I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 16689

Female Classmates, Disruption, and STEM Outcomes in Disadvantaged Schools: Evidence from a Randomized Natural Experiment

Sofoklis Goulas
Rigissa Megalokonomou
Yi Zhang

DECEMBER 2023

I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 16689

Female Classmates, Disruption, and STEM Outcomes in Disadvantaged Schools: Evidence from a Randomized Natural Experiment

Sofoklis Goulas
Brookings Institution

Rigissa Megalokonomou
Monash University, University of Queensland, IZA and CESifo

Yi Zhang
University of Queensland

DECEMBER 2023

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Female Classmates, Disruption, and STEM Outcomes in Disadvantaged Schools: Evidence from a Randomized Natural Experiment*

Recent research has shown that females make classrooms more conducive to effective learning. We identify the effect of a higher share of female classmates on students' disruptive behavior, engagement, test scores, and major choices in disadvantaged and non-disadvantaged schools. We exploit the random assignment of students to classrooms in early high school in Greece. We combine rich administrative data with hand-collected student-level data from a representative sample of schools that feature two novel contributions. Unlike other gender peer effects studies, a) we use a rich sample of schools and students that contains a large and diverse set of school qualities, and household incomes, and b) we measure disruption and engagement using misconduct-related (unexcused) teacher-reported and parent-approved (excused) student class absences instead of self-reported measures. We find four main results. First, a higher share of female classmates improves students' current and subsequent test scores in STEM subjects and increases STEM college participation, especially for girls. Second, a higher share of female classmates is associated with reduced disruptive behavior for boys and improved engagement for girls, which indicates an increase in overall classroom learning productivity. Third, disadvantaged students—those who attend low-quality schools or reside in low-income neighborhoods—drive the baseline results; they experience the highest improvements in their classroom learning productivity and their STEM outcomes from a higher share of female classmates. Fourth, disadvantaged females randomly assigned to more female classmates in early high school choose college degrees linked to more lucrative or prestigious occupations 2 years later. Our results suggest that classroom interventions that reduce disruption and improve engagement are more effective in disadvantaged or underserved environments.

JEL Classification: J16, J24, I24, I26

Keywords: gender peer effects, natural experiment, classroom learning productivity, STEM careers, quasi-random variation, disadvantaged students

Corresponding author:
Rigissa Megalokonomou
Monash University
900 Dandenong Road, Caulfield East
Melbourne, Victoria, 3145
Australia
E-mail: rigissa.megalokonomou@monash.edu

* We thank seminar participants at the University of Queensland, Queensland University of Technology, University of Warwick, the 2023 Western Economics Association conference at the University of Melbourne, 2022 Australian Gender Economics Workshop, 2020 Australian Gender Economics Workshop at the Queensland University of Technology, and the 2021 International Panel Data Conference at the University of Macedonia. This project has been reviewed by the Office of Research Ethics at the University of Queensland and deemed to be exempt from ethics review; Clearance Number: 2020002332. We are grateful to Nancy Gyftopoulou for helping us compile the data. We also thank Dominic Byrne for very helpful work. This paper uses data collected from school archives and administrative data, and we are not at liberty to publish online; however, we can provide the programs as well as detailed instructions and assistance in how to apply and access the data in order to enhance the replicability of our analysis. There is no material or financial conflict of interest on the part of any of the authors with respect to any of the research in this paper. Rigissa Megalokonomou acknowledges research support from the University of Queensland BEL Early Career Grant (No: UQECR1833757).

1 Introduction

Social scientists, educators, and policymakers are interested in preparing students for careers in STEM (science, technology, engineering, and mathematics). STEM careers are attractive, because they are associated with higher lifetime earnings (Fayer, Lacey, and Watson, 2017). On a societal level, a larger STEM workforce may lead to further innovation and long-run economic growth. At the same time, initiatives that prepare females for STEM careers are valuable. Although OECD data show that women attain more education than men (OECD, 2016), they remain underrepresented in STEM fields: Only 28% of college students enrolling in STEM are female (OECD, 2017). These gender differences in STEM degrees translate to gender differences in STEM occupations. This under-representation of females in STEM degrees and occupations accounts for a significant part of the gender pay gap among college graduates (Blau and Kahn, 2017; Buser, Peter, and Wolter, 2017).

What shapes these gender differences in STEM occupations is the focus of much recent research. Much of this gender gap can be traced to differential performance in the related subjects in earlier grades or differential study choices made by boys and girls in school and at university (Buffington, Cerf, Jones, and Weinberg, 2016; Goulas, Griselda, and Megalokonomou, 2022). Some studies emphasize gender differences in preferences (Delaney and Devereux, 2019), while other studies highlight the role of social and environmental factors that influence the gap (Lavy and Megalokonomou, 2023a). Indeed, there is evidence that the gender gaps in performance in STEM subjects is not evident at the beginning of schooling but emerges over time (Fryer and Levitt, 2010; Hyde and Mertz, 2009). Thus, the school environment may contribute to the gender gap in STEM performance and STEM college participation.

Classmates constitute an important social force that shapes academic achievement and educational choices through intensive daily interactions and influences classroom learning productivity. Thus, classmates may substantially influence human capital formation and student decisions. We focus on an observable characteristics of classmates—gender—to study whether exposure to a higher proportion of female peers in the classroom affects students' STEM performance in high school and STEM college participation. For identification of causal effects, we rely on idiosyncratic variation in the proportion of female students within school cohorts across classrooms. To control for unobserved characteristics of students, we exploit an institutional setting with quasi-random peer group formation.[1] Students in Greece are alphabetically assigned to classrooms (based on surname) at the beginning of high school, which alleviates common concerns about selection bias (Manski, 1993). The basic idea is to compare the outcomes and choices of students from different classrooms within the same school-cohort who are exposed to the same classroom environment and have similar characteristics, except for the fact that one classroom has a higher share of female peers than the other for idiosyncratic reasons.[2] We study the impact of classroom gender composition on end of year student test scores in high stakes exams, subsequent test scores

[1] This identification methodology is different from that of Hoxby (2000); Lavy and Schlosser (2011); Hill (2017); and Brenøe and Zölitz (2020), who exploit variation across cohorts within schools, and that of Goulas, Megalokonomou, and Zhang (2022), who exploit variation across cohorts within neighborhoods.

[2] We also account for student and classroom characteristics, and thereby investigate how a higher share of female peers affects outcomes for males and females with similar characteristics and in similar classrooms. This identification approach is similar to that of Anelli and Peri (2019), and Zölitz and Feld (2021), who use classroom-to-classroom variation in gender composition within school-cohorts and teaching section-to-section variation in female share within course-cohorts.

at the end of high school, and tertiary education choices and enrollments. High school performance is the only determinant for university admission in Greece. We focus on test scores in STEM subjects and enrollments in STEM college degrees, but we also study gender composition effects on test scores in humanities subjects and enrollments in humanities college degrees.

We focus on an important dimension in this investigation, which is the socioeconomic background of students. This is a dimension which has not been explicitly studied in other gender peer effect studies. Students from lower socioeconomic backgrounds often perform poorly in STEM courses, which are a necessary step in pursuing fast-growing and lucrative STEM careers (Rozek, Ramirez, Fine, and Beilock, 2019). We hypothesize that exposure to a higher share of female classmates has a more pronounced effect on students who come from a lower socioeconomic background. The motivation is that students' outcomes may be more responsive to peer interactions when there is a lack of resources in their environment and a potential shortage of financial support from the family for children's education and public educational resources. Therefore, gaining new insights about whether educational inputs or interventions (such as exposure to a higher share of female classmates) are more effective in improving educational outcomes for disadvantaged students could make a potential contribution to the design, targeting, and implementation of new interventions, and a better resource allocation.

We use novel student-level data from more than 40,000 students in Greece from a large number of schools for the period 2002-2009. We combine hand-collected test scores, transcripts, and attendance records from 104 public high schools in Greece, which cover more than 40,000 students, with administrative data obtained by the Ministry of Education on university admissions for all students in Greece for several years. Our dataset offers two main advantages. First, we use a rich sample of schools and students that contain a large and diverse set of school qualities, and household income. This allows us to investigate heterogeneous gender composition effects by school quality, and socioeconomic profile. Second, it contains rich information on students' performance in different subjects, classrooms, and grades, and different categories of class absences. In particular, the data distinguish between parent-approved student class absence and truancy or disciplinary teacher-reported expulsion from class, and thus measure student engagement and disruptive behavior.

The primary identification assumption is that classroom-to-classroom differences in the gender composition of students are exogenous to the drivers of STEM outcomes, conditional on school-by-cohort fixed effects. To assess the validity of this identification assumption, we conduct a battery of balancing tests and simulation exercises.[3] First, we show that the classroom-to-classroom variation in gender composition within a school-cohort follows a normal distribution. Second, a series of balancing tests show that the within-school-cohort variation in gender composition is orthogonal to student characteristics. Third, using Monte Carlo simulations, we derive evidence that the within-school-cohort variation in gender composition is consistent with that generated by a random process.

We assess the widely claimed statement that human capital investment and knowledge accumulation for females in STEM—but not all—subjects is enhanced in an environment with a higher share of females (Lavy and Schlosser, 2011). To study this, we split the available sub-

[3]Our approach is similar to that of Lavy and Schlosser (2011); Goulas, Megalokonomou, and Zhang (2022); Anelli and Peri (2019); Gong, Lu, and Song (2021); and Mouganie and Wang (2020).

jects into STEM and humanities subjects, and examine the impact of a higher share of females on student performance and participation in STEM and humanities, separately. If anything, we expect to find less pronounced effects on test scores and participation in humanities. This is because humanities-related learning may use memorization more intensely than problem solving, and classroom-based learning may be more crucial for problem solving than memorization. For instance, memorization during class time may be more easily substitutable with memorization at home compared with problem solving. Also, there is evidence that teacher instruction and experience is more effective in mathematics rather than reading (Hanushek and Rivkin, 2010; Papay and Kraft, 2015). Thus, humanities-related learning may rely less on classroom-based learning. Indeed, we find little (or negative) influence of female classmates on humanities performance and participation.[4]

We find that a higher proportion of female classmates positively affects STEM performance and STEM college degree enrollments for all students, and effects are larger for females. We show that a 10 percentage point increase in the share of female classmates is associated with an increase in grade 11 test scores by 3% and 1% of a SD for females and males, respectively, and an increase in the likelihood of females enrolling in STEM fields in college by 0.4 percentage points (relative to a mean of 12%). These effects are of comparable magnitude to those in the literature on improving school inputs. For instance, our effects on STEM outcomes are comparable to being taught by a teacher between 1.5 and 2 SDs above the average (Hanushek, Kain, O'Brien, and Rivkin, 2005; Lavy and Megalokonomou, 2023a) or to reducing the class size by 20% (Angrist and Lavy, 1999; Krueger, 1999). Classroom gender composition has a smaller and statistically insignificant effect on males STEM degree choice, which suggests a decrease in the gender gap in STEM degrees. We also find a nonlinear structure in gender composition effects, with their magnitude increasing substantially when the share of female peers is over 65%. Exploiting variation in the quality and socioeconomic characteristics of our sampled schools, we investigate heterogeneous gender com- position effects along those dimensions. We find that the overall effects are driven by students in disadvantaged settings, i.e., lower-quality schools, and schools in lower-income neighborhoods. These results have longer-run implications, since we find evidence (although we have limited power) that females exposed to more females in disadvantaged high schools are more likely to choose college degrees linked to more lucrative or prestigious occupations a few years later. In line with our initial hypothesis, this suggests that gender composition effects are more salient in disadvantaged or underserved contexts.

