NBER WORKING PAPER SERIES

THE LONG RUN EFFECTS OF HIGH-SCHOOL CLASS GENDER COMPOSITION

Massimo Anelli
Giovanni Peri

Working Paper 18744
http://www.nber.org/papers/w18744

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
January 2013

The Data Collection for this research was partially Funded by the Fondazione Rodolfo De Benedetti, Milano, Italy. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

The Long Run Effects of High-School Class Gender Composition
Massimo Anelli and Giovanni Peri
NBER Working Paper No. 18744
January 2013
JEL No. I21,J16,Z18

ABSTRACT

The long run earnings and career potential of individuals are strongly affected by their education. Among college educated individuals, the choice of college major is a very important determinant of labor market outcomes. In most countries men and women exhibit significant differences in this choice which is responsible for a large portion of the gender gap in earnings. In this paper we analyze whether the gender composition of peers (classmates) in high school affects the choice of major and hence long run earning potential. We use a newly collected and unique dataset covering 30,000 Italian students graduated from high school between 1985 and 2005. We exploit the fact that students are assigned to classes whose gender composition, within a school over time, varies exogenously. Moreover we are able to control for family, cohort, teacher and school effects in assessing the effect of peer-gender ratio on outcomes. We find that the gender ratio of peers in high school significantly affected the choice of major. A larger share of same-sex peers increases the probability of choosing majors associated to high earning jobs (Economics/Business, Medicine, Engineering). For women we also find that a large percentage of female high school classmates increases their long run performance in college and their earnings.

Massimo Anelli
Department of Economics
University of California, Davis
One Shields Avenue
Davis, CA 95616
manelli@ucdavis.edu

Giovanni Peri
Department of Economics
University of California, Davis
One Shields Avenue
Davis, CA 95616
and NBER
gperi@ucdavis.edu

1 Introduction

Two simple and important facts motivate this paper. First, among college educated individuals, the field of study (college major) is strongly correlated with earnings and career opportunities. According to the U.S. Digest of Educational Statistics (2011) the average salary of a full-time employee one year after graduation was \$ 54,900 if she had a bachelor degree in Engineering but only \$ 31,500 if she had a bachelor degree in the Humanities[1]. The choice of college major is not random, hence those earning differences certainly include differences in quality, family background and motivations. However differences in starting salary in the order of 70-80% call for a careful analysis of the determinants of the choice of major across students. Second, just as dramatic as the differences in earnings, are the differences in the gender composition of college graduates across college majors. The same U.S. Digest of Educational Statistics (2011) shows that only 18% of recent graduates in Engineering, but full 64% of graduates in the Humanities, were women [2]. These differences are not specific to the US. Flabbi (2011) shows that fields such as Engineering and Business exhibit a large initial salary advantage as well as a dramatic over-representation of males in all OECD countries. The opposite is true for college majors in the Humanities.

In spite of the fact that women have overtaken men in their college graduation rates in the US (e.g. Goldin 2006; Turner and Bowen 1999) and in most of the other developed countries, the dramatic gender difference in fields of study is responsible for a significant part of the stubbornly persistent gap in the initial salary of men and women. Understanding why students of different gender choose different fields of study is a complex and multi-faceted issue (see for instance Xie and Shauman 2003). It certainly involves understanding differences in preferences, attitudes and abilities. In this paper we look at a simple related extension of that question, namely: does the gender ratio of *peers* during high school affect that choice? In particular we focus on the gender composition of classmates in their last year of high school for a very large sample of Italian students and we analyze what effect this had on their choice of college major and on subsequent measures of performance in college and on the labor market. While the question of the effect of gender on the choice of major is important and heavily researched, the effect of peer-gender on the choice of major is as interesting and much less researched. The "counterfactual" and the policy implications of this question are clear: changing the share of women among peers may have an effect on the choice of major of a person, all else equal. To the contrary the counterfactual of the effect of gender on major choice is harder to think about (i.e. changing a person's gender leaving all else equal).

There are three reasons why we choose Italian high school students as ideal subjects to study the effect of peer gender on the choice of college major. First, in Italian high schools students are assigned to a class (usually

[1] These data are available at U.S. Digest of Education Statistics 2011 website - table 404

[2] See table 301 in the U.S. Digest of Educational Statistics (2001) at Digest of Education Statistics 2011 website.

made of 20 to 30 peers) when they enroll in high school and, except for small rates of attrition, they share with them the same teachers and in a pre-determined academic curriculum for 5 years (from 13 to 18 years old). These peers, therefore, interact with each other intensely during five years of their school life, a period in which their personal and gender identity are fundamentally shaped. We only observe individuals and their peers in the last year of high school, hence we can be sure that they have been with those same peers at least for that year, but in most cases they have shared most of the five years with the same group. Second, individuals do not choose their class or their classmates since students are assigned to classes by the principal of the school. Different classes in a school have different sets of teachers and we can identify students having the same set of teachers. Year after year these teachers were assigned a group of students (class) whose gender ratio varied significantly in an exogenous way (as we will show). As there was no prescription in Italian high schools to balance the gender ratio in a class, several accidents (e.g. different gender ratio in cohorts, different size of cohorts and different size of classes, different attrition) produced variation of gender ratio across classes that appear totally exogenous. Certainly, as we will show, the class gender ratio was uncorrelated to any of the predetermined and observable characteristics of the students and of the class. This is the ideal context to separate school, cohort and teacher effects from effects of the class gender-ratio. Finally thanks to an unprecedented data collection effort[3] that we organized and coordinated, we now own data on 30,000 college preparatory high school students in the city of Milan (Italy) who graduated between 1985 and 2005. Those data include information on their school, cohort, class, identity of their peers, year of graduation, residence address when in high school, exit-test scores. Moreover we have merged this information with that on their college career (choice of major, graduation date, university attended) on labor market outcomes (earnings in year 2005, occupational choice) and on their family background. These data allow us to address the above-specified questions and learn a lot about high school, college and labor market choices and performance of these individuals.

We have simplified the choice of majors by splitting them into two groups only. One is denoted as "high earning majors" and it includes Engineering, Business/Economics and Medicine, the three majors associated with highest earnings in our sample. The other includes all the other majors and can be considered as "non-high earning majors"[4]. A dummy indicating the choice of a high earning major will be one of our variables of interest. These three college majors are also characterized by a very low percentage of female graduates. Specifically only 33% of graduates in high earning majors were women in our sample while more than 60% of graduates in the other majors were women. One could identify alternative dimensions in the choices of major. Numerous studies, related to gender, focus on the STEM (Science, Technology, Engineering and Mathematics) versus non-STEM choice of major. As we will document below our partition of majors in "high earning" and "non-high earning"

[3]We are grateful to the Fondazione Rodolfo Debenedetti and to Universita' Bocconi for contributing to the funding of the data collection.

[4]The majors included in the "non-high earning" group are the following: Law, Mathematics and Statistics, Natural Sciences, Social Sciences, Agriculture, Humanities, Architecture and Design and Education.

is much more strongly correlated with potential earnings, male bias, gender gap and selectivity in terms of student quality than the STEM vs. non-STEM partition. Hence, in our sample we can talk about the high earning majors as very competitive, associated to high returns and "typically male", which are important and interesting features for our findings.

The most relevant findings of our empirical analysis are three. First, we find that, controlling for academic quality and family background, being assigned to a high school class with a larger share of people of the same gender increases the probability of choosing high earning college majors. This finding is true for both men and women. In particular women in high school classes with very high percentage of women have a 5-6% higher probability to choose high-earning majors than women in classes with high percentage of men. Similarly men in classes with high percentages of men have a 6-7% higher probability of choosing high earning majors. Second, those women who attended classes with large percentage of women and then choose high earning majors performed better than or as well as other women in those majors in terms of dropout rates and time to graduation. Hence the long run effect of attending high school classes with high percentage of women is, for a woman, a higher probability of enrolling and graduating from high earning majors and therefore higher expected wage. On the other hand men who attended high school classes with high percentage of men had higher probability of enrolling in high-earning majors but the final probability of graduating from high earning major was not different from that of other men due to higher attrition. Finally, women who attended high school classes with high percentage of women and then enrolled in high earning majors had higher wages on the labor market relative to other women above and beyond (i.e. even controlling for) the effect driven by the major.

These findings can be rationalized with a simple human capital model in which the choice of major is affected by its costs and benefits. If the individual cost of attending a high earning major depends on abilities (negatively) and on the self-confidence of a person (positively), the low enrollment of women in those majors may be due to a low degree of self-confidence. The share of women in the same high school class may increase the self-confidence of a woman, as also found in the experimental evidence presented in the behavioral economic literature (Gneezy, Niederle and Rustichini 2003, Niederle and Vesterlund 2007) and as hypothesized in the theoretical literature on gender identity introduced by Akerlof and Kranton (2000). In this frame, the presence of a high percentage of female in the class may partially remedy the lower self-confidence of women driving them to choose the more competitive, high earning, typically male college majors.

The rest of the paper is organized as follows. In section 2 we frame our paper within the literature. In section 3 we present our dataset and we describe the set-up of our empirical test, describing in some detail the relevant features of the Italian high school system. In section 4 we show that our dichotomous partition of majors has a very important role in determining earnings and in explaining the gender gap in earnings. In section 5 we present our identification strategy and illustrate the methodology of our empirical estimation. In section 6 we show results on the effects of high school class share of female on the choice of major and on college and labor market outcomes. Section 7 provides some robustness checks for our analysis. In section 8 we suggest a basic theoretical model to interpret our results. We conclude our analysis discussing implications and policies in the final section 9.

2 Literature Review

The literature on the effect of peer gender on school (and labor market) choices and outcomes is not very large. Most of the literature analyzes differences between students in same-sex versus coed schools leaving unsolved the issue of selection (versus causation) and analyzing only the extreme case of all same-sex peers. Solnick (1995) considers data on the anticipated and the final college majors of 1700 female students at eight single-sex colleges and compares those with the choice of 818 female students at seven coed colleges. She tests whether women at single-sex colleges are more likely than their counterparts in coed institutions to stay in traditionally male-dominated fields. She also tests whether they are more likely to shift from traditionally female-dominated to male-dominated majors. The analysis finds that women who begin in traditionally female majors are more likely to shift to typically male majors if they attend an all-women college than if they attend a coed one. Similarly women who start in traditionally male majors are more likely to persist in those majors if they attend an all-women college. Thus she argues that single-sex schools benefit female students by providing them with incentives (possibly by lowering the psychological cost) for moving to majors in traditionally male-dominated fields.

Billger (2002) exploits the case of a female college becoming coed to estimate the effects of attending a women's college on the choice of college major, on the probability of degree attainment, and on the occupation. The study compares women in the all-female cohorts with those in the coed cohorts to see whether women pursued different fields and careers. The results of this analysis suggest significant differences between the women in the all-female (earlier) versus coeducation (later) cohorts. After the admission of men, female students were significantly less likely to pursue male dominated college majors and occupations. The interpretation of these results is that coeducational settings might reinforce gender stereotypes, while single-sex schooling might give more freedom in exploring interests and abilities beyond socially constructed roles, especially for female students. A more recent paper by the same author however, Billger (2008), questions this earlier evidence by using more refined econometrics techniques. In this work Billger exploits the fact that "Title IX" allowed for the first time the creation of single-sex public schooling in the US to investigate whether single-sex schooling leads to improved labor market outcomes. Results show that single-sex education graduates are no more likely to pursue college degrees, and are less likely to meet their own educational expectations then graduates from coed educational environments.

The contributions described above, however, have not been able to deal convincingly with the selection issue. The choice of one-sex or coed education is not random and can be correlated with academic, personal and family attributes of women (and men). Better labor market outcomes, or specific major choices, might be the result of the fact that more career focused women might choose single-sex institutions in the first place. A recent working paper by Favara (2012) looks more in detailed to educational choices in single-sex schools, showing how they may attenuate the influence of gender-stereotypes. Favara uses UK data and focuses on students in lower and upper secondary education. Results of this work suggest that gender stereotyping affects educational choices from the age 14 onwards and such effect is stronger for girls than for boys. Attending a single-sex school leads students to a less stereotyped educational choice. Finally also related to our paper is a recent paper by Schneeweis and Zweimuller (2012) analyzing the causal impact of the gender composition in coeducational classes on the choice of school type for female students attending primary schools in two Austrian cities. Using natural variation in the gender composition of adjacent cohorts within schools, they show that girls are less likely to choose a traditionally female dominated school type and more likely to choose a male dominated school type if they were exposed to a higher share of girls in previous grades. The small size of the sample, the lack of data on long-run outcomes, the inability to control for school-cohort and teacher effects and the smaller range of variation of the explanatory variable, however, make those results rather weak.

Also related to our paper is the literature focussing on the determinants of major choice in college and its gender component (Xie and Shauman, 2003) and the literature on the effect of the gender of teacher on school choices and student performance (e.g. Carrell, Page and West 2010, Dee 2007). Interesting and relevant in interpreting our results is also the experimental and theoretical literature that analyzes differences between men and women in terms of their identity and their attitude towards competition. Akerlof and Kranton (2000, 2002) postulate that identity is a fundamental determinant of social and economic behavior because it provides prescription on how to behave. If the strength of one's gender identity is shaped in relation to peers during the years of high school then the gender composition of those peers may affect the cost of deviating from the prescriptions of one's gender identity. Also Gneezy, Niederle and Rustichini (2003) and Niederle and Vesterlund (2007, 2010) show in experimental evidence that the gender composition of the opponents affect significantly the performance in tournaments, especially of women. Women perform significantly better in tournaments against other women than against men. Considering the class-environment as the natural "competition" environment in high school, and following those results, protracted interaction with competitors of the own sex may reinforce self-confidence in women, while competition with males can inhibit it.

3 Data and Details on Italian High School

Our sample is a group of highly educated individuals who graduated from college-preparatory high schools (Licei) between 1985 and 2005 in the city of Milan, Italy. These individuals are currently between the ages of 25 and 45. We gathered manually (from hard copies of documents) information about their high school exit test, the type of high school attended, their class in high school and the location of their home during high school. While some missing and destroyed records prevented full coverage we were able to cover more than 90% of all records for students who graduated between 1985 and 2005 in thirteen college preparatory high schools in Milan (Italy). This includes about 30,000 individuals. Milan is a large service-oriented metropolitan area in the richest part of Italy. The college educated individuals from this city often become part of the northern Italian elites in business, finance and academia. Hence they are an interesting group that allows us to analyze the gender gap at the top of the income and educational distribution in Italy. For these individuals we have information on the year of graduation, the grade in the high school exit exam, the school attended, the location where they lived during high school and the identity of their parents. Most importantly we know about the identity of their class-mates and we know who shared the same set of professors.

