom quartile of the within school/cohort/gender rank in high school exit score, (3) women in the top quartile. Spec. (4) on men only, (5) on men in the bottom quartile of the within school/cohort/gender rank in high school exit score, (6) men in the top quartile. **Dependent variables:** In panel A dummy=1 if individual completed a college degree. In B dummy=1 if ever completed a PM degree. In C a dummy=1 if dropout of PM majors, conditional on being enrolled in a PM major. In D time elapsed between high school graduation and completion of the degree chosen at first college enrollment (in months). In E college final score (computed as a weighted average of GPA and thesis evaluation and discussion) and conditional on time to graduation. **Treatment:** In first row of each panel, dummy for classes with more than 80% of own-gender classmates. In second row, dummy for classes with more than 90%. **Controls:** as in Table 7

Table 10: The effect of high-school class gender composition on post college outcomes

Treatment↓ Sample→	Females			Males		
	(1)	(2)	(3)	(4)	(5)	(6)
	All	Bot Qual	Top Qual	All	Bot Qual	Top Qual
Panel A - Dependent Variable: (Enrolled in College=1)						
Own Gender Share>90%	-0.002 (0.029)	0.098 (0.062)	-0.084 (0.064)	-0.032 (0.040)	-0.009 (0.116)	0.042 (0.040)
Dep. var. mean	0.777	0.734	0.821	0.786	0.737	0.832
Observations	11,762	2,509	3,340	11,049	2,342	3,072
Panel B - Dependent Variable: (Income is observed==1)						
Own Gender Share>90%	-0.011 (0.053)	0.131* (0.073)	-0.105 (0.073)	-0.040 (0.047)	-0.024 (0.102)	0.011 (0.063)
Dep. var. mean	0.766	0.740	0.781	0.779	0.761	0.785
Observations	11,762	2,509	3,340	11,049	2,342	3,072
Panel C - Dependent Variable: Log of Annual Earnings						
Own Gender Share>90%	-0.018 (0.143)	-0.235 (0.331)	0.343 (0.333)	0.143 (0.095)	0.196 (0.259)	-0.004 (0.115)
Dep. var. mean	9.474	9.282	9.657	9.930	9.758	10.083
Observations	7,615	1,543	2,247	7,613	1,558	2,149
Panel D - Dependent Variable: Log of Annual Earnings (only enrolled in college)						
Own Gender Share>90%	-0.076 (0.133)	0.143 (0.391)	0.152 (0.316)	0.125 (0.078)	0.222 (0.288)	-0.065 (0.092)
Dep. var. mean	9.488	9.292	9.667	9.907	9.735	10.064
Observations	6,325	1,229	1,957	6,419	1,246	1,890
Panel E - Dep. Var.: Log of expected annual earning, imputed based on major of enrollment						
Own Gender Share>90%	-0.017 (0.046)	0.074 (0.111)	-0.043 (0.068)	0.097*** (0.028)	0.245*** (0.088)	0.106*** (0.029)
Dep. var. mean	9.988	9.926	10.053	10.184	10.091	10.258
Observations	9,136	1,839	2,741	8,684	1,725	2,555
<i>Individual Controls:</i>						
Fam. wealth proxy	X	X	X	X	X	X
<i>Class-level controls:</i>						
Size	X	X	X	X	X	X
Geo. Concentration	X	X	X	X	X	X
Fam. wealth proxy	X	X	X	X	X	X
SchoolXCohort FE	X	X	X	X	X	X
SXTeachersX5yrs FE	X	X	X	X	X	X

**Method:** Each coefficient corresponds to one specification. OLS, Standard errors clustered at school/cohort level in parenthesis, \*\*\* p<0.01, \*\* p<0.05, \* p<0.1. Specification (1) on women only. Specification (2) on women in the bottom quartile of the within school/cohort/gender rank in high school exit score, (3) women in the top quartile. Specification (4) on men only. Specification (5) on men in the bottom quartile of the within school/cohort/gender rank in high school exit score, (6) men in the top quartile.