Gender composition may affect classroom learning productivity in two ways. First, if females are less prone to misconduct than males, a higher female share may improve learning productivity by lowering the share of disruptive peers. Second, having more female classmates may lower individual disruptive behavior by causing all students to be more tranquil and compliant (Cohen and Strayer, 1996). We directly study whether a higher share of female peers affects the quality of the classroom learning environment by changing individual disruptive behavior, measured by classroom misconduct-related (unexcused) absences. We also study the impact of gender composition on parent-approved class absences, which measure student engagement. We find that a higher proportion of female classmates is associated with less disruptive behavior for boys and

[4]A negative effect on humanities performance or participation could be due to substituting studying effort away from humanities and possibly toward STEM.

higher engagement for females, which suggests an increase in overall classroom learning productivity. Using a mediation analysis, we quantify the role of disruptive behavior and engagement in accounting for classroom gender composition effects (Chung and Zou, 2020; Gong, Lu, and Song, 2018, 2021). In particular, we find that the impact of female peers that may be transmitted through improved classroom learning productivity (due to lower misconduct and disruption) accounts for up to 8% of the total gender composition effect on STEM performance. Students in lower-quality schools and lower-income neighborhood schools experience the highest improvements in their classroom learning productivity due to having more female peers in the classroom. An improved classroom environment can explain up to 20% of the gender peer effects on STEM-related subject scores in disadvantaged contexts.

Our study moves beyond prior literature in three important ways. First, we are able to investigate gender composition effects in smaller peer groups than previously considered. Much of the related literature exploits across-cohort variation in demographics, including gender (Bifulco, Fletcher, and Ross, 2011; Brenøe and Zölitz, 2020; Carrell and Hoekstra, 2010; Goulas, Megalokonomou, and Zhang, 2022; Hoxby, 2000; Lavy, Paserman, and Schlosser, 2012; Lavy and Schlosser, 2011; Mouganie and Wang, 2020). The caveat in previous studies is that cohort peers may only serve as a rough approximation of actual student interactions (Xu, Zhang, and Zhou, 2020). Students may be more likely to interact intensively in a small peer group, such as that of their classroom peers, rather than in the broader group of cohort peers (Chetty, Friedman, Hilger, Saez, Schanzenbach, and Yagan, 2011; Duflo, Dupas, and Kremer, 2008; Feld and Zölitz, 2017; Gong, Lu, and Song, 2021; Hu, 2015; Megalokonomou and Zhang, 2022; Sacerdote, 2001; Whitmore, 2005; Zimmerman, 2003; Zölitz and Feld, 2021).

Second, the broad coverage of our data allows us to study gender composition effects in disadvantaged settings, since the role of female peers may be more pronounced among students who have limited resources. To the best of our knowledge, this is the first paper that explicitly studies gender peer composition effects on disadvantaged students and discusses the implications. This group is of particular interest, since the efficiency of interventions may be larger in settings in which there is lack of financial and family educational resources. For instance, Anelli and Peri (2019) study the impact of being in a male-dominated environment on students' study choice in college using data from Milan; their analysis pertains to a group in Italy's upper tail of the income and educational distribution. Zölitz and Feld (2021) study the impact of gender composition across tutorial sections in one business school in the Netherlands. Again, their study sample is rather a group at the upper tail of the income and educational distribution, since they focus on university students in a selective institution. In contrast, our results are drawn from a broader distribution of neighborhood income, and school quality, and cover a large and diverse set of schools in the country. Gong, Lu, and Song (2021) examine gender composition effects using a more representative sample of middle schools in China, however there is no emphasis on the socioeconomic profile of the schools or students, or explicit focus on disadvantaged groups. The authors find that a higher proportion of female classmates improves females' contemporaneous test scores and noncognitive outcomes using the China Education Panel Survey (CEPS).[5]

[5] Although the CEPS collects rich information on students, a few restrictions apply due to the stratified sampling design (Gong, Lu, and Song, 2018, 2021; Megalokonomou and Zhang, 2022). CEPS does not sample all classes within a school, but instead it collects data from 2-3 classes within each school-cohort, and thus not all classes within each school-cohort are included in Gong, Lu, and Song (2021). This may limit the variation in the proportion of females within school-cohorts, especially given

Third, we study the effects of high school gender composition on the choice of field of study at the university level, i.e., STEM and non-STEM university degrees. This is among the first papers that studies how high school peers influence postsecondary education choices and in particular, the field of study chosen at the university level, which has important implications for students' later occupational choices. Brenøe and Zölitz (2020) exploit across-cohort variation in the share of females in high school math track and find that a higher share of female peers reduces the likelihood for females to enroll in STEM university programs using administrative data from Denmark. We exploit random variation in the share of female classmates in students' general education classes and thus, examine whether the interest of untracked students—before they make any specialization decision—shifts towards STEM. The general education classes are compulsory for all students and better represent the "general population" compared with the more homogeneous track classes. Students usually self-select into tracks and this may limit the within-track variation in the subject preparedness and personality traits of males and females. For instance, females who select into the math track may be more willing to compete and discourage other females, while males who select into the math track may not be as disruptive as males in the general population. Anelli and Peri (2019) show that high school gender composition has an effect only on males (but not on females') choice of study at the university level using data from Italy. In particular, they find that males attending high school classes with over 80 percent of male peers are more likely to enroll in predominantly male college majors.[6] We use a continuous measure of female classmates instead and a more balanced sample in terms of gender and socioeconomic characteristics.[7]

Forth, we contribute to the literature on the mechanisms of gender peer effects on STEM performance. Previous literature has focused on students' perceived classroom environment, inter-student and teacher-student relations, and teachers' and students' behaviors (Gong, Lu, and Song, 2021; Lavy and Schlosser, 2011; Schøne, von Simson, and Strøm, 2016). Such outcomes are self-reported and inevitably subject to measurement error that may potentially correlate with the variable of interest and could bias the estimates. We overcome this limitation by using student attendance information from administrative sources. Excused and unexcused absences are verified and recorded by teachers. While prior studies have provided suggestive evidence of a mechanism of gender peer effects (Anelli and Peri, 2019; Brenøe and Zölitz, 2020; Lavy and Schlosser, 2011), we quantify the relative importance of changes in individual disruptive behavior and engagement in explaining gender composition effects. Given the overall differences in disruptive behavior and engagement for students across socioeconomic backgrounds, we repeat this exercise for disadvantaged and non-disadvantaged schools. Our findings on the mechanism of gender composition effects also contribute to our understanding of the direct relationship between peer

the large class size in middle schools in China. We contribute to the literature by exploiting primary-collected data combined with administrative data on all classes of the sampled schools. This probably gives us larger variation in the share of female classmates. Second, CEPS collects rich information on student attainment, but all outcomes are measured in the same grade. We contribute by examining the impacts of current classroom gender composition, not only on contemporaneous outcomes, but also on longer-term outcomes and expand our understanding on how classroom gender composition may affect occupational sorting and the gender pay gap.

[6] Anelli and Peri (2019) define as predominantly male college majors programs in Engineering, Economics, and Business. Brenøe and Zölitz (2020) consider all STEM programs which also include programs in Science, Technology, and Mathematics. We define as STEM programs all programs in Science, Technology, Engineering, and Mathematics. In robustness exercises, we also include Health Sciences and Economics in the definition of STEM.

[7] In our sample, the share of male students in the classroom is at least 80% in only 10% of the classrooms.

effects and disruptive behaviors (Carrell and Hoekstra, 2010; Figlio, 2007; Lavy, Paserman, and Schlosser, 2012; Lazear, 2001; Xu, Zhang, and Zhou, 2020).

More broadly, our paper contributes to a better understanding of the origins of the gender gap in STEM performance and STEM choices. While we do not offer a universal explanation for the gender composition effect on STEM outcomes for males and females, we identify one relevant channel and quantify its contribution to the total effect. This study highlights the fact that gender peer effects operate partially by affecting disruptive behavior and engagement—the effects of which, in turn, impact school and professional careers.

2 Institutional Framework

The educational system in Greece is highly centralized (OECD, 2018a). Students are assigned to public schools through zoning based on the proximity of their residential address to schools.[8] At the beginning of high school (grade 10), students are quasi-randomly (alphabetically, based on surname) assigned to physical classrooms in which they take all core courses (Goulas, Griselda, and Megalokonomou, 2022, 2023; Lavy and Megalokonomou, 2023a,b).[9] Students come from different middle schools, and thus usually most of their grade 10 classmates are brand new. Students are not allowed to switch classrooms and must remain in their assigned class for all grades in high school. The alphabetical classroom assignment, together with some small fluctuations in school enrollment from one year to the next, cause quasi-random fluctuations in classroom gender composition which we exploit. Students remain in the same class for their courses and extracurricular activities throughout all years of high school. In Section 5, we provide evidence that student characteristics are randomly assigned to classrooms and that on average classroom characteristics are very similar within school-cohorts. Teachers in each school are also randomly assigned to classrooms to facilitate teachers' schedules, while taking into account their subject specialization.[10] Teachers rotate between classrooms to teach courses in their specialization.

All grade 10 exams are school exams. School exams are designed by classroom teachers and are requested to be of the same format and average difficulty, cover identical content, and test the same skills. Usually teachers coordinate and design their exams together. The school principal is responsible for ensuring that the teachers follow the Ministry of Educations grading guidelines for each subject when grading the school exams. The school principal receives the marked exam papers for the school exams from each teacher within five days after the corresponding exam. Then, the regulation requires the principal to read the marked exam papers, approve the marks, write them in the school log, and enter them into the school computer (if available). Through the physical process of reading the exam papers and documenting the marks, the principal ensures they are following the grading guidelines as required by the regulation (Lavy and

[8] 92% of schools are public, and 93% of students in Greece attend public schools (Goulas and Megalokonomou, 2020). Families are unable to enroll their children in a different public school than the one assigned, since they are required to submit proof of their residential address and utility bills.