The dataset links (using names and date of birth) these data with the student records from all Universities in Milan (there are five of them, two private universities, Universita' Cattolica and Bocconi, and three public universities, Politecnico, University of Milano, University of Milano-Bicocca). Therefore, we can reconstruct the university career of the individual who graduated from a college-preparatory high school in Milan, including the following information: whether they graduated, in what year, in what field of study, in what university and their exit score. In a further step we linked these records to data on the personal income of the individuals in year 2005, as revealed to the internal revenue service. This is the information on their total income, reported on which individual pay taxes. There are some advantages and one main disadvantage in using these income data. The advantage is that the administrative file of reported income includes all individuals in the national territory as it is mandatory to report any income. Hence if a person does not appear is because he/she has no income. Self-employed are included in the sample. The disadvantage is that we do not have a measure of hours worked.

Data on income or wages are very rare in Italian datasets. This data provides a very important new tool in the analysis of income differential and gender gap especially as it is linked to high school and college performance. We use these income data only for people who graduated from high school before 2000. Considering an average college attendance of five years, people in our sample would have been on the labor market for up to 15 years in 2005. Hence this provides a good assessment of the long-run consequences of one's high school experience on the labor market outcomes during the early career (up to 20 years later). As mentioned above, the total number of individuals for which we have data on the high school and University career is around 30,000. For

14,000 of them, those graduated between 1985 and 2000, we match the information on income in year 2005. For a stratified 10% random sub-sample of the initial universe (equal to 3,069 individuals) we also collected much more detailed information from telephone interviews conducted in June 2011 by the professional company "Carlo Erminero & Co.". The additional information covers several variables regarding the family background, parental income, job and education, current employment and current family situation of the individual. We will use those data in some robustness checks.

The setting of our sample, Italian high school in the period 1985-2005, is excellent to identify peer effects. First of all, we can identify in the data individuals belonging to the same class in the last year of high school. In public high schools, during this period, classes were formed in the first year by pooling all new entrant students and drawing a certain number in each class. Students could indicate a preference for being with some other specific student (e.g. an old classmate in middle school) but there was no insurance of this happening. While the criteria to form individual classes may have departed from pure randomness there was no provision for balance in gender composition. To the best of our knowledge the "Licei" in Milan followed an exogenous assignment rule. However, we check for the orthogonality of the gender ratio in each class to all observable student characteristics. Second, a class in an Italian high school was a set of students with the same identical curriculum and the same professors. Hence students in the same class were exposed to a long period of interactions at a crucial age (13 to 18). Some student were lost and other gained to a class by attrition. The students we observe have at least shared the same class and professors for the last year of high school and, likely, for most of the five years. This protracted interactions at a crucial age are good candidates to shape taste and some personality traits of students.

In Table 1 we present descriptive statistics for our data. We divide variables in individual- and class-level variables. For individual-level variables we present statistics for the whole sample and the mean separately for men and women. We have information about individuals' academic career in high school and in college. We also know pre-treatment characteristics such as the house value where the students used to live at the time they attended high school. As house was the primary asset of families and the value of real estate varies significantly within the city of Milano, the home value is a good proxy for family's income. We also have a significant number of siblings within the sample so that we can control in some specification for a family fixed effect.

In our sample of 29,382 high school graduates 52.3% are female and thus 52.3% is also the average female-ratio faced by the individuals. We have re-scaled the high school exit test score to be between 0 and 1, with 0 being the minimum passing score (60 out of 100 in the "post 2000 Maturita' " test score) and 1 the maximum score achievable (100 out of 100). Summary statistics show that the distribution of scores is skewed to the left (mean score is 0.416). Also the women's score distribution has substantially higher mean than the men's (the t-statistics of the difference in means is 11.481). In terms of the ranking within school and within the cohort, females perform substantially better scoring on average at 0.516 versus 0.471 of men with a t-statistic for the difference of 12.4.

Of these 29,382 high school graduates, 22,842 students enrolled in one of the Universities in Milan. Of them 32.5% enrolled high earning majors (Engineering, Economics & Business and Medicine). The gender difference however is substantial: only 20.8% of women chose one of the high earning majors while 45% of men did. The t-statistic for the difference in means is 40.7. Of the 22,842 students enrolling in a university, 17,050 actually graduated (by 2011, when we collected the data) implying a 25% attrition rate. Out of all college graduates, 24% of women earned a degree in a high earning major, while 50% of men did. Interestingly, conditional on enrolling in a high earning major, women have a substantially lower drop out rate than men: 13% versus 20% with a t-statistic of the difference equal to 6.9. To complete the list of college outcomes in our dataset, women take on average 3 months less than men (t-statistic is 7) to graduate from college when the average time to completion was 7 years [5]. 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[6]. We have re-scaled this final score to be between 0 and 1 (with 0 being the minimum passing score) and also for this outcome women perform better than men (0.88 versus 0.81 with a t-statistic of 29.6).

We linked the address where students used to live at the time of high school to the average house value of their specific neighborhood. The log of house value is on average 8.05 and women and men have no statistically different means for this wealth proxy. This shows that the average family background of female and male students in college preparatory high school was roughly the same.

For what concerns labor market outcomes we observe the wage for 17,008 students graduating from high school before 2001: since we have only 2005 income data available, we exclude individuals that are still attending college by 2005 since their earnings might not be representative of their potential earnings (i.e. part-time job while in college). The average log wage is 9.68 and the statistically significant difference (t-statistics is 23) across gender is 0.45 log points in favor of men. For the randomly stratified selected 10% sub-sample that was interviewed we are able to determine more specific outcomes like the type of occupation. We find, for instance that women had 30.8% probability to reach a top occupation (defined as manager, professional or director) versus 43.2% of men with a t-statistic of 7.1.

As for classroom data we observe a total of 1371 classes with an average size of 21.4 students and a standard deviation of 3.8. The average share of female per class is 0.52 with a large standard deviation of 0.18. Classes range from female-only to male-only ones. Interestingly, the average students' exit score by class also range from 0.12 to 0.78 showing high variance in the ability composition of classes. For what concerns the average

[5] Italian students especially during this period (1985-2005) had typically very long spells between college enrollment and graduation.

[6] The exception are engineering students who get a score out of 100. We re-scaled their scores accordingly.

socioeconomic status of classes we show in Table 1 the average percentage of students in the class that live in houses at the bottom 10% of the house value distribution. While this average is 0.12, it is interesting to see that there are classes composed by only students coming from families in the bottom decile of the wealth distribution (in schools located in poor neighborhoods). On the other hand there is a lower tendency to have pure “elite” classes since the highest percentage of students in the class that live in houses at the top 10% of the house value distribution is only 55%.

4 High Earning Majors

4.1 The long-run effects of high earning majors

In this section we justify our choice of splitting college majors into high-earning and non-high earning ones, rather than (for instance) into STEM or non-STEM. In Table 2 we regress two long-run labor market outcomes (the logarithm of observed income in 2005 and the probability of working in a top-occupation in 2011) on individual characteristics: the standardized high school exit score, the final college exit score, a proxy for family wealth and a series of school-by-cohort fixed effects to control for cohort-specific and high school quality variation. In columns 2 and 3 we add to these explanatory variables a dummy for STEM majors or one for high earning major, respectively. Graduating from one of the STEM majors increases significantly the log of income by 0.179 (19% in level earnings). Graduating from a high earning majors (Medicine, Economics & Business, Engineering), however, implies a much larger 0.636 premium in log earnings (equivalent to an 89% increase in level earning). Notice also that relative to Column 1 the inclusion of the high earning dummy increases significantly the R-square of the regression while adding the STEM dummy does not. Moreover, controlling for high earning majors in column 3 reduces to insignificant the effect of high school exit score and turns the effect of the college exit score to positive and significant. This shows two things: first, high school exit scores affect long run earning mainly by increasing the chance of choosing a high earning major. Second, within high earning majors high college exit scores affect log earning positively.

Columns 4-6 show the effect of the same covariates on the probability of working in a “top occupation” [7]. Graduation in STEM majors corresponds to a decrease in the probability of working in top occupations while graduating in a high earning major increases that probability by 12 percentage points with respect to an average probability of 36.7%. This discrepancy between STEM and high earning majors (despite some overlap of two majors) is mainly driven by the fact that a large share of graduates in Mathematics, Physics and Natural Sciences (STEM but not high earning) in our sample worked as teachers, while graduates of Engineering, Economics and Medicine did not. Positive labor market outcomes in our sample are much more strongly associated to high

[7] Defined as managers, professionals, business owners in our phone survey.

earning than to STEM majors.

4.2 Gender gap and high earning major

High earning majors also play an important role in explaining the gender gap in earnings. Specification 1 of Table 3 shows a regression of the logarithm of income in year 2005 on year of graduation dummies, a dummy for high school track (1 for scientific high school track, 0 for classical studies) and a female dummy. The regression includes all the high school graduates over the period 1985 and 2000. These individuals were mostly between 23 and 39 years of age in 2005 when we record their income, and hence well into their working career.

In the first column of Table 3 the coefficient on the female dummy is an estimate of the average gap in (logarithmic points) of gross income between women and men (controlling only for age and type of high school attended). Remarkably, this difference equals -0.38 logarithmic points (about -32 percentage points[8]) which is a very large average difference. On the one hand our measure of yearly earnings is broad and it includes salary, bonuses, overtime pay and other incentives to production. It also reflects differences in hours worked. On the other hand the earning gap is calculated in a group of college graduates, from a rich city in Northern Italy, with relatively homogeneous family backgrounds. Hence this should be a group for which prejudice and discrimination on the workplace is as low as possible within Italy and for which access to education and to opportunity is high and similar for men and women. Recent estimates of the gender earning gap in Italy reveal values that are comparable with those estimated in column 1 of Table 3. For instance the Global Gender Gap report of 2011, by the Global Economic Forum indicated that the earning of women were 46% below those of men in Italy overall. 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 for all workers it was 40%. Our figures of the gender wage differential therefore are close to the previously measured wage gender gap for college educated in Italy.

Interestingly, when we restrict the sample to individuals who enrolled to college (specification 2) and to those who graduated from college (specification 3) the gender income gap does not change much and in specification 3 it is still equal to 0.38 logarithmic points. The selection into college and then into the group that graduated from college does not affect the gender gap. Then, in specification 4, we include some controls for the academic quality of the individual, namely the standardized high school exit test score and the college exit test score. This specification reveals that the high school grade is highly significant in affecting the logarithmic wage of an individual. A student graduating at the top of the class (score of 1) would on average earn 0.37 logarithmic points more than a student graduating at the bottom (score of 0). Academic quality strongly affect labor market outcomes. The inclusion of these controls increases somewhat the gender gap (from 0.38 to 0.42 log points)

[8]The conversion from logarithmic points into percentage points is always calculated as the exponential of the logarithmic points minus one.

as women have, on average, higher high school exit scores and hence academic quality goes in the opposite direction than the gap. Controlling for school-cohort effects (column 5 of Table 3) leaves the estimated gender gap at 0.39 logarithmic points.

In columns 6 and 7 we simply add a dummy equal to one for STEM-major graduates or (alternatively) for high earning major graduates to the previous controls. Including the STEM dummy in specification 6 of Table 3 reduces the gender gap by only 1 logarithmic point, while introducing the high earning majors dummy reduces the gender gap by 10 logarithmic points, one fourth of the whole gap. Notice also that when controlling for the high-earning-major dummy, the college exit-score becomes very significant and, at the same time, the predictive power of the high school exit-score becomes much smaller. This is consistent with the interpretation that the high school score affects labor market outcomes mainly by affecting the choice of college major and our partition between high earning and non-high earning majors captures most of the effect.

5 Discussion and Tests of the Identification Strategy

During the considered period and still today the vast majority of Italian students, from rich and poor families, attended public (rather than private) high schools. The criterion for the choice of the school was based on the residence (school district) and the type of school chosen (classic or scientific type). Students usually attended the closest high school of the chosen type. As they were all public, there was virtually no difference in the marginal monetary cost of attending different schools [9].

Conditional on the choice of high school, students were assigned to one classroom (coded with a letter: A, B, C...) at the beginning of the five year cycle in an exogenous way. While the allocation of students to classes was ultimately decided by the principal in consultation with the teachers, the general principle was that allocation of students was random. We cannot verify directly if the principals followed this random assignment practise. Anecdotal evidence suggests that assignment to classrooms within school might be affected by some external factors. In particular families might exert pressure on the schools to have their children assigned to specific teachers (or more rarely to the same classroom as friends). These mechanisms could potentially introduce correlation between individual characteristics in a classroom (as wealth, ability) if families of a certain type pressure the principal to have their children in specific classes. This might generate classes with similar unobservable characteristics which may affect identification if those are correlated to the class gender ratio. We will show that the gender composition of a class is orthogonal to wealth, ability and spatial concentration of homes of individuals in the class. If present at all, most of the pressure of families was probably exerted to increase the probability that their children were assigned to "high quality" teachers (however their quality can be defined). Our identification strategy isolates this problem by exploiting a feature of the Italian system.

[9] All public schools charge minimal fee per year. Currently it is $150 per year. It was far less during the years 1985-2005.

In our sample a set of teachers is assigned to a classroom (e.g. "classroom A") when hired and then teach all the cohorts of students assigned to that specific "letter-classroom", year after year. Our data allow us to identify within each school and cohort what specific classroom each student in our sample was assigned to. As a consequence, we are able to control for the group of teachers who taught multiple cohorts over time and we only use the variation in gender composition of classrooms *across cohorts, within the same group of teachers.* Figure 1 presents a sketch of this organizational structure. We include dummies to account for school/cohort effects and dummies to account for teacher effects[10] ("letter-class" or "sezione" as called in Italy) in order to implement this method. Even more conservatively, in several specifications, as the set of professors in a "letter-class" changed slowly due to natural turnover and attrition, we include "letter-class" by five-year dummies to be sure that each fixed effect captures the same set of professors. This type of identification is reminiscent of the method used in other education contexts where there was no certainty of random assignment (see Hoxby 2000, Carrell and Hoekstra 2010 and Lavy and Schlosser 2011).

We also perform multiple checks to make sure that the variation of the gender composition within school and group of teachers and across cohorts is orthogonal to observable characteristics.

First, we check that observable pre-determined characteristics of individuals are not correlated across students in a class. This is a stronger condition than what we need for identification (which only requires orthogonality of gender ratio to other characteristics) but gives us a sense of the exogeneity of class formation along observable characteristics.

We regress individual family characteristics $X_{i,l,c,t}$ for individual i in school l, in teacher-group c and year t, on the average background characteristics of her/his classmates (excluding the individual). We run the following specification:

$$X_{i,l,c,t} = \alpha + \beta \frac{\sum_{k \neq i} X_{k,l,c,t}}{n_{l,c,t} - 1} + \varphi_{l,t} + \delta_c + \epsilon_{i,l,c,t} \tag{1}$$

where $\varphi_{l,t}$ are the school-year effects and δ_c are the teacher effects. The term $\frac{\sum_{k \neq i} X_{k,l,c,t}}{n_{l,c,t}-1}$ represents the average background characteristic of classmates (calculated on all individuals in the class except for i) and the variable $n_{l,c,t}$ is the total number of students in that class.