**Sample:** High school graduates enrolling in college by the year 2000.

**Dependent variables:** In panel A dummy=1 if individual enrolled in university. In panel B dummy=1 if individual tax return data matched. In Panel C logarithm of personal income, as revealed to the internal revenue service in year 2005. In panel D logarithm of personal income as in C restricting the specification to individuals who enrolled in college. In panel E logarithm of personal income calculated on our sample and imputed by choice of major

**Treatment:** Dummy for classes with more than 90% of own-gender classmates.

**Individual Controls:** dummy=1 if student used to live in a house in top (and another dummy for bottom) 10% of house value distribution, dummy=1 if student used to commute from outside the city.

**Class-level controls:** dummy=1 if class size in the bottom 25% and one for class size in top 25%, class geographical concentration (measured as Herfindhal index of concentration in city-blocks), Class share in bottom 10% of house value, Class share in top 10% of house value, School/cohort fixed effects, Group of teachers fixed effects every five years (i.e. school-section fixed effects every five years).

Table 11: The effect of high-school class gender composition on the size of friends network

		Females					Males				
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Treatment↓	Sample→	All	Bot Qual	Top Qual	Low SES	High SES	All	Bot Qual	Top Qual	Low SES	High SES
Panel A - Dependent Variable: Number of peers with same choice of major											
Own Gender Share > 80%		-0.076 (0.174)	0.155 (0.316)	-0.114 (0.236)	-0.085 (0.331)	0.383 (0.300)	0.749*** (0.269)	0.870* (0.484)	0.665** (0.324)	1.352*** (0.443)	0.727 (0.441)
Own Gender Share > 90%		-0.028 (0.361)	0.285 (0.438)	-0.511 (0.561)	-0.518 (0.576)	-0.186 (0.517)	2.446*** (0.369)	3.280*** (0.481)	2.086*** (0.455)	3.545*** (0.624)	1.669*** (0.445)
Observations		10,318	2,088	3,074	2,426	2,106	9,770	1,961	2,857	2,189	2,239
Dep. var. mean		2.640	2.678	2.670	2.521	2.653	2.769	2.485	2.912	2.857	2.798
Panel B - Dependent Variable: Number of peers with same choice of major and university											
Own Gender Share > 80%		0.134 (0.146)	0.313 (0.232)	-0.018 (0.195)	0.235 (0.332)	0.446** (0.206)	0.659** (0.253)	0.844* (0.448)	0.546* (0.313)	1.318*** (0.435)	0.706* (0.404)
Own Gender Share > 90%		0.298 (0.397)	0.625 (0.459)	-0.165 (0.518)	0.073 (0.522)	0.019 (0.313)	2.140*** (0.457)	2.495*** (0.766)	1.659*** (0.469)	3.318*** (0.671)	0.994** (0.429)
Observations		10,318	2,088	3,074	2,426	2,106	9,770	1,961	2,857	2,189	2,239
Dep. var. mean		2.026	2.071	2.043	1.958	2.066	2.336	2.004	2.552	2.458	2.385
<i>Individual Controls:</i>											
Fam. wealth proxy		X	X	X	X	X	X	X	X	X	X
<i>Class-level controls:</i>											
Size		X	X	X	X	X	X	X	X	X	X
Geo. Concentration		X	X	X	X	X	X	X	X	X	X
Fam. wealth proxy		X	X	X	X	X	X	X	X	X	X
SchoolXCohort FE		X	X	X	X	X	X	X	X	X	X
SXTeachersX5yrs FE		X	X	X	X	X	X	X	X	X	X

**Method:** Each coefficient corresponds to one specification. OLS. Standard errors clustered at school/cohort level in parenthesis, \*\*\* p<0.01, \*\* p<0.05, \* p<0.1. Spec. (1) on all women, (2) on women in the bottom quartile of the within school/cohort/gender rank in high school exit score, (3) women in the top quartile. Spec. (4) on women whose families used to live in a house in the bottom quartile of house price distribution of Milan at the time of high school attendance, (5) women in the top quartile of the same distribution. Spec. (6) on all men, (7) on men in the bottom quartile of the within school/cohort/gender rank in high school exit score, (8) men in the top quartile. Spec. (9) on men whose families used to live in a house in the bottom quartile of house price distribution of Milan at the time of high school attendance, (10) men in the top quartile of the same distribution.

**Dependent variables:** In panel A, number of high school classmates who decided to enroll in same college major. In B, number of high school classmates who enrolled in same university and major.