[9] The regulation regarding alphabetical classroom assignment is strictly enforced across all schools. See Government Gazette of the Hellenic Republic 167 A/1566/1985. See also Education Ministry Bulletin of the Hellenic Republic 100749/Γ2/17-09-07.

[10] According to the law, if there are disagreements within the school board about teacher assignment to classrooms within a year, the school authority and the Ministry of Education intervene. Extensive evidence of this quasi-random assignment of teachers to classrooms in the same educational system can be found in Lavy and Megalokonomou (2023a), Lavy and Megalokonomou (2023b), and Dinerstein, Megalokonomou, and Yannelis (2022).

Megalokonomou, 2023a). We will also later use information on students' GPA in grade 10. This is derived as the average test score of several school exams that students take throughout the year.

Students take several compulsory general education subjects in each high school grade. They submit their track preferences at the end of grade 10, and are assigned to a track at the beginning of grade 11. This is students' first instance of track specialization. The majority of students remain in the same track in grade 12. Each student takes compulsory specialization-track courses in grades 11 and 12. There are three specialization tracks—(1) Classics, (2) Science, and (3) Exact Science—and students choose based on their desired field of study at the university level. There is no minimum performance threshold for students to enroll in any track, and all schools offer exactly three tracks. We identify two general education subjects as STEM-related in grades 11 and 12 (Mathematics, and Physics), and two subjects as humanities-related in grades 11 and 12 (Language and History).[11] These subjects are compulsory for all students in both grades 11 and 12. Students take standardized national exams on those subjects (as well as other subjects) at the end of the school year. Performance in those exams is the only determinant for university admission. University admission is administered by the Ministry of Education.[12] These are high-stakes exams that are blindly graded by external assessors. We examine the classroom gender composition effects on the average standardized performance in STEM- and humanities-related subjects in grades 11 and 12 as outcomes.

After students take the national exams in grade 12, they submit a list of their preferred tertiary degree programs to the Ministry of Education (OECD, 2018a).[13] A centralized system compares average exam scores of students and assigns candidates to degree programs based on their preferences and degree availability.

We consider two college-related outcomes: admission to a degree program in STEM and admission to a degree program in the humanities. We consider all degree programs offered by Mathematics, Science, Engineering, or Computer Science departments to be STEM degree programs.[14] We consider degree programs in Languages, Literature, Philosophy, History, Religion, or Art to be humanities degree programs. We then assign occupation wages to each college degree based on the exact or closest correspondence between available occupations and college degrees.

3 Data

Our novel dataset combines information from various sources. We conducted primary data collection by visiting and retrieving administrative data on 43,451 students across 2,517 classrooms in 104 high schools.[15] Our data collection was planned and executed in close collaboration with the Ministry of Education and the local School Authorities. Prior to visiting the schools, we sought

[11] In grade 11, Mathematics consists of Algebra and Geometry.

[12] For more information on the national exams and the university admissions process, see Goulas, Megalokonomou, and Zhang (2022); Goulas and Megalokonomou (2019); and Goulas and Megalokonomou (2020).

[13] By *degree program* we mean a department at a specific university. Each university department offers exactly one bachelor's degree program.

[14] In a robustness exercise, we broaden the definition of STEM degrees to include programs in Health Sciences, such as Medicine and Biology, and Economics and Business (Goulas, Griselda, and Megalokonomou, 2022).

[15] We exclude classrooms with five or fewer students (260 observations dropped) because these small classrooms are likely to be atypical in a number of dimensions. We also exclude school cohorts in which students are of a single gender (268 observations), because then there is no variation in our main variable of interest. We also drop school cohorts with only one classroom (487 observations), since our main identification comparison is between classrooms within school-cohorts.

approval from the Ministry of Education and obtained ethics approvals from the Hellenic Data Protection Authority and the Institute of Educational Policy. We also obtained consent from the local school authorities and school principals. Our sample corresponds to roughly 10% of public schools in Greece. Table B.1 compares the sampled schools with the population. Schools in our sample are nationally representative with regard to several student characteristics—such as female share, average age, and share of students born in the 1st quarter of the birth year[16]—average school performance in the university admission exams, and youth unemployment in the vicinity of the school.[17] The locations of the sampled schools are shown in Figure A.1. Schools in our sample are distributed throughout the country and cover a diverse set of areas. Our sample includes schools from the most and least deprived areas in Greece and schools that score at the bottom and top of the distribution in the university admission exams. The map indicates that there are schools in big cities as well as in smaller urban areas and islands.

Our data span eight cohorts that graduated between 2001-2 and 2008-09 and include transcript and attendance information for all high school grades (from 10th through 12th grade) and all classes of the sampled schools. Each record contains a student identifier, a school and classroom identifier, demographic information on the student (year of birth, gender, 1st quarter of the birth year indicator), track enrollment, graduation status, and test scores for each student in each subject and grade. Our student-grade-level attendance data separate misconduct and truancy-related (unexcused) teacher-reported student class absences and parent-approved (excused) class absences. A student's total absences equal the sum of excused and unexcused absences. Excused absences proxy engagement and are usually authorized by parents or guardians, often with a note signed by a doctor or the parent for some short-term illness. They involve entire school days. Unexcused absences indicate a student's suspension from class and are initiated by subject teachers (Goulas, Griselda, and Megalokonomou, 2023; Lavy and Megalokonomou, 2023a,b). We also obtained university admissions information for each student from the Ministry of Education. For each student, we have records on the postsecondary degree to which they were admitted. The information we obtained on admission to tertiary education allows us to examine the impact of female peers on longer-term outcomes.

We rank schools based on the average school-level 12th-grade national university admissions exam scores across all years to construct a measure of school quality.[18,19] We also obtained household income data (in 2009 in euros) at the postcode level from the Ministry of Finance for the whole country and use this as a proxy for neighborhood in each school's postcode. We use this information later to conduct heterogeneity exercises by postcode income and school quality (above and below the median values).

We also obtain occupation-related earnings data from the Labor Force Survey and the National Statistical Authority for 2003, and we map college occupations into annual earnings. This is a

[16]Students born in the 1st quarter of the birth year begin school before the age of six.

[17]Although we have a slightly smaller share of students enrolling into the classics track in grade 12 in our sample compared with the school population. Youth unemployment is measured in 2003.

[18]The quality of school s is calculated as $School\ Quality_s = \frac{Ordinal\ Rank_s - 1}{N-1}$, where $Ordinal\ Rank_s$ is the ordinal rank of school s's mean national exam performance across cohorts. School quality ranges from zero to one, with one being the highest. N is the number of schools in the sample.

[19]The school ranking remains very similar if instead of the across-all-years school-level 12th-grade national university admissions exam scores, we use the year 2002 (first year in the data) to rank schools, and then we perform the analysis for all years except 2002.

proxy for expected annual salary earnings from each occupation or a proxy for how lucrative each occupation is.

Table 1 presents descriptive statistics of our sample at student level (Panel A), classroom level (Panel B), and school level (Panel C).[20] Roughly 56% of students in the average classroom are female, with a SD of 0.148.[21] On average, each school has four classrooms in grade 11, with an average classroom size of 16 students.[22] Figure 1 shows the distribution of the proportion of female classmates in the data (histogram bars) and a simulated normal distribution with the actual mean and SD (in dashes). The actual distribution of the share of female classmates closely resembles a normal distribution.

4 Empirical Strategy

We estimate the impact of female classmates on academic performance and college degree choice using the following linear specification separately for male and female students:

$$Y^{g}_{i,c,s,t} = \beta^{g} PropFemalePeers_{-i,c} + \boldsymbol{X}_i' \gamma^{g} + \boldsymbol{W}'_{-i,c}\delta^{g} + \lambda^{g}_{s,t} + \epsilon^{g}_{i,c,s,t} \tag{1}$$

where $g \in \{Males, Females\}$. $Y^{g}_{i,c,s,t}$ denotes average subsequent performance and enrollment in STEM and humanities for student i of gender g in cohort t in classroom c and school s. The treatment variable of interest is $PropFemalePeers_{-i,c}$, which represents the proportion of female peers (excluding student i) in student i's classroom in grade 11. Vector $\boldsymbol{X}_i'$ includes student-level controls—such as prior GPA, age, born in the 1st quarter of the birth year indicator, track enrollment intention[23]— and share of female classmates in the previous grade.[24] $\boldsymbol{W}'_{-i,c}$ includes leave-out means of all student characteristics at the classroom level and classroom size.

We include school-cohort fixed effects $\lambda_{s,t}$ to control for potential school sorting and school-cohort-specific unobservables. The basic idea is to compare the outcomes and study choices of students who are exposed to the same school environment and have similar characteristics, except for the fact that one classroom has a higher share of female students than the other in the same school-cohort for idiosyncratic reasons. Standard errors are clustered at the classroom level, which is the level of randomization, to allow for heteroskedasticity and serial correlation in the outcomes of students in the same classroom.

[20] Table B.2 provides descriptive statistics by cohort.

[21] In the past, men were overrepresented in different levels of education, but this trend has been changing in recent decades. Since the 1990s, in many countries around the world, females have started to become the dominant gender in terms of education participation at various levels of education. In 2005, women represented 55% of the higher education student population in the OECD area (OECD, 2018b). If these trends continue, in countries such as Austria, Canada, Iceland, Norway, and the United Kingdom, there will be almost twice as many female students as males in higher education in 2025 (OECD, 2018b). In 2005, 62% of higher education degrees in Greece were awarded to females, while the OECD average was 57%. Other countries with similar female shares are Canada, Finland, Ireland, Italy, Spain and New Zealand (OECD, 2018b). In Greece, we find that the proportion of females who graduate from high schools and take the national exams–a prerequisite for university admissions–is on average 57%. This is just 2 percentage points above the average OECD higher education student population.

[22] Table B.3 shows the distribution of classroom numbers in grade 11 by cohort.

[23] Students submit their desired 11th grade track choice enrollment at the end of grade 10, and thus prior to the realization of the variable of interest.

[24] Classrooms that have a higher share of females also have a lower share of males. Because of the mechanical relationship between the share of females and share of males within school-cohorts, it is not feasible to disentangle the effect of having more females from the effect of having less males.

Coefficient β^g captures the impact of classroom gender composition on student standardized test scores and study choices of students of gender g, holding all other variables constant. We transform raw test scores in each subject into z-scores to facilitate interpretation. Then, we average the standardized test scores across STEM and humanities-related subjects in each grade. The main identifying assumption to obtain causal estimates for β^g in specification (1) is that there are no omitted variables that are correlated with both classroom gender composition and the outcomes of interest, and are not included in our extensive set of student- and peer-level sets of controls. To assess the existence of such factors and thus the validity of our empirical strategy, we conduct an extensive set of checks in Section 5.