In Table 4 we regress the house value where the student used to live at the time of high school on its average value for the classroom. Following Guryan, Kroft and Notowidigdo (2009) we also include the mean house value for the cohort in the regression in order to avoid a negative bias that is shown to arise in this type of regressions, especially when the group from which individuals are randomly extracted is small (the school-cohort in our case). No matter if we use the log of the house value or dummies for the top (column 3) and the bottom decile

[10] As one class, year after year, shares an identical group of teachers the dummy capture the "set-of-teachers" effect. We will call this "teacher-effect" for brevity.

(column 4) of the house value distribution, the individual value is uncorrelated to the average value of the classmates. We also analyze whether there is an association in the geographic distribution of homes of students in the same class. If all students in a class come from the same block, one may think that groups of friends (who know each other from earlier years) pressure the system to be in the same class. Column 3 shows the correlation in the location of homes using distance to school as measure and finds that it is not significantly different from 0. Hence our data do not reveal any evidence that there is clustering of student homes within class more than for the school-cohort as a whole. Overall in terms of predetermined characteristics such as of home location and home value of students, our data does not reveal any correlation across students within a class, once we control for school-year and group of teacher effects.

We now move to more direct tests that the class gender composition, our key explanatory variable, exhibit significant variation that appears totally exogenous across classes. First we inspect the data. The histogram in Figure 2 shows the distribution of female shares in our sample of classes. That share has substantial variation, ranging from less than 10% to more than 90%. In Figure 3, left panel, we plot the kernel distribution of the female share in our sample versus a normal distribution and in the right panel we show a diagnostic plot that shows quantiles of female ratio distribution against quantiles of normal distribution. It is clear that the empirical distribution of female ratio in our sample of classes is quite close to a normal distribution, although with somewhat thicker tails. More troublesome for our strategy would be departure from randomness in some specific way. For instance if a group of teachers consistently pressured the principal over the years to obtain low (or high) female share in their classroom and, at the same time, they have a particular gender bias, this could generate correlation between teacher and gender ratio effects. Although this case sounds unlikely, and we only use within teacher group variation so that we would be controlling for the correlation, we verify that there is no autocorrelation in the share of women in classes within teacher-group and across cohorts.

In Figure 4 we plot for each classroom in our sample the female share for cohort t versus the female share for the previous cohort $t-1$. We split the two high school tracks since they have different average female ratio. The top panel shows the scatter plot for classes in classical-track high schools and the bottom panel for classes in the scientific-track. In both panels the plots show no correlation whatsoever. The share of female in a class-school one year has no predictive power at all over the share of women in the same class-school the following year. In order to test this in a regression setting, in Table 5 we collapse our data at the classroom level and we regress the female share of cohort t on the female ratio of previous cohort $t-1$ for each class-school. In column 1 we control for school fixed effects. In column 2 we control for school/cohort fixed effects and in column 3 we add also the dummy for group of teachers. Results show that the autocorrelation of female shares in each of these regressions is indistinguishable from zero. As the group of teachers does not change in a letter class year after year, this implies that the same group of teachers, in the same school has classes with very different gender composition year after year.

Another potential threat to our identification arises if female ratio are correlated with the composition of background characteristics in the class (i.e. if the share of kids from wealthy families in the class is correlated with the female-ratio). In Table 6 we analyze whether the gender composition of a class is correlated with the share of students coming from the wealthiest families (those living in homes with house-value in the top decile) and the share of students coming from the poorest families (with house value in the bottom decile). We also analyze if there is any correlation of the gender composition with the size of the class (number of students) and with the concentration of student home location in Milano (measured as Herfindhal index of concentration in city-blocks). After controlling for school/cohort and group of teachers fixed effects we find no significant correlation of the gender ratio with any of these characteristics of the class. This is true when we use the share of women as dependent variable (specification 1) or alternatively, when we use dummies for the class being in the first, second, third or fourth quartile of the distribution of classes in terms of their percentage of women (specifications 2-5). Having a large or small percentage of women in a class is not associated with any of the other class characteristics included. In Figure 5 we also plot the class female share vis-a-vis the other class characteristics (as included in the regressions of table 6), one at the time. The visual inspection confirms the lack of any correlation.

Last we check that the share of women in a class is uncorrelated with average ability in the class. A possible source of misinterpretation of our results arises if classes with high share of women have also women with higher (lower) academic quality. We do not have a standardized measure of pre-high school ability. Hence we use a measure of academic quality taken after high school, namely the high school exit test score, standardized for each school-cohort to account for possible heterogeneous quality of school-cohorts. In column 1 of Table 7 we regress the mean score in a class on the percentage of female in the same class. In column 2 we regress the mean female score in the class on the percentage of women. In specification 3-6 we regress the percentage in the class of women-grade in each quartile of the overall ranking on the share of women in the class. Including the usual control for school-cohort and set of teacher fixed effects we find no evidence of any correlation between female ratio and the ability composition of the class.

6 Main Specification and Results

The main empirical model that we estimate in this section is the following:

$$y_{i,l,c,t} = \phi_{l,t} + \phi_{c,T} + \beta_1 g(share\_female_{ilct}) + \beta_2 g(share\_female_{ilct})(f_{i,l,c,t}) + \alpha f_{i,l,c,t} + \delta(X_{i,l,c,t}) + \varepsilon_{i,l,c,t} \quad (2)$$

The outcome $y_{i,l,c,t}$ in our main specification is an indicator equal to one if individual i, with set of teachers c, in school l, in cohort t chooses a high earning college major and 0 otherwise. The variable $share\_female_{ilct}$ is the share of women among all classmates of individual i, calculated excluding individual i from the class. The operator $g(\cdot)$ is a function of the female share that we will chose to be linear or non-linear to take into account heterogenous effects across the female share distribution. We include $g(share\_female_{ilct})$ directly in the regression to determine the effect of the share of women on the choice of major of men. We include the dummy variable $f_{i,l,c,t}$ relative to individual i and equal to one if she is a woman. Its coefficient captures the female gap in the outcome variable. We also include the interaction of $g(share\_female_{ilct})$ with the female dummy, $f_{i,l,c,t}$, to see if there is a differential effect of the class gender composition on women and men. The coefficient β_1 captures the effect of the share of females in the class on the probability of choosing high earning majors for men. The coefficient β_2 captures the additional effect for women. The sum of the two captures the effect of the share of females in the class on the probability that a woman chooses high earning majors. We control for individual skills by including within $X_{i,l,c,t}$ the rank position of individual i within school and cohort in the final high school exit test-score and the proxy for family wealth based on the home value for each student. The terms $\phi_{l,t}$ and ϕ_c are school-cohort and group-of-teachers fixed effects. In most of the cases c is an index that vary by letter class and every 5 years so as to isolate the effect of the same group of teachers.

We will also consider other outcomes $y_{i,l,c,t}$ such as college outcomes (the graduation rate, time to graduation, drop-out rates and exit test score in college) and labor market outcomes (log of earnings). In all the estimations we cluster the standard errors at school/cohort level. We also test for robustness clustering at the more conservative school level and results are essentially identical.

We first show in section 6.1 how the share of female in the high school class affects the choice of major of individuals using our main specification. Then in section 6.2 we look at the effect of the class share of female on different academic outcomes when in college, conditional on choosing a high-paying major. Finally in section 6.3 we look at the effect of the class share of female on the earnings.

6.1 The effect of high school class female share on the choice of major

Table 8 shows our main results and basic specifications. In this table we focus on the effect of the high school class female share on the probability of enrollment in high earning majors after high school graduation. In column 1 we simply regress the high earning major dummy on a female dummy and the class female share after including dummies controlling for the high school exit test score, standardized within school and cohort, via dummies for belonging to each quartile of the rank distribution in the scores (omitted dummy is the lowest quartile). We also control for the proxy for family wealth and we include school/cohort fixed effects and teacher fixed effects. The average negative female bias in the choice of high earning majors in column 1 of Table 8 is about 22 logarithmic points (20 percentage points) conditional on skills and background characteristics, while the effect of class female share appears to be not significant. In column 2 we add the interaction of the class female share with the female dummy. This reveals the first important fact: the null effect of female ratio in column 1 was the result of two opposite effects for men and women. In this specification the estimated coefficients imply that moving from a class with no females to a class with only females increases the probability of female students to enroll in a high earning major by 8.5 (i.e. the sum of the coefficients $\beta_1 + \beta_2$) percentage points with respect to a baseline probability of enrolling high earning majors for women of 20.8 percent. Alternatively, we can derive the effect of share of women on the female bias in choosing high earning majors. Column 2 shows that the female gap in the probability of enrolling a high earning major $(\alpha + \beta_2(share\_female_{ilct}))$ would decrease from -34.2 percentage points to -9.6 percentage points (a difference of 26.5 percentage points) when, other things equal, we consider the extreme change from a class with no other female classmate to one with only female classmates (i.e. from $share\_female_{ilct} = 0$ to $share\_female_{ilct} = 1$). Interestingly the male share (complement to one of the female share) affects the choice of major by male students in similar fashion. Indeed results in column 2 show that male students are 16.1 percentage points more likely to enrol a high earning major when the male share in a class (i.e. $1 - share\_female_{ilct}$) goes from 0 to 1. In column 3 we test whether the effect of the class female ratio is heterogenous for women of different academic quality. We interact the female dummy and the class female share with a dummy equal to one when the individual is in the top quartile of the test score distribution within school and cohort. Results show that there is no evidence of a significant differential effect of the class female share on females at the top of the test score distribution. The second dimension of heterogeneity/non linearity might show up across the distribution of female shares across classes. In order to check whether the effect of class female share on students is non linear, we split the distribution of female shares across classes [11] between those below and those above the median in Column 4 of Table 8 and then we split the distribution into quartiles in Column 5. The estimates in column 4 show that females facing female ratios above the median are 4.1 (7-2.9) percentage points more likely to enrol in high earning majors, while males are 2.9 less likely. In column 5 we include dummies for each quartile of the individual female ratio distribution (we exclude the lowest quartile) and we interact them with the female dummy. The estimates show that the effect of class female share on the probability of enrolling in high earning majors for female students is monotonically increasing. The difference between top and bottom quartiles of female ratio is 5.7 (0.127-0.07) percentage points with respect to a baseline probability of 20.8% for women. The effect for male students is exactly symmetric: the probability of enrolling in high earning majors for men is 6.8 percentage points higher in classes in the bottom quartile of female shares than in classes in the top quartile.

The functional form and the magnitude of the effect of same-sex ratio on probability of enrolling in high

[11] We actually rank individuals with respect to female ratio of their class and choose the median.

earning majors look very similar for men and women. Figure 6 plots the estimated effect of same-sex class-share on probability of enrolling in a high earning major, by quartile for male and female. The function of same-sex share is monotonic, increasing for both genders with a more concave appearance for men and convex for women.

These baseline results are consistent with the literature analyzing the choice of major in all-women colleges. Those studies claim that an environment without men increases the probability of women to choose typically male (STEM) majors. Our estimates show a progressive effect of the class female share on both the choice probability of male and female. Our results also show for the first time (to our knowledge) that a symmetric/opposite effect exists for males: those in high school classes with large male-shares are more likely to choose high earning majors. In Table 8, column 6 we also add dummies for classes with only females and for classes with only males to test if single-sex environments imply an additional discrete increase in the marginal effect relative to classes in the fourth quartile of the female share distribution. Interestingly, the results show that female-only classes do not exhibit this additional effect relative to prevalently female classes. However male-only classes show an additional positive jump in the probability to attend high earning majors (+9.4 percentage points).

6.2 Class female share and college outcomes

In the previous section we presented evidence on the effect of the class female share on the initial enrollment probability in high earning majors for both women and men. However, as students in the considered college preparatory high schools, typically enrolled into college right after graduation, this outcome might simply represent a short-run effect and may be undone by higher drop-out rates, higher rate of major changes or poor performance and attrition in these majors. Moreover if students from classes with large share of same-sex classmates are "pressured" into majors that turn out to be bad matches for their abilities (De Giorgi, Pellizzari, Redaelli 2010), in the long run their academic performance may suffer. It is therefore important to test if the students with large share of same-sex peers in high school who enrolled in high earning majors, continued in these majors until graduation or switched/dropped with higher probability. We also test whether their college performance was relatively worse than the performance of similar enrolled students. These questions are extremely relevant to understand whether the same-sex share in high school has only a short-run effect or if it affects the long-run perspective of students. Thanks to our unique dataset we are able to test the effect of gender composition on several different college outcomes.

In Table 9, which has otherwise an identical structure as 8, we use a dummy for graduating in high earning majors (rather than for enrolling in high earning majors) as outcome in our main specification. As graduation took on average four to five years after enrollment this already represents a long-run outcome. Out of the 17,346 individual enrolling college in our sample, only 13,000 actually graduated from college. Attrition from College was and is quite relevant in Italy and for this specific sample. The estimates show that higher same-sex class-share significantly increased the probability of graduating from high earning majors. In our preferred specification, Column 5 of Table 9, women from classes in the fourth quartile of the class female-share distribution were 4.4 percentage points more likely to graduate from high earning majors[12]. This value is very similar to their higher probability of *enrolling* in high earning majors enrollment shown in Table 8 and equal to 5.7 percentage points. The only evident difference between the enrollment and the graduation outcomes is for male-only classes. While men from male-only classes were more likely to enroll in high paying major than men from the top quartile of male-share classes, their graduation rate does not show this extra-effect. The possible "confidence boost" that pushes male students in male-only classes to enroll in high earning majors may draw into them some students whose skills are not appropriate (well matched) for these majors.

In Table 10 we restrict our specification to students who enrolled in high earning majors and analyze whether the drop-out rate of those exposed to high female shares in high school is higher. We consider as dropouts those students who enrolled in college and have not graduated by the time of our data collection. Results show absolutely no evidence of a higher drop-out rate of students from high-school classes with large same-sex shares. In columns 6 and 7 of Table 10 we also control for specific major and university dummies (within the high earning group) as some Universities and majors can be more selective. Results are not changed.

In Table 11 and 12 we repeat the same exercise as in Table 10 using time to graduation and college exit score, respectively, as outcomes. For these two outcomes the analysis is restricted to those who graduated in high earning majors. Hence the estimated effects will reveal if women coming from high female share classes perform worse than similar women in high paid majors. Results in Table 11 show no evidence of longer time to graduation (measured in months) for students coming from high school classes with large same-sex shares, even after controlling for major and university dummies (different universities and different majors have very different expected time to graduation). In Table 12 we look at college exit score of those who graduated from high earning majors. This score is a composite measure of college G.P.A. and of a final dissertation[13] and is comparable only within major and university. It is thus important to look at columns 6 and 7 of Table 13 where we control for university and major dummies. Exit college scores show only minimal under-performance of women from high female share classes (-3.3 percentage points). There is also slight under-performance of male students from male-only high school classes (-2.2 percentage points). This result possibly confirms the evidence in Table 9 that students from male-only classes may increased the probability of enrollment in high earning major even when their skills might not be the best match for those majors.