**Treatment:** In first row of each panel, dummy for classes with more than 80% of own-gender classmates. In second row, dummy for classes with more than 90%.

**Individual Controls:** dummy=1 if student used to live in a house in top (and one for bottom) 10% of house value distribution, dummy=1 if student used to commute from outside the city.

**Class-level controls:** dummy=1 if class size in the bottom 25% and one for class size in top 25%, class geographical concentration (measured as Herfindal index of concentration in city-blocks), Class share in bottom 10% of house value, Class share in top 10% of house value, School/cohort fixed effects, Group of teachers fixed effects every five years (i.e. school-section fixed effects every five years).

Table A1: The effect of high-school class gender composition on students' major choice by bin of own gender share

VARIABLES Explanatory Var.↓	Females			Males		
	(1)	(2)	(3)	(4)	(5)	(6)
	PMM=1	GBM=1	PFM=1	PMM=1	GBM=1	PFM=1
Own Gender Share>10% and <20%	0.113 (0.112)	-0.115 (0.175)	0.002 (0.119)	0.033 (0.065)	-0.001 (0.078)	-0.032 (0.081)
Own Gender Share>20% and <30%	0.043 (0.101)	-0.031 (0.161)	-0.012 (0.110)	0.065 (0.062)	-0.014 (0.074)	-0.050 (0.072)
Own Gender Share>30% and <40%	0.059 (0.098)	-0.063 (0.161)	0.004 (0.112)	0.077 (0.061)	-0.014 (0.072)	-0.063 (0.074)
Own Gender Share>40% and <50%	0.052 (0.099)	-0.051 (0.162)	-0.001 (0.112)	0.077 (0.064)	-0.018 (0.074)	-0.059 (0.072)
Own Gender Share>50% and <60%	0.053 (0.100)	-0.040 (0.163)	-0.013 (0.112)	0.084 (0.064)	-0.016 (0.074)	-0.067 (0.073)
Own Gender Share>60% and <70%	0.035 (0.100)	-0.040 (0.163)	0.005 (0.113)	0.078 (0.065)	-0.018 (0.074)	-0.059 (0.073)
Own Gender Share>70% and <80%	0.047 (0.100)	-0.023 (0.164)	-0.025 (0.113)	0.033 (0.066)	0.006 (0.077)	-0.040 (0.074)
Own Gender Share>80% and <90%	0.057 (0.101)	-0.026 (0.164)	-0.031 (0.115)	0.108 (0.072)	-0.032 (0.084)	-0.076 (0.075)
Own Gender Share>90%	0.059 (0.106)	-0.056 (0.168)	-0.002 (0.124)	0.238*** (0.075)	-0.147* (0.078)	-0.091 (0.075)
Observations	10,318	10,318	10,318	9,773	9,773	9,773
Dep. var. mean	0.139	0.586	0.275	0.427	0.460	0.113
<i>Individual Controls:</i>						
Fam. wealth proxy	X	X	X	X	X	X
<i>Class-level controls:</i>						
Size	X	X	X	X	X	X
Geo. Concentration	X	X	X	X	X	X
Fam. wealth proxy	X	X	X	X	X	X
SchoolXCohort FE	X	X	X	X	X	X
SXTeachersX5yrs FE	X	X	X	X	X	X

**Method:** OLS, Standard errors clustered at school/cohort level in parenthesis, \*\*\* p<0.01, \*\* p<0.05, \* p<0.1. Specifications (1)-(3) on women only. (4)-(6) on men only.

**Sample:** high school graduates enrolling in college.

**Dependent variables:** In specifications (1) and (4) dummy=1 if student enrolled in Engineering, Economics and Business at first college enrollment. In (2) and (4) a dummy=1 if student enrolled in Natural Sciences, Mathematics, Statistics, Computer Sciences, Medicine, Agriculture, Architecture, Design, Social Sciences or Law. In (3) and (6) a dummy=1 if student enrolled in Humanities or Education.

**Treatment:** Dummies for each 10% bin of the share of own-gender classmates.

**Individual Controls:** dummy=1 if student used to live in a house in top (and another dummy for bottom) 10% of house value distribution, dummy=1 if student used to commute from outside the city.