5 Evidence of the Validity of the Identification Strategy

5.1 Source of Variation

Understanding how the observed classroom gender composition in 11$^{\text{th}}$ grade is formed is vital to our estimation and analysis. Students are quasi-randomly assigned to classrooms following a lexicographic order based on surnames in 10$^{\text{th}}$ grade, the initial year of high school. The ideal experiment would require the treatment (proportion of female classmates) to be set at the grade of randomization, i.e., grade 10; however, student or classroom level information on grade 9 is unavailable, as students attended different middle schools, and thus we would not be able to include any prior controls for test scores in that case. Thus, we set the treatment in grade 11 and rely on the randomness of the initial classroom composition in 10$^{\text{th}}$ grade and small additional idiosyncratic variation in the classroom composition due to very few reassignments across grades which are unlikely to be systematic or distort classroom composition significantly.[25,26] We discuss those movements below.

The reason for the few reassignments between grades 10 and 11 is the following: Very few students transfer into (on average, 1 student per school-cohort) or out (on average, fewer than 1 student per school-cohort) of the sampled schools between grades 10 and 11. This minimal degree of mobility triggers a slightly larger (but still small) mobility, since schools need to follow specific class size rules. For instance, assume a student transfers into a school one week after the official school starting date and their surname starts with a letter early in the alphabet. This student will be assigned to classroom 1 due to the alphabetic assignment of students to classrooms based on surname. This lexicographic assignment based on students' surnames has to be satisfied for all students at any point during the school year in a school-cohort.[27] However, there are also

[25]We show that there is sufficient across-classroom variation within school-cohorts to obtain consistent estimates of the classroom gender composition effect. Limited variation in group characteristics may cause amplification bias (Angrist, 2014). We perform a variance decomposition for the share of female classmates. Table B.4 shows that the within-school-cohort variation in the share of female classmates accounts for 72% of the total variation. This variation comes from the ample within-school-cohort variation in grade 10 (71% of total variation) and the difference between grade 10 and grade 11 (87% of total variation).

[26]We obtain very similar results if we set the treatment at grade 10 and re-run our analysis (i.e., exploit random variation in the share of female classmates in grade 10) instead of grade 11. In that case, we have no prior controls for test scores, but it may not be very concerning given the across-classroom randomization. However, we prefer to set the treatment to grade 11 in our baseline results to include a full set of controls.

[27]During our data collection, we had the opportunity to discuss the random assignment of students with the school principals. We have kept notes from our discussion with the principal during every school visit. Our notes indicate that every school we visited followed the law requiring alphabetical assignment to classrooms. During the data extraction process, privacy restrictions did not allow us to request first and last names from the schools. However, schools shared student names with us in ten instances

classroom size rules that need to be satisfied. After the incoming student is assigned to classroom 1, the size of classroom 1 has to be readjusted by dropping one student. Then, the student whose surname is alphabetically ranked as the last in classroom 1 must move to classroom 2 to satisfy the class size rules. This will potentially lead to a further readjustment, since the student whose surname is alphabetically ranked as the last in classroom 2 must move to classroom 3. This adjustment may be carried over to all classrooms until the highest classroom number has a new student enrolled. We checked whether this institutional feature is accommodated in the data. Figure A.2 shows that in 80% of the cases, there is no student changing classroom between grades 10 and 11, and thus the proportion of females in grade 10 is identical to the proportion of females in grade 11.[28] Also, in the majority of cases in which a student changed classrooms between grades 10 and 11 in the sample, the student was moved to the adjacent classroom, which reassures us that the institutional features are satisfied. A limitation of our data is that we are unable to follow students across schools to check at which point these changes occur or follow students across schools.

In Subsections 5.2, and 5.3 we provide extensive evidence of the randomness of the variable of interest. In Subsection 5.4 we show that any movements between grades 10 and 11 are small and orthogonal to student or classroom characteristics and that the classroom assignment of students in grade 10 is also random.

5.2 Randomness of Class Gender Composition

The identification strategy exploits the classroom-by-classroom variation of female classmates share within school-grade. To verify that the variation in the share of female classmates is as good as random and uncorrelated with individual characteristics, we perform the following tests.

First, we show that the residuals of the share of female classmates after controlling for school-cohort fixed effects are normally distributed. In particular, we regress the share of female classmates on school-cohort fixed effects and obtain the predicted residuals (Anelli and Peri, 2019). Figure 2, Panel (A) shows the kernel density of the residuals of the classroom-level female peers proportion obtained from the data and a simulated normal distribution with the same standard deviation (0.148) and number of unique classrooms (2,517) as the actual data. Clearly, the empirical residual distribution closely matches with the simulated normal distribution. Figure 2, Panel (B) shows the standardized normal probability plot. The 45-degree line (black) represents the benchmark where the empirical female peers share residuals follow a normal distribution, while the scatter points (in gray) show the extent to which the residuals of the actual female classmates share depart from the normal distribution. We observe that there is no significant deviation of the residuals from the hypothetical normality, including the tails of the distribution. Therefore, this result supports that the variation in the female classmates share in 11$^{\text{th}}$ grade—driven by

under a separate data-sharing agreement. We have gone back to the data from these ten schools to investigate whether students are assigned alphabetically to classrooms in grade 11. Our investigation of these ten schools and eight cohorts (between 2001-02 and 2008-09) revealed that 99.85% of students were assigned to classrooms alphabetically. We were able to look closely into the limited cases of misassigned students. Those cases relate to differences in the order of the letters between the Greek and Latin alphabets. We do not have reason to believe that this kind of nuance is of any consequence for our empirical investigation of the impact of female classmates on students' later outcomes.

[28] Our results remain very similar if we only keep students in school-cohorts in which there was no across-classroom change. In that case, we do not include controls for the share of female classmates in grade 10, since it would be identical to that in grade 11.

the randomized classroom assignment in 10^{th} grade—is as good as random, after controlling for school-cohort fixed effects.

Second, our key identifying assumption would be violated if, for instance, female students could select into classrooms based on the expectation of a higher or lower proportion of female classmates. If the classroom assignment is truly random conditional on school-cohort factors, we should find a weak association between students' observed characteristics and the share of female peers in the classroom. Table 2 presents the results of the balancing test in which we regress the proportion of female students in grade 11 (our variable of interest) on each student's characteristics, or prior test scores—such as gender, age, born in 1^{st} quarter of birth year indicator, or 10^{th}-grade GPA—conditional on school-cohort fixed effects. The practically zero and statistically insignificant estimate on all characteristics and prior test scores, as well as the small F-statistic on the joint significance of all characteristics and prior test scores, suggests that students' characteristics and student prior test scores are not associated with the gender classroom composition.[29] Table B.6 shows estimated effects for a similar exercise conducted for males (Panel A) and females (Panel B), separately. Estimated effects are again practically zero and insignificant. This is very reassuring given that the main specification is estimated separately by gender.

5.3 Monte Carlo Simulation

We also conduct Monte Carlo simulations to examine whether the within-school-cohort deviation in the proportion of female classmates in grade 11 is idiosyncratic. We do so by examining whether the observed within-school-cohort variation in the proportion of female peers resembles the variation that would stem from a simulation of a randomly generated gender composition. For each school-cohort, we randomly generate the gender of each student using a binomial distribution with p equal to the actual proportion of females in the school-cohort.[30] We then compute the within-school-cohort standard deviation of the proportion of female peers across classrooms. We repeat this simulation process 1,000 times to obtain a 95% empirical confidence interval of the within-school-cohort standard deviation for each school-cohort. Table 3 summarizes the results from 1,000 simulations. We find that 87%, 91%, and 96% of the school-cohorts with the observed standard deviations in the proportion of female peers fall within the corresponding 90%, 95%, and 99% empirical confidence intervals, respectively. This provides evidence that the actual classroom assignment is consistent with a random process.[31] We also visually show one

[29] We also check whether the predetermined characteristics (gender, age, and born in 1^{st} quarter of birth year indicator) of students at classroom level and the female drop out share are balanced in all classrooms within a school-cohort. We regress each classroom-level mean characteristic on classroom number indicators 1, 2, 3, 5, 6, and 7 (with classroom number 4 being the omitted group as a point of comparison), conditional on school-cohort fixed effects. Table B.5 shows that only 3 out of 24 coefficients are statistically significant. The p-values of F-statistics for the joint significance of the regressors suggest balanced classroom-level characteristics within school-cohorts. We conduct similar exercises in which we exclude a different classroom number other than 4, and the pattern remains the same. Results are available upon request.

[30] Mouganie and Wang (2020) and Lavy and Schlosser (2011) conduct similar exercises in which they simulate cohort-by-cohort variation of the proportion of female peers within each school. We simulate classroom-by-classroom variation of the proportion of female peers within school-cohort.

[31] Our results are in line with those of Mouganie and Wang (2020), in which 93% of the observed standard deviations are located within 95% empirical confidence intervals; and Lavy and Schlosser (2011), in which 89% of the observed standard deviations fall within 90% empirical confidence intervals. In Figure A.3, we visualize the simulation results of 95% empirical confidence intervals of the within school-cohort standard deviations in the proportion of female peers across classrooms. Due to space constraints, we randomly select 50 out of 728 school-cohorts. Figure A.3 shows that around 90% of the actual standard deviations fall within their corresponding confidence intervals. Visualized results for other school-cohorts and 90% and 99% simulated

simulation in Figure 3. Clearly, the actual within school-cohort variation in the share of female peers is meaningfully different from the simulated variation.

5.4 Other Checks for Randomness

In this subsection, we show that a) the few cross-classroom transfers between grade 10 and 11 are uncorrelated with students' characteristics or prior test scores and b) the grade 10 classroom assignment of students is indeed random.[32]

We start by examining whether the difference in the proportion of females from grade 10 to 11 and the likelihood of transfer is associated with student characteristics. In Table B.7, column 1, we regress the change in the proportion of female classmates from grade 10 to 11 on students' characteristics and students' prior GPA. All estimates are practically zero and statistically insignificant. This finding suggests that the change in the share of female classmates from 10^{th} to 11^{th} grade is independent from student characteristics and prior performance. In columns 2 and 3, we use as outcome variables a binary indicator that takes the value of 1 if a student was transferred to another class in grade 11 and 0 otherwise, and the 10^{th} grade proportion of female classmates, respectively. We do not find any association between students' characteristics and the reported outcomes. This implies that the class transfer status and the initial gender class composition in 10^{th} grade is not associated with students' characteristics or prior test scores. Small F-statistics values also provide evidence for the joint insignificance of the regressors.