[12]This value is obtained by adding the coefficients on the 4th quartile (-0.073) with the one on the interaction of female and the fourth quartile (0.117).

[13]We re-scale the score to be 0 and 1.

6.3 Class female share and long-run earnings

Results in Table 9 to 12 have shown that the gender composition in the high school class affected student choice after high school, and this effect persists through college up to graduation. Are these effects also discernible on the labor market outcomes of those students? To assess this we use the fact that we have matched the school career data for those individuals with the income data as of 2005. In Table 13 we estimate our main specification using the logarithm of income in 2005 as dependent variable. We restrict our sample to students who enrolled in high earning majors after graduating from high school between 1985 and 2000 [14]. The estimates on the interaction between the top quartile female share and the female dummy capture the differences between the earnings of the average individual enrolling high earning majors and those women who attended high school class with large female shares.

In column 1 the gender gap in log of wage conditional on skills and ability characteristics and within high earning major graduates is -0.14 and the average effect of female ratio across gender is not statistically different from zero. In column 2 we let female ratio to interact with the dummy female and it is clear that the zero average effect of female ratio in column 1 was the result of averaging opposite effects for men and women. Furthermore, individual expected log earnings increases in the high school class same-sex ratio: moving from a class with female-ratio equal 0 to a class with female-ratio equal 1 makes female expected log income increase by 8.4% (8.7% in income levels) and male expected wage drop by 29.4% (25.5% in level wage). Alternatively we can look at the gender gap and estimate that it ranges from a -26% in level expected wage when female share is 0 to +4.5% for female share equal 1. Even though we are extrapolating this range from a linear function in female-ratio, it is still remarkable that just due to the peer gender effect, for very high female share classes, the gender gap turns positive. In columns 3 to 7 of Table 13 we estimate heterogenous effects of female ratio across the distribution of female share and we control for major and university dummies to get rid of differences in expected wage across them. Column 7, our preferred specification, shows that females from high school classes in the top quartile of the female-ratio distribution (and who graduated from high earning majors) have expected wage 14.6 percentage points higher than females from high school classes in the bottom quartile. This estimate reinforces the findings that exposure to a high female-share class in high school adds a positive effect to academic and labor market productivity, for those who choose a high earning major.

7 Robustness Checks

In our dataset we do not have information on pre-high school ability of students. Hence, in order to control for relative skills in our main specification we use the relative ranking within school and cohort in the

[14] We exclude individuals that did not graduate yet when we observe their wage.

standardized exit test score. One might be concerned that class female ratio might endogenously change the relative performance of women in the class. If this class-level endogenous effect were to influence the relative ranking in performance across classes within school and cohorts, then including the ranking as control in our main specification might confound the estimated treatment effect of female share on choice of major. We thus run our main specification and replicate results of Table 8 excluding the relative ranking from the controls. Table 14 shows that results are robust to the exclusion of the relative ranking control when compared with results of Table 8. This suggests that individual relative ranking within school and cohort is not endogenously affected by the female ratio in the class. Hence the channel we are identifying is independent from endogenous changes in relative academic performance. The share of same-sex classmates, especially for women, affects the choice of major and the following college and labor market performance without affecting relative high school performance per se. We can think of this effect as a "confidence boost" for women in a less gender stereotyped environment. Our data suggest that this may bring persistent change in the potential productivity and human capital accumulation of women.

A concern would arise if one were to suspect that families with specific characteristics, associated with better (or worse) outcomes of their children, were to influence the assignment of their kids to classes with specific female shares. Interestingly (and uniquely) in our dataset we can identify siblings who graduated from one of Milan high schools (no matter if they attended the same institution or not) between 1985 and 2005. We exploit the exact address of where students lived at the time of high school and their last name. In Table 15 we run our main specification (as of Table 8) restricting the analysis to siblings and including family fixed effects. This is a very important check as we are only using the within-family variation of major choice between same-sex siblings. While we were able to control in detail for school, teacher and cohort effects in the previous regressions we had only rough family income controls. The results of Table 15 instead, include the most comprehensive set of family controls (family effects) and are indeed virtually identical to those of Table 8. Unobserved family characteristics do not affect at all our results, implying that they are likely orthogonal, as we already shown for observable characteristics, to the class gender composition.

A final robust check is presented in Table 16, in which we exploit the exogenous variation of female share in each cohort, in each school. To use a different margin of variation of peer gender composition which cannot be affected by choices of the principal or of teachers we consider the female share in each cohort, in each school. Indeed principals take the cohort of kids enrolling their institution as given and determined mainly by demographics in the relevant school district. When a school receives a cohort with large share of women is more likely to have classes with high share of women. In Table 16 we thus replicate our main specification replacing the female ratio in the classroom with the female ratio of the entire school cohort and we control for school fixed effects (rather than school-cohort effects). The effects of the share of female in the cohort are very close to the estimated effects of class female share and reported in Table 8. Both margins of variation of the gender composition (exogenous variation in gender composition of classes within a school and idiosyncratic variation of gender across cohorts) confirm that women from classes with large shares of women chose high earning college major with significantly higher frequency.

8 Framework to Interpret the results

The results presented above are consistent with the idea that a large share of same sex peers in high school increases the freedom and confidence of individuals to choose high paid major. This may be because peers of the same gender reinforce one's self confidence or because they reduce the gender "component" of a choice allowing individual to choose based on tastes and ability. A possible model to interpret our results is one in which individuals choose highly competitive majors based on their academic quality and on their self confidence. If the high school environment, and the gender ratio in it, affects one's self-confidence this could be the channel through which the choice of major is affected. The following model formalizes simply this idea.

Consider an individual i of gender $f = 1, 0$ (with female= 1) with skill level s, who is choosing a college major m. Conditional to choosing a specific college major m her/his returns Y are a function of his/her own skills

$$Y_{ifsm} = Y_i(f, m, s) : \text{ for : } m = L, H \tag{3}$$

For simplicity we reduce the choice of major to high earning majors ($m = H$) or low earning ($m = L$). We assume the distribution of skills and returns to major to be equal between gender:

$$Y(f = 1, s, m) = Y(f = 0, s, m) : \ foralls : \text{ and for : } m = H, L \tag{4}$$

Function $Y(\cdot)$ is such that $\frac{\partial Y}{\partial s} > 0$ and $\frac{\partial Y^2}{\partial^2 s} < 0$ and by definition $Y(f, s, m = H) > Y(f, s, m = L)$ for every skill level s [15].

We also model the psychological cost of attending major m as a function of their confidence, assuming that the cost is higher if confidence is lower:

$$C^m_{ifs} = C^m_{is}(conf_i|f, s) \tag{5}$$

with $\frac{\partial C}{\partial conf} < 0$.

Given that in our model returns to majors are the same for males and females, the observed equilibrium in the choice of major is the result of differences in the cost of choosing such majors. In this setting, individual i

[15] We can relax this assumption by specifying two different sets of major-specific skills.

will choose major m if the net returns to it are higher than the net returns to any other major. Simplifying this choice to the dichotomous case of high vs. low earning major, the probability p^H_{isf} that individual i of gender f and skill s chooses a high earning major $m = H$ can be expressed as follows:

$$p^H_{ifs} = P[Y^H_{is} - C^H_s(conf_i|f,s) > Y^L_{is} - C^L_s(conf_i|f,s)] \tag{6}$$

with Y_{is} independent of gender f since we assumed returns to a specific major conditional on any given skill level s is the same between gender. Rearranging equation (6) we obtain:

$$p^H_{ifs} = P[Y^H_{is} - Y^L_{is} > C^H_s(conf_i|f,s) - C^L_s(conf_i|f,s)] \tag{7}$$

Given the assumption about the distribution of returns to major and conditional on skill level s and gender f, individual i's probability of choosing a high earning major p^H_{ifs} is thus a function of the cost of attending a high earning major $C^H_s(conf_i|f,s)$, of the cost of attending a low earning major $C^L_s(conf_i|f,s)$ and of an individual-specific idiosyncratic preference ε_i:

$$p^H_{ifs} = h(C^H_s(conf_i|f,s) - C^L_s(conf_i|f,s), \varepsilon_i|f,s) \tag{8}$$

with $\frac{\partial h}{\partial C^H_s} < 0$ and $\frac{\partial h}{\partial C^L_s} > 0$. Hence the difference $C^H_s(conf_i|f,s) - C^L_s(conf_i|f,s)$ decreases as $conf_i$ increases. Assuming that the level of confidence $conf_i$ for each gender is increased by the high school class same-gender share, then this model predicts that higher same-gender share implies higher probability of choosing a high earning major. Substituting $share\_female_{ilct}$ as determinant of $conf_i$ and linearizing (8) we obtain the estimated equation 2.

9 Discussion and Conclusions

In this paper we estimated the effect of the gender composition of high school classes on the choice of college major and then on the performance in college and in the labor market. We simplified the choice of major to a dichotomous one between high earning ones (Engineering, Economics & Business and Medicine) and non-high earning ones. High earning majors have very competitive admission tests, are characterized by a strong male-dominated competition and give access to the best paid occupations in our sample. We exploit the exogenous variation in gender composition of high school classes and we show that the share of same-gender classmates affects significantly and positively the probability of choosing a high earning major. We speculate that this effect could be the result of higher confidence and higher willingness to compete for women who are in mainly female environments, as it has been pointed out by previous experimental literature. The interesting novelty of our result is that we find a similar effect at work for men and that we find long-lasting effects on academic performance and labor market outcomes from being exposed to a large same sex share of peers in high school. Indeed female high school students exogenously assigned to classes with high female shares are on average 5.7 percentage points more likely that other women to enroll in a high earning major with respect to a baseline probability of 20.8 percent (while higher male shares increase male students' probability to enrol high earning majors by the same amount). Women are 4.4 percentage point more likely to graduate from a high earning major. They take one month less to graduate from those majors relative to other women and they earn similar final grades. Finally they have a earning premium of 20%, relative to other women five to twenty years after high school graduation.

Overall, the expected wage premium for a women who attended a high school class with high female share (in the top quartile of the female-share distribution) relative to an identical one who attended a class with a few female classmates (in the bottom quartile of the female-share distribution) is an annual EUR 1,110 (about $ 1,470). This premium is the result of two different effects. An expected wage premium of EUR 760 deriving from the increased probability to graduate in high earning majors (+4.4 percentage points) times the average wage premium of graduating in high earning majors (+89%) with respect to an average annual wage of EUR 19,412 for individuals graduated in non-high earning majors. The second component of the wage premium derives from the fact that among women graduating in high earning majors (with a 4.4 percentage point higher probability), those who attended high school classes with high female ratio have an additional 24% higher wages relative to an average annual wage of EUR 33,176 for individuals graduated in high earning majors, which amounts to EUR 350. The sum of these two effects is EUR 1,110. This expected wage premium of EUR represents an expected 5.3% wage increase (with respect to an average annual wage of EUR 18,983.65 for college graduated women) for women exogenously attending high school classes with high female share. This is a return as large as the return of one extra year of schooling in Italy.

Let us notice, as a final remark, a very simple policy implication of our results. If the policy objective is to increase women (and men's) enrollment in high earning majors, by reducing the psychological costs of enrolling and attending those majors, then our results would suggest that gender segregated classrooms would be an effective step. Gender segregated classrooms would increase the probability of choosing high earning majors for both women and men. At the very least, schools could offer to students the possibility of choosing single sex classes. This policy would probably increase the number of women applying to high earning major, and it would also increase the average quality of the admitted pool of students in those majors.

References

Akerlof, George A., Kranton, Rachel E. (2000). "Economics and identity." Quarterly Journal of Economics 115, 715-753.

Akerlof, George A., Kranton, Rachel E. (2002). "Identity and schooling: some lessons for the economics of education." Journal of Economic Literature 40 (4), 1167-1201

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.

Carrell, Scott E., and Mark L. Hoekstra (2010). "Externalities in the Classroom: How Children Exposed to Domestic Violence Affect Everyone's Kids". American Economic Journal: Applied Economics, 2(1): 211-2

Carrell Scott, Marianne Page and James West (2010). "Sex and 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.

Dee, Thomas S. (2007). "Teachers and the Gender Gaps in Student Achievement." Journal of Human Resources 42(3): 528-554

Favara, Mara (2012). "The cost of acting 'girly'. Gender stereotype and educational choices." Working paper.

Flabbi, Luca (2011). "Gender Differences in Education, Career Choices and Labor Market Outcomes on a Sample of OECD Countries." Organization for Economic Co-operation and Development Background Paper for the WDR 2012.

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.

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

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, Muriel, Vesterlund, Lise (2010). “Explaining the gender gap in math test scores: the role of competition.” Journal of Economic Perspectives 24 (2), 129-144. .

Sacerdote, Bruce L. (2001). “Peer Effects with Random Assignment: Results for Dartmouth Roommates.” Quarterly Journal of Economics 116 (2):681-704.

Nicole Schneeweis, Martina Zweimiller (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.

Schneeweis, Nicole and Martina Zweimuller (2012) ”Girls, girls, girls: Gender composition and female school choice” , in *Economics of Education Review*, Vol. 31, Number 4, page 482-500.

Tidball, Elizabeth M. (1986). ‘Baccalaureate origins of recent natural science doctorates’, Journal of Higher Education 57(6), 606-620.

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.

Xie, Y. and K. Shauman (2003), “Women in Science: Career Processes and Outcomes”, Cambridge, MA: Harvard University Press

Appendix: Details of the Dataset

In this work we present for the first time a unique database collecting information on the high school career, university career and labor market performance of young Italians who have graduated from high school between 1985 and 2005. This dataset involves different sources that have been carefully matched in a complicated merging process. The collection of the database has involved the collaboration of many parties. The help from the following persons and institutions made the collection possible: the directors of the high schools in Milan, the company Ambroscuole the Provincia di Milano, Daniele Checchi (for Universta Statale), Carlo Lucifora (for Universita Cattolica), Francesco Peri (For Universita di Milano, Bicocca), Augusto Sarti and Mauro Santomauro (for Politecnico di Milano). Davide Malacrino and Francesca Barbiero provided excellent assistance in collecting and organizing the data. We summarize the sources of the single datasets merged, the information contained in each dataset and the merging process in figure 10. The diagram representing the merging process must be read from left to right. The core dataset include the universe of all high school graduates attending college-prep schools in the city of Milan between 1985 and 2005 (around 30550 individuals). Data have been collected manually by inputting the information contained in hard copies of the school records. The list of the 13 college-prep high schools in the city of Milan involved in the data collection process by type of school (Classical Studies vs. Scientific Studies) is included in Table A.1.