**Class-level controls:** dummy=1 if class size in the bottom 25% and one for class size in top 25%, class geographical concentration (measured as Herfindhal index of concentration in city-blocks), Class share in bottom 10% of house value, Class share in top 10% of house value, School/cohort fixed effects, Group of teachers fixed effects every five years (i.e. school-section fixed effects every five years).

Table A2: Effect on major choice by academic quality (above/below median quality).

Sample→ Treatment↓	Females			Males		
	(1)	(2)	(3)	(4)	(5)	(6)
	All	Below Med Quality	Above Med Quality	All	Below Med Quality	Above Med Quality
Panel A - Dep. Var.: (PM major=1)						
Own Gender Share>80%	0.013 (0.011)	0.038** (0.017)	0.002 (0.021)	0.063** (0.031)	0.090 (0.055)	0.011 (0.037)
Own Gender Share>90%	0.010 (0.037)	-0.032 (0.036)	0.051 (0.056)	0.154*** (0.043)	0.240*** (0.060)	0.095 (0.069)
Dep. var. mean	0.139	0.080	0.187	0.427	0.314	0.519
Panel B - Dep. Var.: (GB major=1)						
Own Gender Share>80%	0.005 (0.022)	-0.002 (0.037)	0.008 (0.026)	-0.045 (0.035)	-0.044 (0.063)	-0.017 (0.036)
Own Gender Share>90%	-0.024 (0.039)	0.051 (0.050)	-0.109** (0.053)	-0.127*** (0.038)	-0.180*** (0.063)	-0.115** (0.057)
Dep. var. mean	0.586	0.630	0.549	0.460	0.565	0.375
Panel C - Dep. Var.: (PF major=1)						
Own Gender Share>80%	-0.018 (0.020)	-0.036 (0.034)	-0.010 (0.025)	-0.018* (0.011)	-0.046 (0.028)	0.006 (0.016)
Own Gender Share>90%	0.015 (0.053)	-0.019 (0.063)	0.058 (0.065)	-0.027 (0.016)	-0.060** (0.027)	0.020 (0.022)
Dep. var. mean	0.275	0.290	0.264	0.113	0.121	0.107
Observations	10,318	4,641	5,677	9,773	4,393	5,380
<i>Individual Controls:</i>						
Fam. wealth proxy	X	X	X	X	X	X
<i>Class-level controls:</i>						
Size	X	X	X	X	X	X
Geo. Concentration	X	X	X	X	X	X
Fam. wealth proxy	X	X	X	X	X	X
SchoolXCohort FE	X	X	X	X	X	X
SXTeachersX5yrs FE	X	X	X	X	X	X

**Method:** OLS, Standard errors clustered at school/cohort level in parenthesis, \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ . Specification (1) on women only. Specification (2) on women below the median of the within school/cohort/gender rank in high school exit score, (3) women above the median of the same distribution. Specification (4) on men only. Specification (5) on men below the median of the within school/cohort/gender rank in high school exit score, (6) men above the same distribution. This table replicates table 7 but stratifies the sample in below vs. above the quality distribution median instead for bottom and top quartile.

**Sample:** High school graduates enrolling in college.

**Dependent variables:** In panel A dummy=1 if student enrolled in Engineering, Economics and Business at first college enrollment. In panel B a dummy=1 if student enrolled in Natural Sciences, Mathematics, Statistics, Computer Sciences, Medicine, Agriculture, Architecture, Design, Social Sciences or Law. In panel C a dummy=1 if student enrolled in Humanities or Education.

**Treatment:** In first row specifications dummy for classes with more than 80% of own-gender classmates. In second row of each panel dummy for classes with more than 90% of own-gender classmates.

**Individual Controls:** dummy=1 if student used to live in a house in top (and another dummy for bottom) 10% of house value distribution, dummy=1 if student used to commute from outside the city.

**Class-level controls:** dummy=1 if class size in the bottom 25% and one for class size in top 25%, class geographical concentration (measured as Herfindhal index of concentration in city-blocks), Class share in bottom 10% of house value, Class share in top 10% of house value, School/cohort fixed effects, Group of teachers fixed effects every five years (i.e. school-section fixed effects every five years).