Lastly, we perform some additional backward verification exercises of the randomness of gender composition in 11^{th}-grade classrooms. To do so, we examine the randomness of the 10^{th}-grade proportion of females and the change in the proportion of females between grades 10^{th} and 11^{th}. Figure A.4 presents the histogram of the share of females in grade 10 (Panel A) and evidence of residuals' normality of the 10^{th}-grade share of female classmates net of the school-cohort fixed effects (Panels B and C). These exercises are similar in spirit to Figures 1 and 2, but are produced now for grade 10 instead of grade 11. The simulation exercise, which is summarized in Figure A.5 and Table B.8, also provides evidence that the variation in the 10^{th}-grade proportion of females and the difference between the 10^{th}- and 11^{th}-grade proportions of females follow an idiosyncratic pattern.[33] Taken together, the variation in the proportion of female classmates in 11^{th} grade is indeed random.

confidence intervals are available upon request.

[32]Only a few students in the sample drop out from 10^{th} to 11^{th} grade (around 1.5% of the full sample). We do not include these observations of dropouts in our estimation sample since we do not have information on their outcomes. To examine whether students may be more or less likely to drop out as a response to the share of female classmates, we run the following specification. We report the estimate and standard error from a regression that has the likelihood of observing a dropout as the dependent variable and the proportion of female classmates in 10^{th} grade as the independent variable, conditional on school-cohort fixed effects. The estimated coefficient is equal to -0.011 and the standard error is equal to 0.009. Column 4 of Table B.5 also suggests that the female share of dropouts is well balanced across classrooms. Taken together, there is little evidence that drop outs would significantly contaminate the randomness of the 11^{th}-grade proportion of females originated from the 10^{th}-grade random assignment.

[33]These exercises are similar in spirit to Figure 3 and Table 3.

6 Results

We first present the baseline results of gender classroom composition by gender for the full sample. We assess the robustness of these results and test for non-linearities. Then, we conduct exercises in which we examine whether the baseline results change when we consider students in disadvantaged schools and non-disadvantaged schools, which we define based on postcode household income and school quality.

6.1 Main Results

Table 4 shows the main estimated effects of classroom gender composition in 11$^{\text{th}}$ grade on average test scores in STEM- and humanities-related subjects in 11$^{\text{th}}$ grade (Panel A), 12$^{\text{th}}$ grade (Panel B), and university degree major choices (Panel C). Columns 1-2 and 4-5 report the results for males and females, respectively. All results account for student-level controls and school-cohort fixed effects. Results in columns 2 and 5 additionally include classroom-level controls. We also present the *Means of Y* for each gender below standard errors.

We find that being assigned to a classroom that has a higher proportion of female classmates increases performance at the end of 11$^{\text{th}}$ (Panel A) and 12$^{\text{th}}$ grades (Panel B) in STEM subjects for both males and females. Same grade (grade 11) effects are larger than subsequent (grade 12) effects, especially for females. Our results show that a higher share of female classmates increase female students' performance more than that of males. In particular, a 10-percentage-point increase in the share of female peers in the classroom increases females' test scores in STEM subjects by 0.026 of a SD in grade 11, while the effect on males' test scores is half as large—i.e., 0.013 of a SD.[34]

We assess the widely accepted claim that human capital investment and knowledge accumulation for females in STEM subjects is enhanced in environments with a higher share of females (Lavy and Schlosser, 2011). Thus, we look at the effects of the treatment variable on STEM and humanities subjects, separately. Performance in humanities subjects may rely less on classroom-based instruction than performance in STEM subjects. Previous studies find that interventions that target classroom learning are more effective for improving performance in math than in reading (Abdulkadiroğlu, Angrist, Dynarski, Kane, and Pathak, 2011; Angrist, Pathak, and Walters, 2013; Behrman, Fan, Wei, Zhang, and Zhang, 2020; Black, Doolittle, Zhu, Unterman, and Grossman, 2008; Dobbie and Fryer Jr, 2013). Additionally, classroom-based instruction has been found to be more effective in improving calculation abilities and boosting performance in math rather than language skills and performance in reading (Hanushek and Rivkin, 2010; Lavy and Megalokonomou, 2023a; Papay and Kraft, 2015).

We investigate the impact of classroom gender composition on average performance in humanities subjects in both Panels A and B of Table 4. We mostly find negative, smaller, and insignificant effects of the share of female peers in the classroom on end of the same grade (grade 11) or subsequent (grade 12) performance in humanities subjects. In fact, the estimated effects on females' performance in grade 12 are negative and significant (coefficient=-0.074, s.e.= 0.036).

[34] Results using a broader definition of STEM degrees that includes Health Sciences and Economics/Business are shown in Table B.10 and are very similar to when we use the baseline definition of STEM degrees.

A negative effect on humanities performance could be due to substituting studying effort away from humanities and toward STEM.[35,36]

We then find substantial longer-term effects of classroom gender composition on university degree choice (Panel C in Table 4). Females assigned to classrooms with more females in grade 11 are more likely to choose a STEM college degree 2 years later. In particular, a 10-percentage-point increase in the proportion of female peers increases females' likelihood of enrolling in STEM degrees by 0.43 of a percentage point. The estimated effect on humanities degrees is negative and statistically significant, indicating that females switch away from humanities degrees and move toward STEM degrees. The estimated effect of the share of female classmates on males' degree choices is indistinguishable from zero, which suggests a decrease in the gender gap in STEM study.[37]

Our results are in line with those of Schneeweis and Zweimüller (2012) and Schøne, von Simson, and Strøm (2016), who find that having more female peers in lower secondary school increases females' likelihood of studying STEM in upper secondary school. Our findings also point to the same direction as that of Gong, Lu, and Song (2021), who find that a 10-percentage-point increase in the proportion of female classmates in middle school increases a student's end of year test score by around 10% of a SD. Contrary to our findings, Anelli and Peri (2019) use data from high schools in Milan and find that a higher share of females in high school has a small and insignificant effect on females' probability of choosing male-dominated majors. They also find positive effects on males' test scores from being assigned to a very male-dominated environment, i.e., at least 80 percent of male peers. Opposite to our findings, Zölitz and Feld (2021) leverage cross-cohort variation in a Business School in the Netherlands and find that a 10 percentage point increase in the share of female tutorial peers at the university reduces females' probability of choosing a male-dominated major by 0.8 percentage points and increases their probability of choosing a female-dominated major by 1 percentage point.

6.2 Placebo Exercise

To test for the validity of the research design, we estimate the model based on a sample with placebo treatment. To do so, we reshuffle students within the same school-cohort so that stu-

[35] We also examine the effects of classroom gender composition on test scores in each STEM and humanities subject. Table B.11 shows consistent classroom composition effects across STEM and humanities individual subjects.

[36] The data structure so far is such that we consider one row per student and we calculate the average performance across STEM and non-STEM subjects. A different estimation approach would be to explicitly account for differences in each STEM or humanities subject and stack the data for the various subjects after converting the raw test scores into z-scores by subject, school, and year. Table B.13 presents the estimated results while using this alternative specification. We now have multiple rows per student—one row for each STEM subject (Mathematics, Physics) in each grade and one row for each non-STEM subject (Language, History) in each grade. This stacked panel data permits regression analysis with controls for subject-by-school-by-cohort fixed effects. A similar pattern is observed as in Table 4, in which a higher share of female classmates improves student test scores in STEM subjects in both grades 11 and 12, especially those of females.

[37] In Table B.9 we present the estimated effects of the share of female classmates in grade 10 on test scores in grade 10. We present these results for the curious reader, since this is the only exam in the whole high school career of students that is determined before students progress to grade 11. Although, we made a decision to set the treatment in grade 11 in order to include a full set of student controls, we still provide evidence that the immediate gender composition effect on grade 10 test scores is mostly in line with the remaining effects. Females who are initially assigned to a higher share of female classmates in grade 10 end up performing significantly better in STEM subjects in exams at the end of grade 10. The effects for females' performance in humanities subjects females are small and imprecise. The same applies for male students. The only student-level controls here are students' age and an indicator for whether a student is born in the $1^{s}t$ semester. We control for this in the main specification since we include student GPA in grade 10 as a control in the main analysis.

dents are placed in different classrooms, in which they were never actually placed in reality. We then compute the updated gender composition as a *placebo* female classmates share. Similar to Goulas, Megalokonomou, and Zhang (2022), we conduct an exercise in which we replace the share of female classmates in the main specification with the placebo share of female classmates. The placebo estimates are reported in Table 4 in columns 3 and 6 and are practically zero and statistically insignificant. This suggests that our baseline classroom gender composition effects do not reflect spurious correlation between the proportion of female classmates and confounding factors at the school-cohort level.

6.3 Robustness Checks

6.3.1 Reshuffling Exercise

Our identification of classroom gender composition effects relies on the orthogonality between the share of female classmates and student characteristics within school-cohorts. To examine whether the baseline results are driven by some schools that may not follow the random assignment, we create random restricted samples by iteratively dropping schools from two random draws of the pool of sampled postcodes and re-estimate the effects of classroom gender composition (Gong, Lu, and Song, 2018, 2021).[38] We estimate a total of 3,403 regressions for each outcome variable and plot the distributions of the estimates in Figure A.6. We find that all distributions are concentrated around the baseline estimates, which suggests that the baseline classroom gender composition effects are not driven by specific schools. These findings suggest that our baseline results are unlikely to be driven by schools that do not follow the random classroom assignment.

6.3.2 Spillover Effects from Female Peers in Other Classrooms within School-cohorts

We further examine the robustness of our results to spillover effects from females in other classrooms. The proportion of female peers in one's classroom may be mechanically and negatively correlated with the proportion of females in other classrooms in the same school-cohort. In Table B.14 we present the baseline estimates when we include controls for the share of female peers in all other classrooms within school-cohorts in the specification. These estimates (columns 3-4) are very similar to those obtained in the baseline results (columns 1-2), suggesting that our results are not driven by female share spillovers from other classrooms.

6.3.3 Spillover Effects from Female Peers' Ability

Another concern is that females outperform males and thus, the effects of a higher share of female classmates may reflect females' higher academic performance (Whitmore, 2005). To account for gender differences in ability, we include controls for each student's prior test scores and the average classroom level prior achievement in the baseline specification. To further mitigate this concern, in Table B.14 we also include controls for the average prior test scores (GPA 10) of each student's female classmates. Columns 5-6 in Table B.14 show that our estimates of classroom gender composition are very similar to the baseline estimates in columns 1 and 2.

[38] We sampled schools from a total of 83 postcodes and dropped two postcodes each time and re-estimated the baseline specification.

6.4 Nonlinear Effects

We use two approaches to investigate the nonlinear effects of the proportion of female peers on student outcomes. In the first, we replace the share of female classmates in the main specification with indicators for four quantiles of the share of female classmates. We construct those four quantiles based on the overall distribution of the share of female classmates in the whole sample.[39] Table B.15 presents the estimated nonlinear classroom gender composition effects. The effect associated with the first quantile of the proportion of female classmates is the benchmark and thus omitted from the specification. We find substantial classroom gender composition effects on STEM performance when the proportion of female classmates is 65% or higher (Quantile 4). Our results corroborate the findings of Hoxby (2000) and Goulas, Megalokonomou, and Zhang (2022), who document larger female peer effects when the proportion of female peers exceeds 66% and 58%, respectively.