Among the five major universities of Milan involved in the collection of our data, three are public universities (Universita' degli Studi di Milano, Universita' degli Studi di Milano Bicocca, Politecnico di Milano) and two are private (Universita' Bocconi, Universita Cattolica di Milano) . The first two public universities mentioned have a very broad offer of majors while Politecnico di Milano offers degrees in Engineering, Architecture and Design only. Among the private universities Universita Cattolica di Milano has a broad offer of majors comparable with Universita degli Studi di Milano while Universita Bocconi is a school of Business, Economics and Law. The list of majors offered overall by universities located in Milan is very large. For our analysis we thus aggregated all these majors in 11 broader Fields of Study as shown in Table A.2.

Figures

Figure 1: School structure and identification strategy

	School 1				⋯	School N			
Group of teachers Cohorts	1	2	⋯	7	⋯	1	2	⋯	9
1985	A	B	⋯	G	⋯	A	B	⋯	I
1986	A	B	⋯	G	⋯	A	B	⋯	I
1987	A	B	⋯	G	⋯	A	B	⋯	I
⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯
2005	A	B	⋯	G	⋯	A	B	⋯	I

Note: dashed rectangles identify the levels for which we include dummies in our identification strategy.

Figure 2: High school class female share distribution

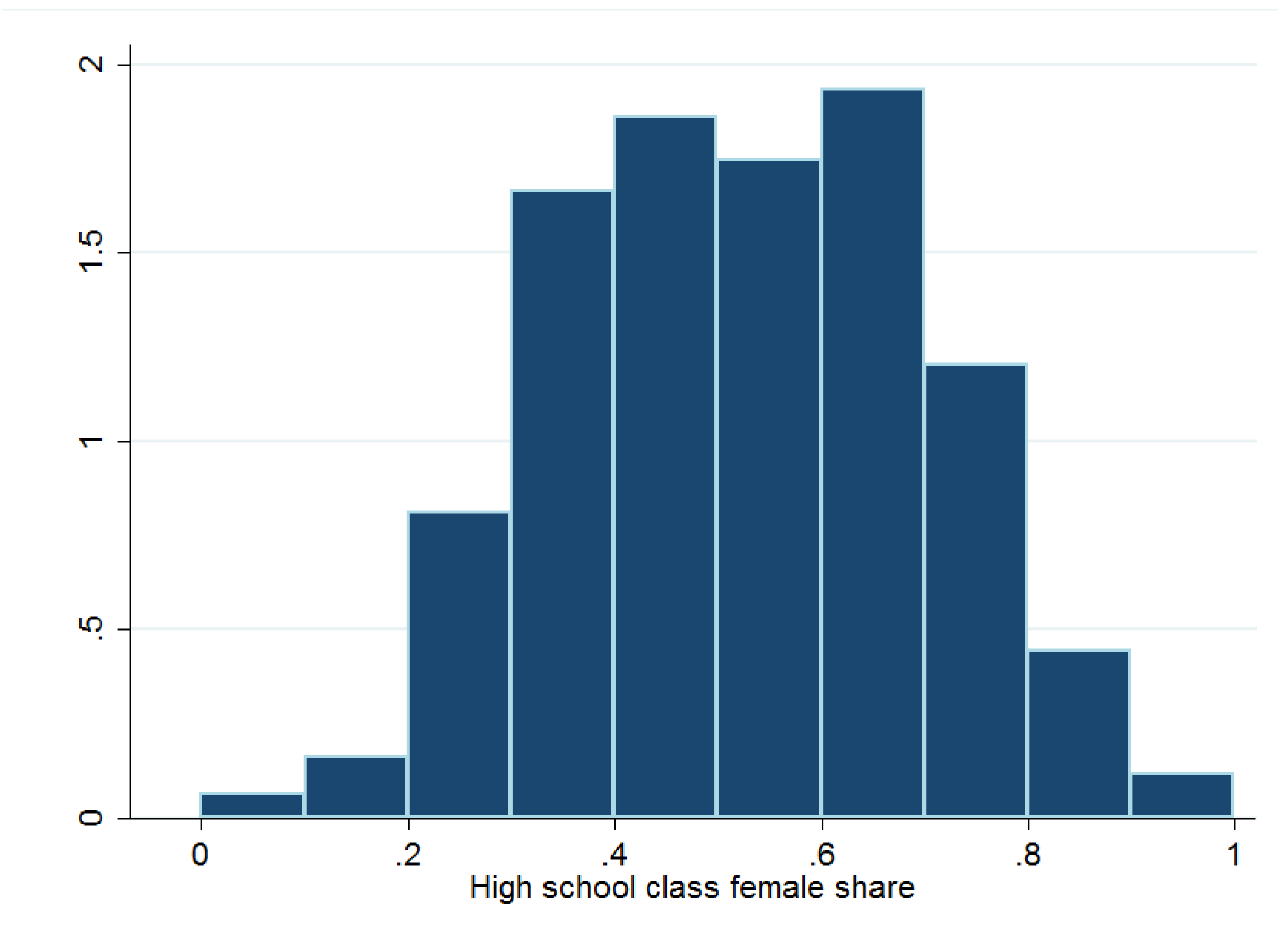


Figure 3: The high school class female share distribution is normal

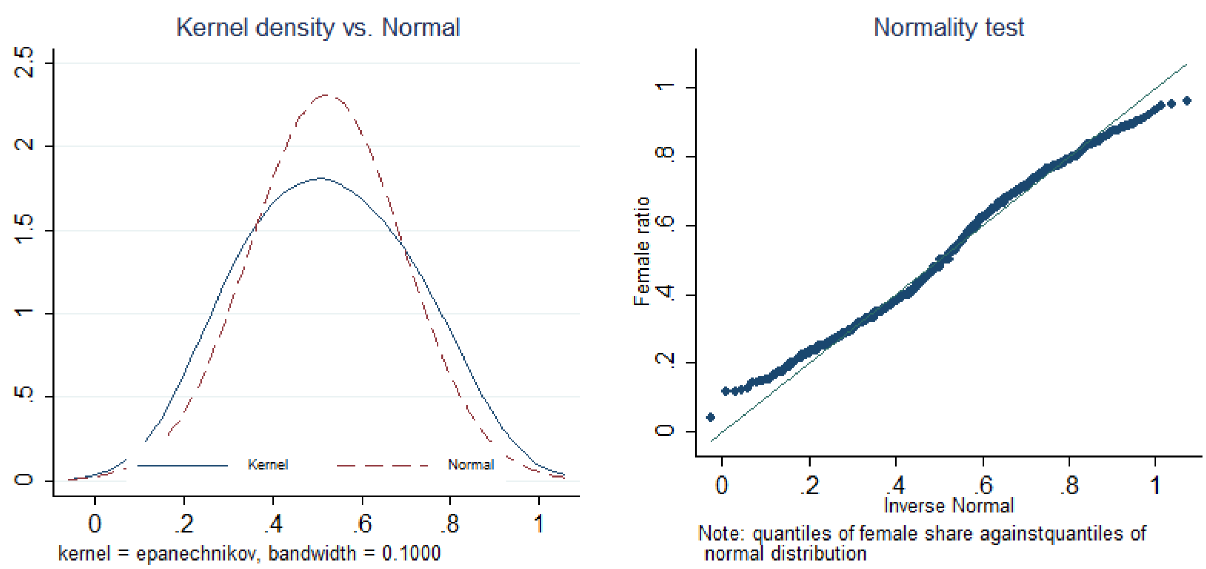


Figure 4: High school class female share autocorrelation within group of teachers across cohorts

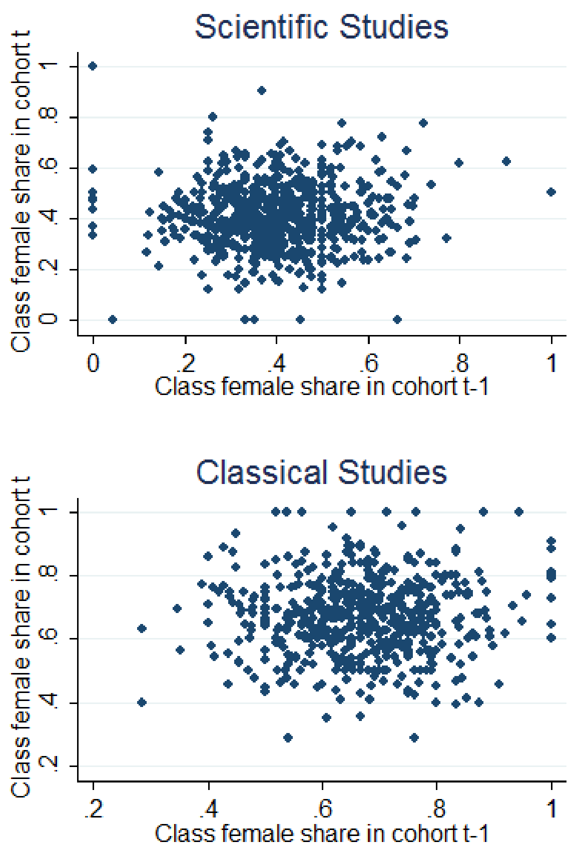


Figure 5: High school class female share vis-a-vis class characteristics

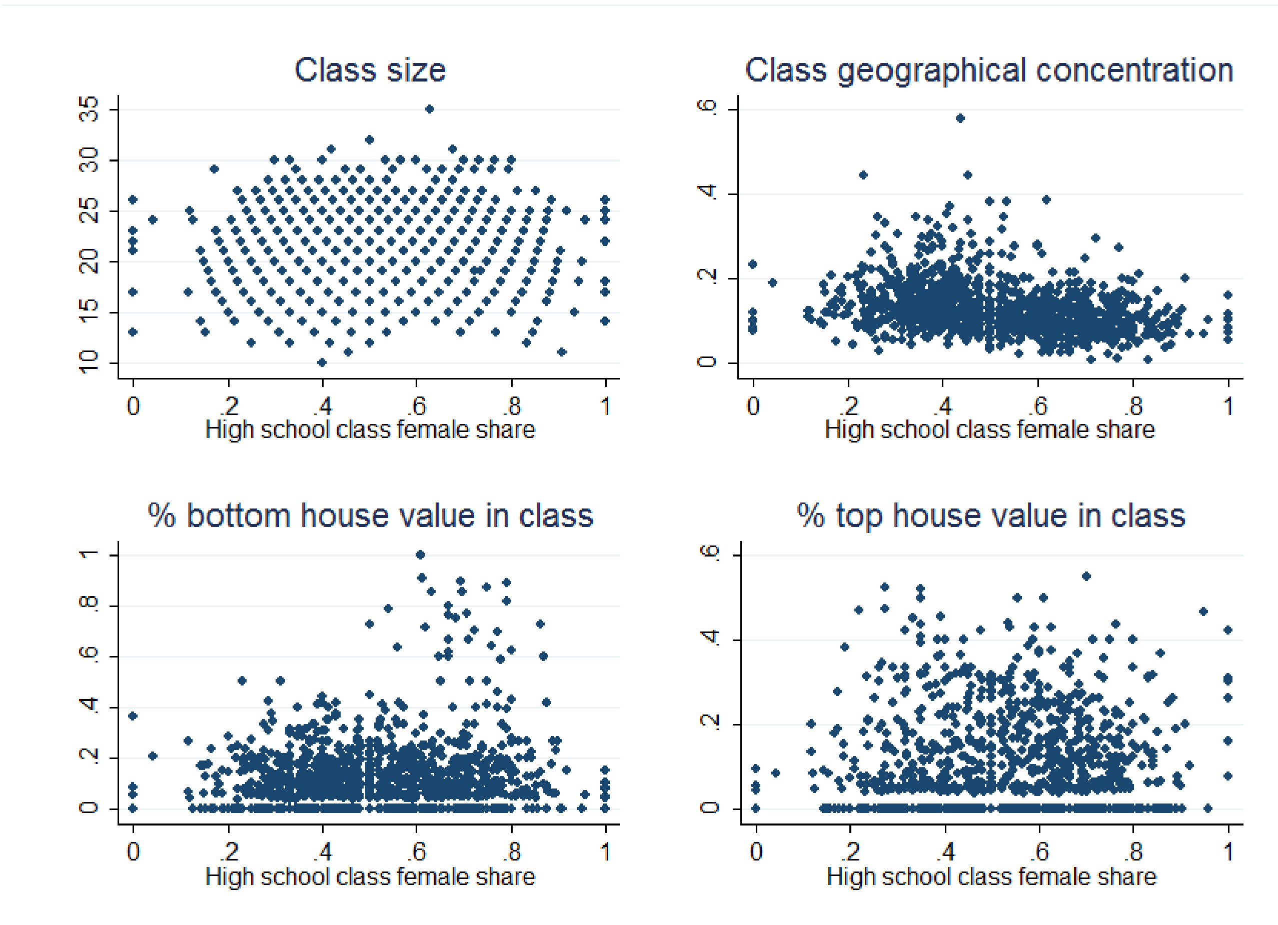


Figure 6: Treatment effect of high school class same-sex share on probability of enrolling in high earning majors by quartile of same-sex share