In the second approach, we follow Anelli and Peri (2019) and iteratively estimate classroom gender composition effects when the proportion of female classmates exceeds a cutoff that ranges between 50% and 90% (in increments of 5 percentage points). This procedure allows us to identify the proportion of female classmates that maximizes classroom gender composition effects. Figures 4 and 5 show the results for males and females, respectively. We find that the classroom gender composition effect on females' STEM performance peaks when the proportion of female classmates is in the range of 80%-85% (Panels A and C in Figure 5). The likelihood of enrolling into a STEM study for females at the university level increases monotonically when the proportion of female classmates increases above 70% (Panel E in Figure 5). At the same time, females' test scores in humanities and their likelihood of enrolling into degrees in humanities at the university level decrease monotonically when the proportion of female classmates increases above 75% (Panels B, D, and F in Figure 5). For males, classroom gender composition effects on test scores and enrollments in STEM or humanities do not become indistinguishable from zero at almost any point of the proportion of female classmates distribution.

6.5 Heterogeneous Effects

6.5.1 By Household Income and School Quality

In this section, we investigate heterogeneous classroom gender composition effects by neighborhood income and school quality. Presence of resources for students may be an important factor for the effectiveness of educational interventions. In disadvantaged environments in which there is a lack of educational, financial, and other resources, targeted interventions may be more effective,

[39] Table B.12 shows the allocation of the share of female classmates across quantiles by cohort. Each classroom has a unique share of female classmates that falls in a unique quantile. However, schools usually have several classrooms within each cohort, and each classroom may be assigned to a different quantile. As we can notice in column (4), around half of the schools in each cohort have a share of female classmates that fall in three quantiles. Overall, more than 90% of the sampled schools have a share of females that falls in at least two quantiles in each cohort (column 7). We note that we obtained access to primary collected data from 104 schools, but not all of the schools appear in all years. The reason behind this is that some schools opened later than 2002 and some schools stopped operating before 2009. In 3 cases, the schools' computer labs together with all computers were destroyed by a fire or another disaster and the school data were all lost. For these schools, we digitized student records from books for early years and merged with the electronic version of the data for the later years. An example of a student record is shown in Figure A.7. For 6 additional schools, student records were kept in a book format for a few years. Our baseline results remain unaffected if we only use schools that remain in the sample for the whole sample period.

since students may not already be overloaded with information and resources.[40] School quality is also associated with availability of school resources, teacher quality, and overall learning environment (Landaud, Ly, and Maurin, 2020).[41] We stratify the sampled schools based on whether schools are located in neighborhoods above or below the median neighborhood household income, as well as whether schools are above or below the median school quality. We measure school quality based on the school's ordinal rank, which is computed based on the school-level average performance in standardized exams across cohorts 2002-2009.

Table 5 presents the estimated effects for those different subgroups by gender. In columns 1-2, 3-4, 5-6, and 7-8 we present the estimates of the gender classroom composition for students in schools below the median income, above the median income, below the median quality, and above the median quality, respectively. Overall, the estimates for the female classroom share are significantly positive and larger among students residing in neighborhoods with lower average household income, and in schools with lower quality. This indicates that students attending low SES schools are more affected by school peers. For instance, the estimated gender composition effect on the choice to enroll in a STEM university degree for females in low income neighborhoods is 0.045 (s.e.=0.023), while for females in high income neighborhoods it is 0.003 (s.e.=0.034). The same variable's coefficient for females in low quality schools is 0.051 (s.e.=0.024), while for females in high quality schools the effect is -0.003 (s.e.=0.030). The estimated effects for females who reside in rich neighborhoods and attend high-quality schools are much smaller in magnitude, and are statistically insignificant in most of the cases. We also find that for males any positive effects are mainly driven by males in low household income schools and below the median school quality. The implications of these results seem clear: Students' educational outcomes are more responsive to peer interactions and collaborative learning when there is a potential shortage of financial support from the family for children's education and public educational resources. This is an important finding and contributes to the literature about the effectiveness of education policies in settings with different levels of resources and socioeconomic backgrounds. This result is also in line with the literature on class size, which finds that the benefits of reducing the class size are much larger for black students and free lunch students (Krueger, 1999) or for low socioeconomic background students (Kara, Tonin, and Vlassopoulos, 2021).

6.5.2 By Prior Performance

We also examine heterogeneous effects by students' prior performance. We segment the sample to quartiles of student's prior performance in grade 10 and then estimate gender peer effects by using the baseline specification (1). Figure A.8 shows the estimated effects and the corresponding 90% confidence intervals for males and females with prior performance at each quartile of the distribution. Panel A in Figure A.8 shows that a higher share of female classmates improves all students' STEM performance, with the effects being larger for female students. As we can see, the treatment line for females is shifted higher up compared with males, indicating that effects are more pronounced for females across the whole distribution of prior performances compared

[40] For instance, in contemporaneous work we find large benefits for female students who are randomly assigned to a female (compared with male) peer role model, and these effects are driven by schools in the bottom of the income distribution (Goulas, Gunawardena, and Megalokonomou, 2023).

[41] School quality is positively correlated with neighborhood income (ρ=0.328; p-value<0.01).

with males. Effects are pointing to the same direction in Panels B and C, but with a less obvious heterogeneity by prior performance.

7 Identifying the Mechanisms

In this section, we explore the potential mechanisms through which gender peer effects operate. In particular, we focus on how students' absenteeism behaviors may change when they are exposed to more female peers in the classrooms and how students' outcomes vary in response to changes in absenteeism behaviors induced by the random variation of female peers share.

7.1 Association Between Female Classmates and Absenteeism Behavior

Studies on peer effects have shed light on how female peers could affect students' outcomes. For example, Lavy and Schlosser (2011) and Gong, Lu, and Song (2021) suggest that a higher proportion of female peers in the classroom improves students' perceived classroom environment, inter-student relationships, teacher-student relations, and students' and teachers' self-assessed behaviors, which drives the positive effect of female peers on students' achievement. However, those studies rely on self-reported data on attendance and absences, and are subject to measurement error that might be correlated with the proportion of females. For example, a classroom with more female students may report more leniently on average with regard to the classroom environment, own behaviors, and teachers' behaviors, which might bias the estimates.

We use attendance records obtained from schools' archives that measure disruption and engagement using misconduct-related (unexcused) teacher-reported and parent-approved (excused) student class absences instead of self-reported measures. Some benefits of using more objective measures of student attendance to measure disruption and engagement are that—unlike survey or questionnaire data—they (1) include every student, (2) offer measurement variation beyond that of Likert scales, (3) do not rely on self-reporting, and (4) reflect the disruptive behavior or engagement of the specific student they correspond to, and not just an assessment of those behaviors by parents or guardians.

We examine two novel mediating factors—excused and unexcused absences—through which the proportion of female peers could affect students' academic performance outcomes. The sum of excused and unexcused absences is the total hours of absence a student receives during the school year. Excused absence is an idiosyncratic absence pattern of students, which is recorded by teachers when there is parental consent for absence. Unexcused absence refer to unauthorized absence or classroom disruption, which is recorded when students are expelled from class by teachers because of misconduct. Aggressive and antisocial behaviors are associated with lower empathy, and boys tend to score lower on measures of empathy than girls (Cohen and Strayer, 1996; Keenan and Shaw, 1997). Our hypothesis is that students randomly assigned to a higher share of female classmates may reduce their disruptive behavior, which may be reflected in fewer unexcused absences, and create a more conducive learning environment that could motivate more students to attend classes, which would trigger a drop in excused absences.

To test this hypothesis and examine the proposed mechanisms, we investigate whether a higher female classmates share affects students' excused absences and unexcused absences in

11th and 12th grades, respectively.[42] Table 6 shows the estimated effects for specification (1), in which we replace the outcome variable with the different types of student absences (total, excused, and unexcused). Indeed, we find that male students who are assigned to classrooms with a higher share of females tend to reduce their unexcused absenteeism behavior persistently in both 11th and 12th grades.

The estimated effect of classroom gender composition on males' unexcused absences is -2.138 (s.e.= 1.219) in grade 11 and -2.887 (s.e.=1.027) in grade 12. This is a considerable effect given that the mean unexcused absences per boy for the whole school year is 30 and 35 hours in grade 11 and 12, respectively. The implication is that a higher share of female classmates is associated with lower classroom disruption in the same and the subsequent year. This conducive learning environment tends to motivate more females to attend classes, which is reflected in the decrease in females' average excused and total absences. Indeed, females reduce their total absences by 4.558 hours (s.e.= 2.157) and 3.295 hours (s.e.=1.965) in grades 11 and 12, respectively (column 4 of Panels A and B in Table 6). Overall, a 10 percentage point increase in the female classmates share increases test scores in STEM subjects by 2.6% of a SD for females and improves their engagement, class attendance, and instruction time by half of a school hour. The same increase in the share of female classmates increases males' test scores in STEM by 1.3% of a SD, and reduces males' expulsion and disruption by 0.21 school hours, which improves the classroom environment and increases their instruction time.[43]

In the previous section, we found evidence that estimated gender composition effects are more pronounced for students who reside in relatively poor neighborhoods (with neighborhood income below the median of our study sample) and attend low-quality schools (with percentile achievement ranks below the median of our study sample). We examine whether the channel of the more conducive learning environment in the presence of more females holds in disadvantaged settings. There is evidence that there is greater difficulty to attract and retain quality teachers in low-income schools possibly due to students' higher propensity to disrupt the class (McKee, Sims, and Rivkin, 2015). Thus, intuitively, the presence of females in disadvantaged schools may improve the learning environment substantially.

Tables B.17 and B.18 report estimated gender composition effects by neighborhood income and school quality. These tables show that males from poor neighborhoods (Table B.17) or low-quality schools (Table B.18) tend to decrease their unexcused absences significantly (Column 9), while females in those classrooms tend to reduce their excused absences (Column 6), when there is an increase in the share of female classmates. The estimated effect of a higher share

[42] A survey of 771 high school students that we conducted in 2022 suggests that unexcused absences are indeed perceived by the students as a penalty for student disruptive behavior in the classroom. These 771 students come from high schools in our sample. A relevant item in the student questionnaire is the following "Have you witnessed hourly unexcused absences given as a penalty to disruptive students?". Figure A.9, Panel A, shows that 89.37% of high school students in the sample have seen disruptive students receiving unexcused absences as a penalty for their behavior in the classroom. Another item in the questionnaire ("In which way can a student in your classroom behave and receive unexcused absences as a penalty?") provides the reason why students receive unexcused absences. Figure A.9, Panel B, shows that most students have seen students making noise and disrupting others' attention as the most frequent reasons to receive an unexcused absence. This survey data provides additional reassurance that unexcused absences are used as a penalty for disruptive behavior in the classroom.

[43] Not all schools in the sample gave us access to their student attendance data. To address potential sample attrition bias, we regress an attrition indicator (which takes the value of 1 if data on absences are missing and 0 otherwise) on the proportion of female peers and school-cohort fixed effects. Table B.16 shows that all estimated effects are close to 0 and are also statistically insignificant, which suggests that the estimated effects in the mechanism section are not driven by sample attrition.

of females on males' unexcused absences in low income neighborhoods is -4.286 (s.e.= 1.644) in grade 11 and -4.203 (s.e.=1.366) in grade 12. The effects are very small, almost zero, for males in high income neighborhoods (column 11). These estimated gender composition effects are almost double in magnitude compared with the ones in Table 6. The related estimated effects in Table B.18 are also larger than the full sample effects (-2.826 with s.e.= 1.699 in grade 11 and -4.150 with s.e.=1.405 in grade 12). This probably also explains why disadvantaged students—those who attend low-quality schools or reside in low-income neighborhoods—experience the highest improvements in their STEM outcomes from a higher share of female classmates, as we see in Table 5. A higher share of female classmates in disadvantaged settings significantly reduces classroom disruption, increases instructional time for students, and improves classroom learning productivity, providing additional evidence that education interventions may be more effective in settings with lower resources.

Taken together, a higher share of female peers reduces students' misconduct and motivates class attendance, which creates a conducive classroom learning environment that facilitates students' interaction and benefits students' academic achievement, especially in STEM. This improved learning environment due to students' behavioral changes could foster a collaborative learning atmosphere between genders and generate a positive impact on students' academic performance. This mechanism is even more pronounced and effective in low-quality schools and low-income neighborhoods.

7.2 Association Between Outcomes and Absenteeism Behavior

We study whether absenteeism behavior is a mediating factor through which a higher share of female peers improves students' outcomes by including controls for the absenteeism variables in the main specification (1) and examining whether the estimated effects change. The overall peer effects can be divided into a direct effect $\frac{\partial Y}{\partial Fem.\ Peer}$ and an indirect effect $\frac{\partial Y}{\partial Absenteeism} \times \frac{\partial Absenteeism}{\partial Fem.\ Peer}$. For absenteeism to be a plausible channel that explains some of the positive effects of female peer effects, the second term on the RHS of equation (2), $\frac{\partial Y}{\partial Absenteeism} \times \frac{\partial Absenteeism}{\partial Fem.\ Peer}$, must be positive.

$$\underbrace{\frac{dY}{dFem.\ Peer}}_{Overall\ Effect} = \underbrace{\frac{\partial Y}{\partial Fem.\ Peer}}_{Direct\ Effect} + \underbrace{\frac{\partial Y}{\partial Absenteeism} \times \frac{\partial Absenteeism}{\partial Fem.\ Peer}}_{Indirect\ Effect} \tag{2}$$

In Appendix Table B.19, we examine the sensitivity of the baseline effects to the inclusion of the absenteeism variables (i.e., controls for excused and unexcused absences) in the main specification. We focus on STEM test scores in grades 11 and 12 (Panels A and B), and STEM university major choice (Panel C). Columns 1 and 4 reproduce the baseline estimates. Columns 2 and 5 report estimates from the same specification, but from a reduced sample, since that not all schools gave us access to the student absenteeism data. In columns 3 and 6, we use the reduced sample, and we also include controls for the absenteeism channel in the specification. We find that all estimates in columns 3 and 6 decrease in magnitude except for the STEM major choice outcome (Panel C). The drop in the size of the estimates (between columns 2 and 3 or between columns 5 and 6 of Table B.19) is moderate indicating that it is not an exhaustive mechanism. It is not surprising that the absenteeism channels do not fully explain the positive female peers effects.

College educational choices are also likely to be associated with students' preparedness in a field of study or comparative advantage in a group of subjects rather than class attendance behaviors (Brenøe and Zölitz, 2020; Goulas, Griselda, and Megalokonomou, 2022). The higher likelihood of females choosing a STEM degree is probably an effect of their improved STEM-related test scores when assigned to a higher share of female classmates, which suggests more preparedness in STEM studies. These results suggest that the classroom environment or classroom learning productivity (proxied by the absenteeism behavior) is a plausible mechanism for the effects of female peers on test scores in STEM.

7.3 Decomposition of Mechanisms

To assess the importance of the attendance channels, we use decomposition analysis to quantify how much each of the proposed mechanisms—excused absence or unexcused absence—and their combined explanatory power explains female peer effects.[44] We follow the approach by Gong, Lu, and Song (2021) to assess the significance of excused and unexcused absenteeism as novel channels in explaining female classmates effects. Let $m^j_{i,c,s,t}$ denote the mechanism variables j ($j = 1, 2$, since we consider two mechanisms, excused and unexcused absences in 11$^{\text{th}}$ and 12$^{\text{th}}$ grade). We remove the gender notation g for simplicity.[45]

Table 7 presents the shares of the overall female classmate effects that can be attributed to unexcused (column 1) and excused (column 2) absences in an attempt to explain the effects on STEM-related test scores. Column 3 shows the total share that can be attributed to unexcused and excused absences. Column 4 shows the share that is unexplained by the proposed mechanisms. Panel A shows the shares of the female classmate effects that are explained by the proposed mechanisms for all students. We find that unexcused absences have significantly more explanatory power than excused absences in explaining gender peer effects on STEM-related test

[44] Recent studies such as Chung and Zou (2020); Gong, Lu, and Song (2021); and Gong, Lu, and Song (2018) have also used a decomposition analysis following the method of Heckman, Pinto, and Savelyev (2013) and Gelbach (2016).

[45] First, we replace the outcome variable and use the absenteeism behavior instead of students' scholastic outcomes in the main specification (1). This is summarized in specification (3):

$$m^j_{i,c,s,t} = \lambda^j_{s,t} + \alpha^j PropFemalePeers_{-i,c} + \boldsymbol{X}'_i\phi^j + \boldsymbol{W}'_{-i,c}\theta^j + \epsilon^j_{i,c,s,t} \quad j = 1,2 \tag{3}$$

Next, we include controls for the mechanism variables $m^j_{i,c,s,t}$ (excused and unexcused absences) in (1), so that the baseline equation is now augmented as in (4):

$$Y_{i,c,s,t} = \underline{\lambda_{s,t}} + \underline{\beta} PropFemalePeers_{-i,c} + \sum_{j=1}^{2} \zeta^j m^j_{i,c,s,t} + \boldsymbol{X}'_i\underline{\gamma} + \boldsymbol{W}'_{-i,c}\underline{\delta} + \epsilon_{i,c,s,t} \tag{4}$$

We include the 11$^{\text{th}}$ or 12$^{\text{th}}$ grade absenteeism variables as controls when the outcome is measured in the 11$^{\text{th}}$ or 12$^{\text{th}}$ grade, respectively. Heckman, Pinto, and Savelyev (2013) discuss the rationale of the equation. We illustrate some details in Appendix C. Following the derivation by Gelbach (2016), we show that:

$$\hat{\beta} = \hat{\underline{\beta}} + \sum_{j=1}^{2} \hat{\zeta^j} \hat{\alpha^j} \tag{5}$$

where $\hat{\beta}$ is the baseline estimated coefficient of $PropFemalePeers_{-i,c}$ in (1). Equation (5) suggests that the j^{th} component of the mechanisms in explaining the peer effects is $\hat{\zeta^j}\hat{\alpha^j}$ and the part that is unexplained by the proposed mechanism is captured by coefficient $\hat{\underline{\beta}}$ in (4). The explanatory power for each mechanism j is $\hat{\zeta^j}\hat{\alpha^j}/\hat{\beta}$, and the combined explanatory power is $\sum_{j=1}^{2} \hat{\zeta^j}\hat{\alpha^j}/\hat{\beta}$. Gelbach (2016) discusses the derivation in details. We apply the derivation in our context in Appendix D.

scores: For males, unexcused absences explain around 6.8% of the gender peer effects on STEM-related subject scores in grade 11 and 5.3% in grade 12; for females, unexcused absences explain approximately 2.2% in grade 11 and 1.4% in grade 12. In contrast, excused absences have little explanatory power: mostly around 1.0% for both genders. The remainder is unexplained by the proposed mechanisms. This result is in line with Gong, Lu, and Song (2021), who find that classroom environment alone mediates away around 8.9% of the increase in student test score.

Panels B and C show the decomposition of the gender peer effects to the proposed mechanisms for students in low and high income neighborhoods, respectively. Unexcused absences explain a much larger share of gender peer effects in the low compared with the high household income group especially for males (8.1% compared with 3.6% in grade 11 and 6.4% compared with 0.09% in grade 12). The pattern is the same in Panels D and E. Unexcused absences explain 15% of the overall effect for males in low-quality schools compared with almost 1% in high-quality schools in grade 11. This is also evident in grade 12; unexcused absences explain 7% of the effect for males in low-quality schools and 1% in high-quality schools in grade 12. The channel of a more conducive learning environment due to lower disruption and noise in the classroom explains a large share of the gender peer effects in STEM subjects in disadvantaged schools, but not in advantaged settings. That may be explained by the fact that students may be more disciplined and less disruptive in advantaged schools, while there is much space for improvement in disadvantaged schools. This is important from a policy perspective. It suggests that classroom interventions that reduce disruption are more effective in disadvantaged environments and have the potential to boost outcomes.

8 Gender Composition Effects on Gender Pay Gap

In order to fully assess the implications of a higher share of female classmates in early high school, it is important to consider the longer-term consequences on occupational choices. To do so, we link our university degree enrollments administrative data to occupations-related earnings data using the Labor Force Survey. We then replace the outcome variable in the main specification (1) with the occupation-related earnings and report the estimates of the effect of a higher share of female classmates.

Table 8 presents the estimated effects for the full sample (columns 1-2), and by subgroups based on socioeconomic profile (columns 3-6 by postcode income and columns 7-10 by school quality). The estimated effect of a higher share of female classmates is positive for females in all cases. In contrast, the impact on males is smaller, in most cases negative, and not statistically significant. For females in low income postcodes, exposure to a classroom that has 10 percentage points more females increases expected occupation-related wages by 14.6% of a s.d (s.e.=0.083). This effect size is substantial and around the same in magnitude as other educational interventions in secondary school, estimated based on the Greek context. For example, they are at the higher bound of the effect of providing feedback information to students (Goulas and Megalokonomou, 2021). For females in low quality schools, the effects are positive, although imprecise (estimated coefficient=0.079, s.e.=0.093). This pattern of increased occupation wages for females is a direct consequence of increasing females' likelihood of choosing a STEM university degree, which in turn, leads to a higher STEM occupational participation.