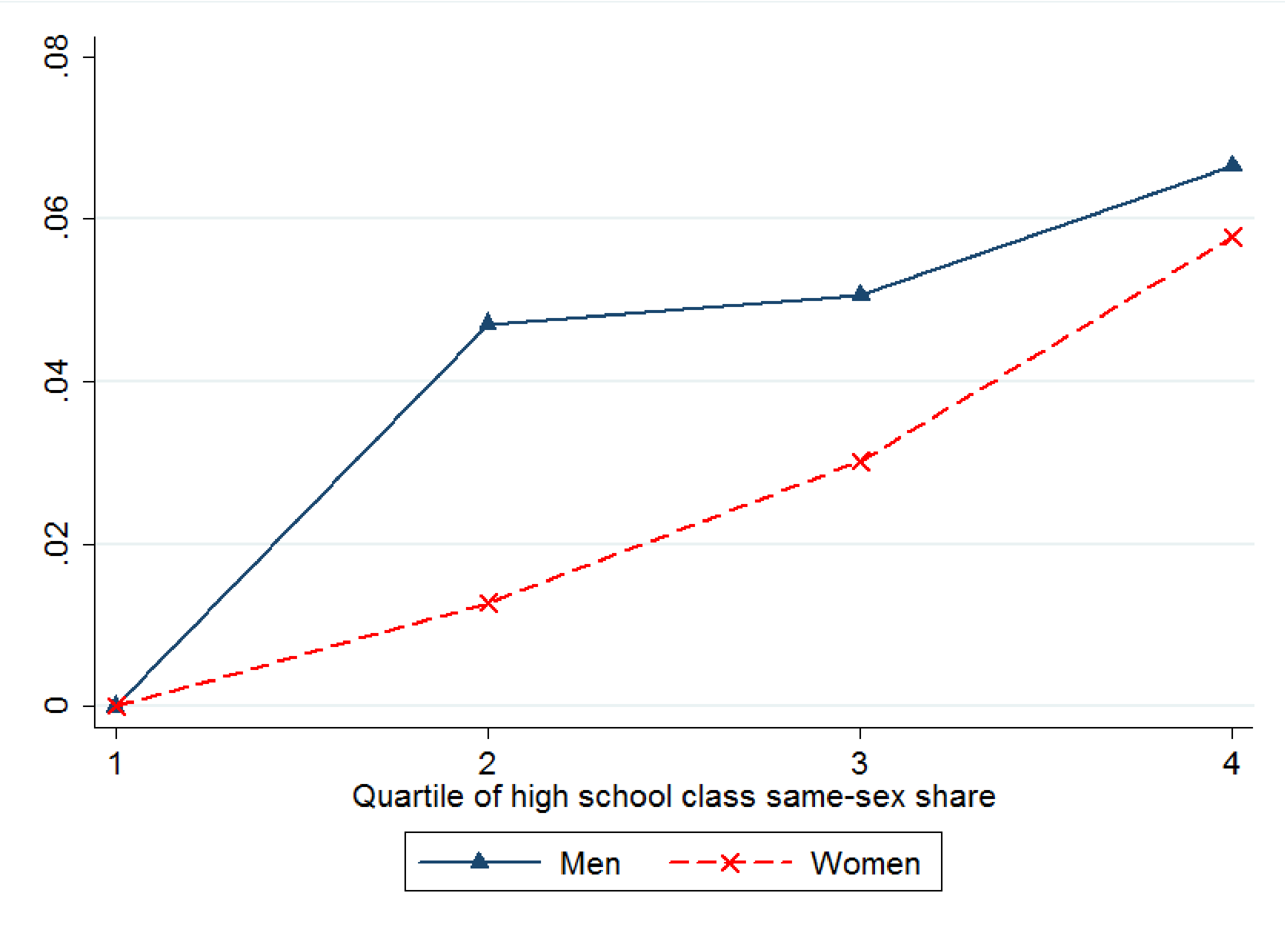
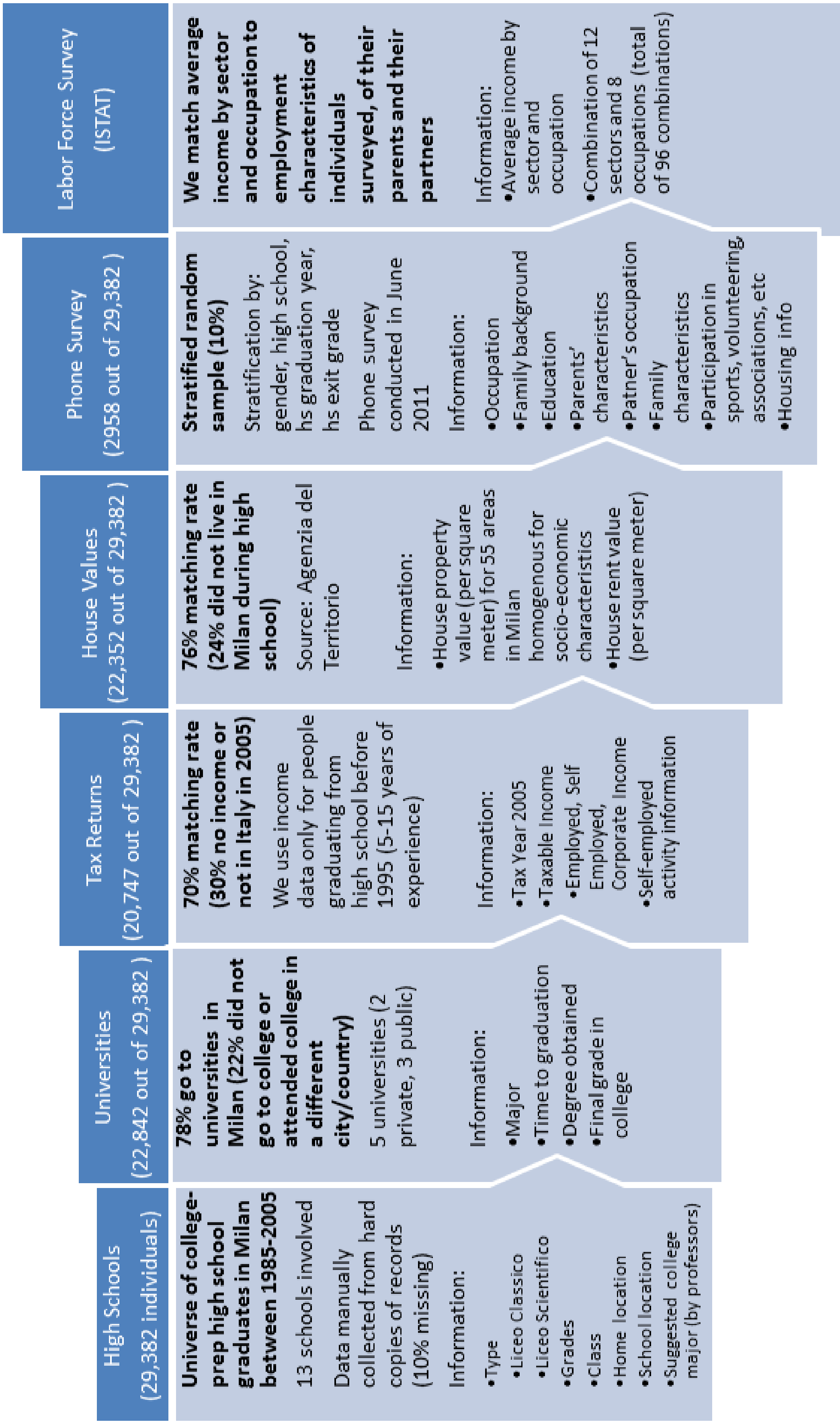


Figure A.1: Dataset Structure



Tables

Table 1: Summary Statistics

Variable	Obs	Mean	All Std.Dev.	Min	Max	Women Mean	Men Mean	Difference | T-stat |
Individual Variables:								
Female	29382	0.523	0.499					
High School Variables:								
High school exit standardized score	29382	0.416	0.299	0	1	0.435	0.395	11.5
Within school/cohort rank	29382	0.495	0.313	0	1	0.516	0.471	12.4
College Variables:								
Enrolled in High Paying majors (on total Enrolled)	22842	0.325	0.469	0	1	0.208	0.452	40.1
Graduated in High Paying majors (on total Graduated)	17050	0.360	0.480	0	1	0.239	0.499	36.7
High Paying majors dropout	7431	0.175	0.380	0	1	0.132	0.197	6.9
College time to graduation	16015	81.124	26.386	32	295	79.781	82.669	6.9
College exit score	16585	0.848	0.155	0	1	0.880	0.810	29.6
Pre-treatment Variables:								
Log(house value)	22352	8.049	0.315	7.409	9.143	8.046	8.052	1.4
Outcome Variables:								
Log(wage)	17008	9.679	1.279	0.693	13.746	9.461	9.906	22.1
Top Occupation	2958	0.367	0.482	0	1	0.308	0.432	7.1
High School Class Variables:								
Class size	1371	21.432	3.801	10	35			
Female share	1371	0.520	0.183	0	1			
Average high school exit score	1371	0.419	0.104	0.118	0.780			
Average log(House Value)	1218	8.042	0.151	7.619	8.498			
% students in bottom decile of house value distribution	1218	0.119	0.139	0	1			
% students in top decile of house value distribution	1218	0.089	0.113	0	0.55			

Table 2: The importance of STEM and High Earning majors on income and career

VARIABLES	(1) Log(wage)	(2) Log(wage)	(3) Log(wage)	(4) Top Occupation=1	(5) Top Occupation=1	(6) Top Occupation=1
Graduated in High Earning major=1			0.639***			0.121***
			(0.031)			(0.033)
H.S. exit score	0.409***	0.409***	0.078	-0.039	-0.039	-0.083
	(0.047)	(0.047)	(0.049)	(0.054)	(0.054)	(0.053)
College exit score	-0.273***	-0.273***	0.310***	-0.097	-0.097	-0.041
	(0.076)	(0.076)	(0.082)	(0.094)	(0.094)	(0.097)
House value in top 10%	0.098*	0.098*	0.085*	0.072	0.072	0.060
	(0.052)	(0.052)	(0.050)	(0.072)	(0.072)	(0.072)
House value in bot 10%	0.001	0.001	0.024	-0.051	-0.051	-0.048
	(0.042)	(0.042)	(0.038)	(0.045)	(0.045)	(0.043)
Constant	10.347***	10.347***	9.784***	0.691***	0.691***	0.629***
	(0.063)	(0.063)	(0.069)	(0.079)	(0.079)	(0.083)
Observations	8,363	8,363	8,354	1,460	1,460	1,457
R-squared	0.193	0.193	0.244	0.140	0.140	0.152
Sample	All Graduates	All Graduates	All Graduates	Interviewed	Interviewed	Interviewed
School X Cohort 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.
Sample: for spec. 1-3 students graduated from high school between 1985 and 2000 and completing college,for spec. 4-6 students graduated from high school between 1985 and 2000 and completing college and interviewed with a phone survey in year 2011.
Dependent variable: for spec. 1-3 logarithm of personal income, as revealed to the internal revenue service in year 2005,for spec. 4-6 a dummy for probability of working in a top occupation, defined as managers, professionals, business owners in our phone survey.

Table 3: Gender differential

	Dependent variable: Log of actual wage 1985-2000						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
VARIABLES							
(Female=1)	-0.382***	-0.356***	-0.380***	-0.422***	-0.395***	-0.395***	-0.286***
	(0.024)	(0.025)	(0.028)	(0.027)	(0.029)	(0.029)	(0.029)
H.S. exit score				0.371***	0.361***	0.361***	0.083**
				(0.038)	(0.040)	(0.040)	(0.041)
College exit score				0.071	0.090	0.090	0.512***
				(0.074)	(0.077)	(0.077)	(0.083)
Graduated in High Earning major=1							0.579***
							(0.028)
Constant	10.316***	10.320***	10.418***	10.269***	10.260***	10.260***	9.772***
	(0.049)	(0.055)	(0.049)	(0.067)	(0.063)	(0.063)	(0.071)
Observations	17,008	14,144	10,733	10,401	10,401	10,401	10,392
R-squared	0.182	0.188	0.216	0.223	0.239	0.239	0.277
Sample	All	Enrolled	Graduated	Graduated	Graduated	Graduated	Graduated
Cohort FE	X	X	X	X			
School X Cohort FE					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.
Dependent variable: logarithm of personal income, as revealed to the internal revenue service in year 2005.

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 house value	0.009			
	(0.019)			
Cohort mean house value	-95.427***			
	(8.161)			
Peers' mean dist from school		-0.014		
		(0.020)		
Cohort mean dist from school		-119.657***		
		(8.988)		
% class in top 10%			0.013	
			(0.020)	
% cohort in top 10%			-109.847***	
			(13.567)	
%class in bot 10%				-0.037
				(0.039)
% cohort in bot 10%				-83.493***
				(8.291)
Constant	748.840***	829.563***	-0.093	65.124***
	(63.343)	(61.853)	(0.077)	(6.391)
Observations	22,351	26,106	22,351	22,351
R-squared	0.789	0.832	0.830	0.706
School X Cohort FE	X	X	X	X
SchoolXTeachersX5years 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: all high school students graduating in Milan between 1985 and 2005 for which the address information is non-missing
Dependent variable: for spec. 1 house market value where the students used to live at the time they attended high school, for spec. 2 linear distance from school to the house where the students used to live at the time they attended high school in meters,for spec 3 a dummy = 1 if the student used to live in a house valued in top decile of the house value distribution, for spec. 4 a dummy = 1 if the student used to live in a house valued in top decile of the house value distribution.

Table 5: Auto-correlation of class female share within school and group of teachers

Dependent variable: High School Class Female Share			
	(1)	(2)	(3)
VARIABLES			
Class female share t-1	0.022	0.036	-0.058
	(0.046)	(0.056)	(0.047)
Constant	0.650***	0.663***	0.723***
	(0.030)	(0.041)	(0.032)
Observations	1,268	1,268	1,268
R-squared	0.537	0.609	0.648
Sample	Classes	Classes	Classes
School FE	X		
SchoolXCohort FE		X	X
SchoolXTeachers FE			X

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 classroom of a given school in a given year.

Dependent variable: high school class female share in cohort t.

Table 6: High school class female share and other class characteristics

VARIABLES	(1) Female share	(2) Class in 1st Quart of fem share	(3) Class in 2nd Quart of fem share	(4) Class in 3rd Quart of fem share	(5) Class in 4th Quart of fem share
Class % in bottom 10 house value	0.069	-0.162	-0.057	-0.090	0.309
	(0.075)	(0.172)	(0.214)	(0.400)	(0.403)
Class % in top 10 house value	0.032	0.030	-0.330	0.466*	-0.166
	(0.070)	(0.088)	(0.313)	(0.255)	(0.136)
Class size	-0.001	-0.006	0.009	0.002	-0.006
	(0.002)	(0.005)	(0.007)	(0.008)	(0.006)
Class geographical concentration	0.038	-0.291	0.136	0.179	-0.024
	(0.097)	(0.386)	(0.598)	(0.320)	(0.148)
Constant	0.772***	0.194	-0.296	0.355*	0.746***
	(0.062)	(0.153)	(0.170)	(0.194)	(0.167)
Observations	1,218	1,218	1,218	1,218	1,218
R-squared	0.717	0.549	0.422	0.455	0.595
School X Cohort FE	X	X	X	X	X
School X Teachers FE	X	X	X	X	X

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 of a given school in a given year
Dependent variable: class female share for spec. 1, each of the four quartiles of the female share distribution for specifications 2-5.
Independent variables: share of students in the classroom who used to live in a house valued in the top or bottom decile of the house value distribution, total number of students in the classroom, Herfindhal index for geographical concentration of students in the classroom.

Table 7: High school class female share and class academic quality

VARIABLES	(1) Mean high school score	(2) Mean female high school score	(3) Fem share in 1st high school score quartile	(4) Fem share in 2nd high school score quartile	(5) Fem share in 3rd high school score quartile	(6) Fem share in 4d high school score quartile
Fem share	0.041	-0.021	0.047	-0.059	0.030	-0.019
	(0.025)	(0.039)	(0.032)	(0.041)	(0.059)	(0.069)
Constant	0.284***	0.291***	0.047***	0.415***	0.360***	0.178***
	(0.011)	(0.017)	(0.014)	(0.018)	(0.026)	(0.030)
Observations	1,371	1,363	1,363	1,363	1,363	1,363
R-squared	0.576	0.528	0.370	0.370	0.328	0.351
School X Cohort FE	X	X	X	X	X	X
School X Teachers FE	X	X	X	X	X	X

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 of a given school in a given year.
Dependent variable: for spec. 1 average high school exit score in the class, for spec.2 average exit score for females only, for spec. 3-6 each a dummy = 1 if the the average score of the class is each of the four quartiles of the class average score distribution.
Independent variable: class female share.

Table 8: The effect of high-school class female share on students' major choice

Dependent variable: High Earning major choice at first college enrollment=1							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
VARIABLES							
(Female=1)	-0.217***	-0.342***	-0.348***	-0.249***	-0.263***	-0.261***	-0.261***
	(0.009)	(0.024)	(0.024)	(0.011)	(0.014)	(0.014)	(0.014)
Female share	-0.037	-0.161***	-0.136***				
	(0.030)	(0.036)	(0.038)				
(Fem share X) (Female=1)		0.246***	0.265***				
		(0.044)	(0.045)				
(Fem share)X(4th Rank Quart=1)			-0.104				
			(0.073)				
FemX(4th Rank Quart=1)			0.019				
			(0.056)				
(Fem share)XFemX(4th Rank Quart=1)			-0.068				
			(0.102)				
(Fem share above median)=1				-0.029**			
				(0.014)			
(Female=1)X(Fem share above median=1)				0.070***			
				(0.015)			
(Fem_share2quartile=1)					-0.019	-0.016	-0.016
					(0.014)	(0.015)	(0.015)
(Fem_share3quartile=1)					-0.023	-0.019	-0.020
					(0.017)	(0.017)	(0.017)
(Fem_share4quartile=1)					-0.070***	-0.067***	-0.068***
					(0.021)	(0.021)	(0.022)
(Female=1)X(Fem_share2quartile=1)					0.030	0.029	0.029
					(0.021)	(0.021)	(0.021)
(Female=1)X(Fem_share3quartile=1)					0.052***	0.050**	0.050**
					(0.019)	(0.019)	(0.019)
(Female=1)X(Fem_share4quartile=1)					0.127***	0.124***	0.125***
					(0.022)	(0.022)	(0.022)
Female-only class=1						0.014	0.011
						(0.034)	(0.035)
Male-only class=1						0.094**	0.093**
						(0.046)	(0.046)
Class size							-0.001
							(0.001)
Class geographical concentration							0.018
							(0.085)
Class % in bottom 10 house value							0.038
							(0.058)
Class % in top 10 house value							0.030
							(0.058)
2nd Rank Quartile=1	0.085***	0.085***	0.085***	0.085***	0.085***	0.086***	0.086***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
3rd Rank Quartile=1	0.172***	0.172***	0.172***	0.172***	0.172***	0.172***	0.172***
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
4th Rank Quartile=1	0.273***	0.273***	0.336***	0.273***	0.273***	0.273***	0.273***
	(0.012)	(0.012)	(0.040)	(0.012)	(0.012)	(0.012)	(0.012)
House value in top 10%	0.012	0.014	0.013	0.014	0.015	0.014	0.013
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
House value in bot 10%	-0.004	-0.005	-0.004	-0.005	-0.005	-0.005	-0.007
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.012)
Constant	0.287***	0.336***	0.324***	0.289***	0.292***	0.292***	0.292***
	(0.009)	(0.013)	(0.016)	(0.007)	(0.009)	(0.009)	(0.054)
Observations	17,346	17,346	17,346	17,346	17,346	17,346	17,346
R-squared	0.184	0.186	0.187	0.186	0.186	0.186	0.187
School X Cohort FE	X	X	X	X	X	X	X
SchoolXTeachersX5years FE	X	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.
Sample: high school graduates enrolling in college.
Dependent variable: dummy=1 if student enrolls one of the high earning majors (Engineering, Economics & Business and Medicine) at first college enrollment.

Table 9: The effect of high-school class female share on students' probability of graduation in High Earning majors

Dependent variable: latest graduation in High Earning majors=1

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
(Female=1)	-0.223***	-0.340***	-0.360***	-0.254***	-0.263***	-0.262***
	(0.010)	(0.026)	(0.027)	(0.013)	(0.016)	(0.016)
Female share	-0.057*	-0.175***	-0.154***			
	(0.034)	(0.040)	(0.043)			
(Fem share X) (Female=1)		0.231***	0.274***			
		(0.049)	(0.048)			
(Fem share)X(4th Rank Quart=1)			-0.073			
			(0.077)			
FemX(4th Rank Quart=1)			0.065			
			(0.058)			
(Fem share)XFemX(4th Rank Quart=1)			-0.144			
			(0.108)			
(Fem share above median)=1				-0.036**		
				(0.015)		
(Female=1)X(Fem share above median=1)				0.070***		
				(0.019)		
(Fem_share2quartile=1)					-0.026	-0.024
					(0.018)	(0.018)
(Fem_share3quartile=1)					-0.040**	-0.038**
					(0.019)	(0.019)
(Fem_share4quartile=1)					-0.076***	-0.073***
					(0.022)	(0.023)
(Female=1)X(Fem_share2quartile=1)					0.018	0.017
					(0.025)	(0.025)
(Female=1)X(Fem_share3quartile=1)					0.049**	0.048**
					(0.024)	(0.024)
(Female=1)X(Fem_share4quartile=1)					0.118***	0.117***
					(0.025)	(0.025)
Female-only class=1						-0.013
						(0.035)
Male-only class=1						0.049
						(0.043)
2nd Rank Quartile=1	0.071***	0.071***	0.071***	0.071***	0.071***	0.071***
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
3rd Rank Quartile=1	0.160***	0.160***	0.160***	0.160***	0.160***	0.160***
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
4th Rank Quartile=1	0.250***	0.250***	0.294***	0.250***	0.250***	0.250***
	(0.012)	(0.012)	(0.040)	(0.012)	(0.012)	(0.012)
House value in top 10%	0.011	0.012	0.011	0.012	0.013	0.013
	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)
House value in bot 10%	0.004	0.003	0.004	0.004	0.003	0.003
	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
Constant	0.040	0.124*	0.093	0.039	0.082	0.080
	(0.070)	(0.071)	(0.072)	(0.067)	(0.066)	(0.066)
Observations	13,000	13,000	13,000	13,000	13,000	13,000
R-squared	0.191	0.193	0.193	0.192	0.193	0.193
School X Cohort FE	X	X	X	X	X	X
SchoolXTeachersX5years 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.
Sample: high school graduates enrolling in college.
Dependent variable: dummy=1 if student graduates from one of the high earning majors (Engineering, Economics & Business and Medicine) at the latest graduation.

Table 10: The effect of high-school class female share on students' probability of dropping out from High Earning majors

	Dependent variable: student's probability of dropping a High Earning major						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
VARIABLES							
(Female=1)	-0.049***	-0.070**	-0.052***	-0.063***	-0.063***	-0.038**	-0.029*
	(0.011)	(0.030)	(0.015)	(0.017)	(0.018)	(0.017)	(0.017)
Female share	0.001	-0.014					
	(0.041)	(0.046)					
(Fem share X) (Female=1)		0.046					
		(0.058)					
(Fem share above median)=1			-0.008				
			(0.017)				
(Female=1)X(Fem share above median=1)			0.008				
			(0.022)				
(Fem_share2quartile=1)				0.003	0.002	0.002	0.001
				(0.016)	(0.016)	(0.016)	(0.015)
(Fem_share3quartile=1)				-0.002	-0.003	-0.006	-0.004
				(0.020)	(0.020)	(0.020)	(0.019)
(Fem_share4quartile=1)				-0.015	-0.016	-0.013	-0.019
				(0.025)	(0.025)	(0.025)	(0.024)
(Female=1)X(Fem_share2quartile=1)				0.026	0.026	0.030	0.014
				(0.028)	(0.028)	(0.029)	(0.028)
(Female=1)X(Fem_share3quartile=1)				0.022	0.022	0.028	0.011
				(0.028)	(0.028)	(0.028)	(0.028)
(Female=1)X(Fem_share4quartile=1)				0.012	0.012	0.007	-0.004
				(0.032)	(0.033)	(0.032)	(0.031)
Female-only class=1					0.017	0.002	0.016
					(0.090)	(0.088)	(0.085)
Male-only class=1					-0.010	0.000	0.003
					(0.033)	(0.031)	(0.030)
2nd Rank Quartile=1	-0.079***	-0.079***	-0.079***	-0.079***	-0.079***	-0.086***	-0.074***
	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)	(0.017)
3rd Rank Quartile=1	-0.170***	-0.170***	-0.170***	-0.169***	-0.169***	-0.181***	-0.160***
	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)
4th Rank Quartile=1	-0.224***	-0.224***	-0.224***	-0.223***	-0.223***	-0.241***	-0.215***
	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)
House value in top 10%	-0.020	-0.019	-0.019	-0.020	-0.020	-0.017	-0.007
	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)	(0.014)
House value in bot 10%	0.019	0.019	0.019	0.019	0.019	0.022	0.009
	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.015)
Constant	0.352***	0.359***	0.360***	0.362***	0.363***	0.329***	0.864***
	(0.039)	(0.040)	(0.034)	(0.037)	(0.037)	(0.039)	(0.049)
Observations	5,807	5,807	5,807	5,807	5,807	5,807	5,807
R-squared	0.141	0.141	0.141	0.141	0.141	0.153	0.200
School X Cohort FE	X	X	X	X	X	X	X
SchoolXTeachersX5years FE	X	X	X	X	X	X	X
Major dummy						X	X
University dummy							X

Method: OLS, Standard errors clustered at school/cohort level in parenthesis, *** $p<0.01$, ** $p<0.05$, * $p<0.1$.
Sample: high school graduates enrolling in High Earning majors.
Dependent variable: dummy = 1 if the student enrolled one of the High Earning majors (Engineering, Economics & Business and Medicine) and dropout.

Table 11: The effect of high-school class female share on students' time to graduation

	Dependent variable: time to graduation						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
VARIABLES							
(Female=1)	-3.512***	-3.557*	-3.572***	-3.245***	-3.188***	-1.080	-1.201
	(0.719)	(1.876)	(0.850)	(0.978)	(0.991)	(1.045)	(1.048)
Female share	2.649	2.617					
	(3.234)	(3.589)					
(Fem share X) (Female=1)		0.095					
		(4.063)					
(Fem share above median)=1			0.921				
			(1.180)				
(Female=1)X(Fem share above median=1)			0.171				
			(1.521)				
(Fem_share2quartile=1)				0.529	0.604	0.895	0.944
				(1.038)	(1.062)	(1.047)	(1.032)
(Fem_share3quartile=1)				1.560	1.675	1.740	1.749
				(1.443)	(1.456)	(1.411)	(1.396)
(Fem_share4quartile=1)				-0.273	-0.201	0.320	0.265
				(1.857)	(1.884)	(1.789)	(1.794)
(Female=1)X(Fem_share2quartile=1)				-0.717	-0.747	-0.631	-0.607
				(1.674)	(1.678)	(1.600)	(1.598)
(Female=1)X(Fem_share3quartile=1)				-1.001	-1.055	-0.968	-1.152
				(2.099)	(2.103)	(2.030)	(2.028)
(Female=1)X(Fem_share4quartile=1)				0.986	0.679	-0.150	-0.249
				(2.234)	(2.265)	(2.146)	(2.181)
Female-only class=1					6.138	6.006	6.429
					(4.814)	(4.919)	(4.820)
Male-only class=1					2.176	2.773	2.792
					(2.785)	(3.130)	(3.224)
2nd Rank Quartile=1	-3.203***	-3.204***	-3.175***	-3.163***	-3.137***	-4.037***	-3.887***
	(1.131)	(1.131)	(1.133)	(1.124)	(1.122)	(1.090)	(1.098)
3rd Rank Quartile=1	-5.889***	-5.889***	-5.893***	-5.871***	-5.846***	-7.591***	-7.060***
	(1.161)	(1.161)	(1.158)	(1.157)	(1.159)	(1.162)	(1.183)
4th Rank Quartile=1	-11.259***	-11.259***	-11.259***	-11.233***	-11.208***	-14.034***	-13.278***
	(1.104)	(1.104)	(1.105)	(1.104)	(1.106)	(1.061)	(1.106)
House value in top 10%	-3.655***	-3.654***	-3.660***	-3.640***	-3.672***	-3.189**	-3.065**
	(1.348)	(1.350)	(1.350)	(1.350)	(1.349)	(1.322)	(1.304)
House value in bot 10%	0.984	0.983	0.965	0.949	0.960	1.143	1.183
	(1.198)	(1.199)	(1.201)	(1.204)	(1.204)	(1.154)	(1.152)
Constant	64.703***	64.712***	64.992***	66.042***	65.777***	202.416***	220.202***
	(7.329)	(7.296)	(7.441)	(7.536)	(7.507)	(28.730)	(21.203)
Observations	4,615	4,615	4,615	4,615	4,615	4,610	4,610
R-squared	0.266	0.266	0.266	0.266	0.267	0.325	0.328
School X Cohort FE	X	X	X	X	X	X	X
SchoolXTeachersX5years FE	X	X	X	X	X	X	X
Major dummy						X	X
University dummy							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 Earning majors.
Dependent variable: time to graduation defined as time between high school graduation and first college graduation in months.

Table 12: The effect of high-school class female share on college exit score of students enrolling in High Earning majors

	Dependent variable: college exit score rescaled between 0-1						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
VARIABLES							
(Female=1)	0.068***	0.084***	0.066***	0.069***	0.069***	0.024***	0.025***
	(0.006)	(0.016)	(0.008)	(0.009)	(0.009)	(0.008)	(0.008)
Female share	-0.018	-0.006					
	(0.021)	(0.022)					
(Fem share X) (Female=1)		-0.034					
		(0.032)					
(Fem share above median)=1			-0.003				
			(0.009)				
(Female=1)X(Fem share above median=1)			0.004				
			(0.012)				
(Fem_share2quartile=1)				0.005	0.005	0.003	0.003
				(0.009)	(0.009)	(0.008)	(0.008)
(Fem_share3quartile=1)				-0.007	-0.008	-0.000	-0.000
				(0.010)	(0.010)	(0.010)	(0.010)
(Fem_share4quartile=1)				0.020	0.019	0.022*	0.023*
				(0.015)	(0.016)	(0.013)	(0.013)
(Female=1)X(Fem_share2quartile=1)				-0.006	-0.006	-0.013	-0.014
				(0.014)	(0.014)	(0.013)	(0.013)
(Female=1)X(Fem_share3quartile=1)				0.025*	0.025*	0.010	0.011
				(0.014)	(0.013)	(0.012)	(0.012)
(Female=1)X(Fem_share4quartile=1)				-0.032**	-0.030*	-0.034**	-0.033**
				(0.016)	(0.016)	(0.013)	(0.013)
Female-only class=1					-0.029	0.007	0.004
					(0.051)	(0.039)	(0.039)
Male-only class=1					-0.014	-0.022**	-0.022**
					(0.012)	(0.009)	(0.010)
2nd Rank Quartile=1	0.047***	0.047***	0.047***	0.046***	0.046***	0.060***	0.059***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.008)	(0.008)
3rd Rank Quartile=1	0.091***	0.091***	0.091***	0.090***	0.090***	0.116***	0.112***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.007)	(0.008)
4th Rank Quartile=1	0.177***	0.177***	0.177***	0.176***	0.176***	0.213***	0.208***
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.007)	(0.007)
House value in top 10%	0.008	0.008	0.008	0.008	0.008	0.008	0.008
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.008)	(0.008)
House value in bot 10%	-0.005	-0.005	-0.005	-0.004	-0.004	-0.011	-0.011
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.008)	(0.008)
Constant	0.246***	0.243***	0.236***	0.193***	0.195***	0.103***	0.102**
	(0.021)	(0.022)	(0.020)	(0.026)	(0.027)	(0.032)	(0.045)
Observations	4,603	4,603	4,603	4,603	4,603	4,603	4,603
R-squared	0.298	0.298	0.298	0.300	0.300	0.464	0.466
School X Cohort FE	X	X	X	X	X	X	X
SchoolXTeachersX5years FE	X	X	X	X	X	X	X
Major dummy						X	X
University dummy							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 Earning majors.
Dependent variable: college exit score assigned on the basis of G.P.A. and a final dissertation, rescaled between 0-1.

Table 13: The effect of high-school class female share on actual wage of students enrolling in High Earning majors

	Dependent variable: log of actual wage						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
VARIABLES							
(Female=1)	-0.135***	-0.294***	-0.161***	-0.147**	-0.147**	-0.198***	-0.189***
	(0.034)	(0.101)	(0.044)	(0.060)	(0.060)	(0.062)	(0.062)
Female share	-0.148	-0.254**					
	(0.113)	(0.122)					
(Fem share X) (Female=1)		0.338*					
		(0.193)					
(Fem share above median)=1			0.024				
			(0.054)				
(Female=1)X(Fem share above median=1)			0.082				
			(0.074)				
(Fem_share2quartile=1)				-0.115**	-0.115**	-0.103*	-0.100*
				(0.048)	(0.049)	(0.052)	(0.052)
(Fem_share3quartile=1)				-0.040	-0.040	-0.033	-0.032
				(0.059)	(0.061)	(0.063)	(0.064)
(Fem_share4quartile=1)				-0.084	-0.086	-0.085	-0.080
				(0.088)	(0.089)	(0.098)	(0.100)
(Female=1)X(Fem_share2quartile=1)				-0.062	-0.061	-0.019	-0.026
				(0.101)	(0.102)	(0.100)	(0.100)
(Female=1)X(Fem_share3quartile=1)				-0.023	-0.023	0.027	0.031
				(0.109)	(0.110)	(0.115)	(0.116)
(Female=1)X(Fem_share4quartile=1)				0.209**	0.204**	0.274**	0.283**
				(0.099)	(0.102)	(0.114)	(0.114)
Female-only class=1					0.092	-0.011	-0.060
					(0.163)	(0.207)	(0.193)
Male-only class=1					-0.002	-0.089	-0.086
					(0.078)	(0.092)	(0.095)
2nd Rank Quartile=1	-0.026	-0.027	-0.027	-0.024	-0.024	-0.006	-0.018
	(0.062)	(0.061)	(0.061)	(0.061)	(0.062)	(0.066)	(0.066)
3rd Rank Quartile=1	0.102**	0.101**	0.103**	0.100**	0.100**	0.084	0.053
	(0.048)	(0.048)	(0.048)	(0.048)	(0.048)	(0.053)	(0.053)
4th Rank Quartile=1	0.197***	0.198***	0.197***	0.197***	0.198***	0.145**	0.098*
	(0.052)	(0.052)	(0.052)	(0.052)	(0.052)	(0.056)	(0.057)
House value in top 10%	0.086	0.086	0.084	0.086	0.085	0.104	0.097
	(0.083)	(0.083)	(0.083)	(0.083)	(0.084)	(0.076)	(0.076)
House value in bot 10%	0.107**	0.104**	0.106**	0.105**	0.106**	0.111**	0.111**
	(0.048)	(0.048)	(0.048)	(0.049)	(0.049)	(0.056)	(0.055)
Constant	10.005***	9.958***	9.777***	9.834***	9.841***	4.712***	4.687***
	(0.104)	(0.106)	(0.119)	(0.139)	(0.141)	(0.199)	(0.340)
Observations	4,016	4,016	4,016	4,016	4,016	3,437	3,437
R-squared	0.310	0.310	0.310	0.312	0.312	0.370	0.374
School X Cohort FE	X	X	X	X	X	X	X
SchoolXTeachersX5years FE	X	X	X	X	X	X	X
Major dummy						X	X
University dummy							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 graduated in High Earning majors.
Dependent variable: logarithm of personal income, as revealed to the internal revenue service in year 2005.

Table 14: The effect of high-school class female share on students' major choice enrollment

Dependent variable: High Earning major choice at first college enrollment=1					
	(1)	(2)	(3)	(4)	(5)
VARIABLES					
(Female=1)	-0.201***	-0.326***	-0.233***	-0.248***	-0.245***
	(0.009)	(0.024)	(0.011)	(0.014)	(0.014)
Female share	-0.042	-0.166***			
	(0.031)	(0.037)			
(Fem share X) (Female=1)		0.246***			
		(0.045)			
(Fem share above median)=1			-0.028**		
			(0.014)		
(Female=1)X(Fem share above median=1)			0.072***		
			(0.015)		
(Fem_share2quartile=1)				-0.020	-0.016
				(0.014)	(0.015)
(Fem_share3quartile=1)				-0.023	-0.018
				(0.016)	(0.016)
(Fem_share4quartile=1)				-0.069***	-0.064***
				(0.021)	(0.021)
(Female=1)X(Fem_share2quartile=1)				0.031	0.029
				(0.021)	(0.021)
(Female=1)X(Fem_share3quartile=1)				0.055***	0.053***
				(0.019)	(0.019)
(Female=1)X(Fem_share4quartile=1)				0.126***	0.124***
				(0.023)	(0.023)
Female-only class=1					-0.002
					(0.035)
Male-only class=1					0.111***
					(0.039)
House value in top 10%	0.009	0.011	0.010	0.011	0.011
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
House value in bot 10%	-0.006	-0.008	-0.007	-0.007	-0.008
	(0.011)	(0.012)	(0.011)	(0.012)	(0.012)
Constant	0.392***	0.442***	0.394***	0.398***	0.397***
	(0.009)	(0.013)	(0.006)	(0.008)	(0.008)
Observations	17,346	17,346	17,346	17,346	17,346
R-squared	0.140	0.142	0.141	0.142	0.142
School X Cohort FE	X	X	X	X	X
SchoolXTeachersX5years FE	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.
Sample: high school graduates enrolling in college.
Dependent variable: dummy=1 if student enrolls one of the high earning majors (Engineering, Economics & Business and Medicine) at first college enrollment.

Table 15: Robustness check: the effect of high-school class female share on choice of major controlling for family fixed effects

	Dependent variable: High Earning major choice at first college enrollment=1						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
VARIABLES							
(Female=1)	-0.220***	-0.349***	-0.372***	-0.250***	-0.278***	-0.274***	-0.274***
	(0.015)	(0.047)	(0.053)	(0.020)	(0.029)	(0.029)	(0.029)
Female share	-0.101	-0.226**	-0.198**				
	(0.076)	(0.088)	(0.093)				
(Fem share X) (Female=1)		0.256***	0.315***				
		(0.087)	(0.097)				
(Fem share)X(4th Rank Quart=1)			-0.090				
			(0.144)				
FemX(4th Rank Quart=1)			0.079				
			(0.103)				
(Fem share)XFemX(4th Rank Quart=1)			-0.205				
			(0.196)				
(Fem share above median)=1				-0.014			
				(0.029)			
(Female=1)X(Fem share above median=1)				0.072**			
				(0.030)			
(Fem_share2quartile=1)					-0.018	-0.013	-0.013
					(0.030)	(0.031)	(0.031)
(Fem_share3quartile=1)					0.011	0.019	0.019
					(0.036)	(0.036)	(0.036)
(Fem_share4quartile=1)					-0.104**	-0.095**	-0.095**
					(0.044)	(0.044)	(0.044)
(Female=1)X(Fem_share2quartile=1)					0.062	0.059	0.059
					(0.042)	(0.042)	(0.042)
(Female=1)X(Fem_share3quartile=1)					0.072*	0.068*	0.068*
					(0.041)	(0.041)	(0.041)
(Female=1)X(Fem_share4quartile=1)					0.127***	0.124***	0.123***
					(0.044)	(0.044)	(0.044)
Female-only class=1						-0.030	-0.028
						(0.121)	(0.122)
Male-only class=1						0.231	0.229
						(0.144)	(0.144)
Class size							-0.000
							(0.003)
Class geographical concentration							-0.068
							(0.224)
Class % in bottom 10 house value							0.105
							(0.128)
Class % in top 10 house value							0.005
							(0.128)
2nd Rank Quartile=1	0.086***	0.087***	0.088***	0.088***	0.086***	0.086***	0.087***
	(0.020)	(0.020)	(0.020)	(0.020)	(0.020)	(0.020)	(0.020)
3rd Rank Quartile=1	0.180***	0.179***	0.179***	0.181***	0.179***	0.179***	0.179***
	(0.020)	(0.020)	(0.020)	(0.020)	(0.020)	(0.020)	(0.020)
4th Rank Quartile=1	0.272***	0.272***	0.333***	0.272***	0.272***	0.272***	0.272***
	(0.021)	(0.021)	(0.071)	(0.021)	(0.021)	(0.021)	(0.021)
House value in top 10%	0.047*	0.050*	0.050*	0.050*	0.049*	0.049*	0.049*
	(0.026)	(0.026)	(0.026)	(0.026)	(0.026)	(0.026)	(0.027)
House value in bot 10%	0.008	0.007	0.009	0.006	0.008	0.008	0.003
	(0.028)	(0.028)	(0.028)	(0.028)	(0.028)	(0.028)	(0.029)
Constant	0.601***	0.644***	0.631***	0.575***	0.566***	0.563***	0.569***
	(0.153)	(0.151)	(0.152)	(0.152)	(0.151)	(0.152)	(0.166)
Observations	4,317	4,317	4,317	4,317	4,317	4,317	4,317
Number of famid	2,374	2,374	2,374	2,374	2,374	2,374	2,374
Family FE	X	X	X	X	X	X	X
School X Cohort FE	X	X	X	X	X	X	X
SchoolXTeachersX5years FE	X	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.
Sample: high school graduates enrolling in college.
Dependent variable: dummy=1 if student enrolls one of the high earning majors (Engineering, Economics & Business and Medicine) at first college enrollment.

Table 16: Robustness check: the effect of high-school cohort female share on students' major choice enrollment

Dependent variable: High Earning major choice at first college enrollment=1						
	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES						
(Female=1)	-0.219***	-0.434***	-0.422***	-0.271***	-0.263***	-0.260***
	(0.009)	(0.026)	(0.027)	(0.010)	(0.012)	(0.012)
Female share	0.136	-0.089				
	(0.086)	(0.094)				
(Fem share) X (Female=1)		0.427***				
		(0.048)				
(Fem share)X(4th Rank Quart=1)			-0.103			
			(0.068)			
FemX(4th Rank Quart=1)			0.016			
			(0.053)			
(Fem share)XFemX(4th Rank Quart=1)			-0.065			
			(0.097)			
(Fem share above median)=1				-0.002		
				(0.021)		
(Female=1)X(Fem share above median=1)				0.117***		
				(0.014)		
(Fem_share2quartile=1)					0.019	0.020
					(0.017)	(0.017)
(Fem_share3quartile=1)					0.016	0.016
					(0.023)	(0.023)
(Fem_share4quartile=1)					-0.000	-0.000
					(0.027)	(0.027)
(Female=1)X(Fem_share2quartile=1)					-0.018	-0.020
					(0.021)	(0.021)
(Female=1)X(Fem_share3quartile=1)					0.085***	0.083***
					(0.017)	(0.017)
(Female=1)X(Fem_share4quartile=1)					0.137***	0.134***
					(0.020)	(0.020)
Female-only class=1						0.031
						(0.033)
Male-only class=1						0.102**
						(0.044)
2nd Rank Quartile=1	0.085***	0.085***	0.085***	0.085***	0.085***	0.085***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
3rd Rank Quartile=1	0.176***	0.176***	0.174***	0.176***	0.176***	0.176***
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
4th Rank Quartile=1	0.277***	0.277***	0.337***	0.277***	0.277***	0.277***
	(0.012)	(0.012)	(0.038)	(0.012)	(0.012)	(0.012)
House value in top 10%	0.018*	0.021**	0.016	0.021**	0.021**	0.021**
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
House value in bot 10%	0.002	0.000	-0.001	0.001	0.001	0.001
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
Constant	0.146**	0.251***	0.276*	0.195***	0.186***	0.186***
	(0.058)	(0.062)	(0.159)	(0.023)	(0.025)	(0.025)
Observations	17,346	17,346	17,346	17,346	17,346	17,346
R-squared	0.154	0.158	0.180	0.157	0.158	0.158
School FE	X	X	X	X	X	X
SchoolXTeachersX5years 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.
Sample: high school graduates enrolling in college.
Dependent variable: dummy=1 if student enrolls one of the high earning majors (Engineering, Economics & Business and Medicine) at first college enrollment.

Table A.1 - List of high schools by track

Liceo classico (focus on classical studies, humanities)	**Liceo scientifico** (focus on math and sciences)
ISTITUTO LSLR PITAGORA	DA VINCI
BECCARIA	DONATELLI-PASCAL
BERCHET	EINSTEIN
MANZONI	GALILEI
OMERO E TITO LIVIO	MARCONI
PARINI	SEVERI
	VOLTA

Table A.2 - Fields of study aggregation

Fields of Study	**College Majors Aggregated into the Field**
Agriculture	Agriculture, Nutrition, Veterinary
Architecture and Design	Architecture, Design
Economics and Business	Economics, Business
Education	Communication Studies, Education, Nursing, Physical education
Engineering	Engineering
Humanities	Literature, Languages, Philosophy, Cultural Heritage, Art History, Music, History, Archeology
Law	Law studies
Mathematics and Statistics	Mathematics, Statistics, Physics, Computer Science.
Medicine	Medicine
Natural Sciences	Biology, Bio Technological Sciences, Pharmacy, Environmental Studies, Geology, Natural Science, Chemistry
Social Sciences	Political Sciences, Social Sciences, Sociology, Psychology