Random exposure to more female classmates positively influences test scores and university degree enrollments and has meaningful economic consequences for occupation sorting and earnings at adulthood, especially for females and students from disadvantaged contexts. If anything, males' occupation-related earnings seem unaffected in the longer run, which would contribute to a potential decrease in the gender pay gap.

9 Conclusion

This paper examines how improvements in classroom gender composition can affect educational individual outcomes in adolescence and early adulthood. Although this is not a research question that is addressed for the time, no evidence exists as to the effects and implications of those effects for disadvantaged students. The present analysis is based on exogenous variation in classroom assignment within schools in early high school in Greece. After initial assignment, students remained with mostly the same classmates throughout high school. We examine the effects and implications of random exposure to a higher share of female classmates on students' human capital formation, and in particular, current and subsequent test scores, and university level degree choices. We split subjects and degrees into STEM and humanities, and examine the widely claimed statement that human capital investment and knowledge accumulation is enhanced for females in STEM, in environments with more females. We also examine the mechanism behind those effects. Throughout the analysis, we remain particularly interested in the distinction between disadvantaged and non-disadvantaged students. Our study combines high school rich transcript data, attendance records, and longer-term educational choices data with novel administrative data obtained from the Ministry of Education for all students in Greece.

The results suggest that students who were exposed to a higher proportion of female classmates, due to the randomized classroom assignment in early high school, had substantially improved their end-of-year and subsequent STEM-related test scores, and are more likely to choose STEM degrees in college 2 years later. The benefits are larger for females than males. These effects persist into early adulthood and manifest in choosing degrees in post-secondary education that are linked to more lucrative or prestigious occupations. These effects are mostly driven by disadvantaged groups: students in low-quality schools, and students in lower-income neighborhoods. We examine a potential avenue through which gender classroom composition effects influences education and careers. We find that as the share of female classmates increases, males reduce their disruptive behavior (reflected in fewer unexcused absences), and create a more conducive learning environment that could motivate more students to attend classes (reflected in fewer excused absences).

This paper adds to the growing economic literature investigating female classmate effects on human capital outcomes. We offer two main contributions. First, most papers in the gender composition literature study those effects in selective or advantageous settings. To the best of our knowledge this is the first paper in this literature that attempts to estimate these effects for disadvantaged and non-disadvantaged groups and discuss those differences. This is because we have conducted a primary-data collection and we obtained data from a rich sample of schools and students that contains a large and diverse set of school qualities and socioeconomic profiles. This distinction is also policy relevant, and gaining new insights about whether educational inputs or interventions are more effective for disadvantaged students could make a potential contribution to the design, targeting, and implementation of new interventions, and a better resource allocation. The implications of our findings may also be relevant for other educational interventions.

Second, unlike other papers in the economic literature that use surveys and thus, self-reported measures of behavioral variables, we rely on administrative teacher-reported data to measure disruption and engagement using misconduct-related (unexcused) and parent-approved (excused) student class absences. We use these more objective behavioral measures to investigate the mechanisms behind gender classroom composition effects and also conduct a medication analysis. In addition, we discuss differences in explaining gender composition effects on human capital outcomes between disadvantaged and non-disadvantaged students. This will widen the array of successful policies that harness gender peer effects and help policymakers predict their success.

The impact on female classmates in early high school carries meaningful economic consequences, and it sharply affects the test scores in STEM subjects in high-stakes exams, postsecondary degree sorting, as well as occupation-related expected earnings in adulthood. Showing how the effect manifests over a persistent set of outcomes during adolescence, and early adulthood, sheds light on how policies can shape individuals' trajectories in life. Policymakers may direct resources to learning settings with fewer females and potentially higher incidence of disruptive behavior and lower student engagement. Our findings suggest that classroom interventions that reduce disruption are much more effective in disadvantaged or underserved environments.

References

Abdulkadiroğlu, A., J. D. Angrist, S. M. Dynarski, T. J. Kane, and P. A. Pathak (2011). Accountability and Flexibility in Public Schools: Evidence from Boston's Charters and Pilots. *Quarterly Journal of Economics 126*(2), 699–748.

Anelli, M. and G. Peri (2019). The Effects of High School Peers Gender on College Major, College Performance And Income. *Economic Journal 129*(618), 553–602.

Angrist, J. D. (2014). The Perils of Peer Effects. *Labour Economics 30*, 98–108.

Angrist, J. D. and V. Lavy (1999). Using maimonides' rule to estimate the effect of class size on scholastic achievement. *The Quarterly journal of economics 114*(2), 533–575.

Angrist, J. D., P. A. Pathak, and C. R. Walters (2013). Explaining Charter School Effectiveness. *American Economic Journal: Applied Economics 5*(4), 1–27.

Behrman, J., S. Fan, X. Wei, H. Zhang, and J. Zhang (2020). After-School Tutoring, Household Substitution and Student Achievement: Experimental Evidence from Rural China. *University of Pennsylvania Population Center Working Paper (PSC/PARC), 2020-36*.

Bifulco, R., J. M. Fletcher, and S. L. Ross (2011). The Effect of Classmate Characteristics on Post-secondary Outcomes: Evidence from the Add Health. *American Economic Journal: Economic Policy 3*(1), 25–53.

Black, A. R., F. Doolittle, P. Zhu, R. Unterman, and J. B. Grossman (2008). The Evaluation of Enhanced Academic Instruction in After-School Programs: Findings After the First Year of Implementation. NCEE 2008-4021. *National Center for Education Evaluation and Regional Assistance*.

Blau, F. D. and L. M. Kahn (2017, September). The Gender Wage Gap: Extent, Trends, and Explanations. *Journal of Economic Literature 55*(3), 789–865.

Brenøe, A. A. and U. Zölitz (2020). Exposure to More Female Peers Widens the Gender Gap in STEM Participation. *Journal of Labor Economics 38*(4), 1009–1054.

Buffington, C., B. Cerf, C. Jones, and B. A. Weinberg (2016, May). STEM Training and Early Career Outcomes of Female and Male Graduate Students: Evidence from UMETRICS Data Linked to the 2010 Census. *American Economic Review 106*(5), 333–38.

Buser, T., N. Peter, and S. C. Wolter (2017, May). Gender, Competitiveness, and Study Choices in High School: Evidence from Switzerland. *American Economic Review 107*(5), 125–30.

Carrell, S. E. and M. L. Hoekstra (2010). Externalities in the Classroom: How Children Exposed to Domestic Violence Affect Everyone's Kids. *American Economic Journal: Applied Economics 2*(1), 211–28.

Chetty, R., J. N. Friedman, N. Hilger, E. Saez, D. W. Schanzenbach, and D. Yagan (2011). How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project STAR. *Quarterly Journal of Economics 126*(4), 1593–1660.

Chung, B. and J. Zou (2020). Understanding Spillover of Peer Parental Education: Randomization Evidence and Mechanisms. Technical report.

Cohen, D. and J. Strayer (1996). Empathy in Conduct-Disordered and Comparison Youth. *Developmental Psychology 32*(6), 988.

Delaney, J. M. and P. J. Devereux (2019). Understanding Gender Differences in STEM: Evidence from College Applications. *Economics of Education Review 72*, 219–238.

Dinerstein, M., R. Megalokonomou, and C. Yannelis (2022). Human Capital Depreciation and Returns to Experience. *American Economic Review 112*(11), 3725–62.

Dobbie, W. and R. G. Fryer Jr (2013). Getting Beneath the Veil of Effective Schools: Evidence from New York City. *American Economic Journal: Applied Economics 5*(4), 28–60.

Duflo, E., P. Dupas, and M. Kremer (2008). Peer Effects and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya. *CEPR Discussion Paper No. DP7043*.

Fayer, S., A. Lacey, and A. Watson (2017). STEM Occupations: Past, Present, and Future. *Spotlight on Statistics 1*, 1–35.

Feld, J. and U. Zölitz (2017). Understanding Peer Effects: On the Nature, Estimation, and Channels of Peer Effects. *Journal of Labor Economics 35*(2), 387–428.

Figlio, D. N. (2007). Boys Named Sue: Disruptive Children and Their Peers. *Education Finance and Policy 2*(4), 376–394.

Fryer, Roland G., J. and S. D. Levitt (2010). An Empirical Analysis of the Gender Gap in Mathematics. *American Economic Journal: Applied Economics 2*(2), 210–40.

Gelbach, J. B. (2016). When Do Covariates Matter? And Which Ones, and How Much? *Journal of Labor Economics 34*(2), 509–543.

Gong, J., Y. Lu, and H. Song (2018). The Effect of Teacher Gender on Students Academic and Noncognitive Outcomes. *Journal of Labor Economics 36*(3), 743–778.

Gong, J., Y. Lu, and H. Song (2021). Gender Peer Effects on Students Academic and Noncognitive Outcomes Evidence and Mechanisms. *Journal of Human Resources 56*(3), 686–710.

Goulas, S., S. Griselda, and R. Megalokonomou (2022). Comparative Advantage and Gender Gap in STEM. *Journal of Human Resources*.

Goulas, S., S. Griselda, and R. Megalokonomou (2023). Compulsory Class Attendance Versus Autonomy. *Journal of Economic Behavior & Organization 212*, 935–981.

Goulas, S., B. Gunawardena, and R. Megalokonomou (2023). Same-Gender Peer Role Models in Education. *Mimeo*.

Goulas, S. and R. Megalokonomou (2019). Which Degrees Do Students Prefer During Recessions? *Empirical Economics 56*(6), 2093–2125.

Goulas, S. and R. Megalokonomou (2020). Marathon, Hurdling, or Sprint? The Effects of Exam Scheduling on Academic Performance. *BE Journal of Economic Analysis & Policy 20*(2).

Goulas, S. and R. Megalokonomou (2021). Knowing who you actually are: The effect of feedback on short- and longer-term outcomes. *Journal of Economic Behavior & Organization 183*, 589–615.

Goulas, S., R. Megalokonomou, and Y. Zhang (2022). Female Neighbors, Test Scores, and Careers. *CESifo Working Paper*.

Hanushek, E. A., J. Kain, D. O'Brien, and S. G. Rivkin (2005). The Market for Teacher Quality.

Hanushek, E. A. and S. G. Rivkin (2010). Generalizations about Using Value-added Measures of Teacher Quality. *American Economic Review 100*(2), 267–271.

Heckman, J., R. Pinto, and P. Savelyev (2013). Understanding the Mechanisms through Which an Influential Early Childhood Program Boosted Adult Outcomes. *American Economic Review 103*(6), 2052–86.

Hill, A. J. (2017). The Positive Influence of Female College Students on Their Male Peers. *Labour Economics 44*, 151–160.

Hoxby, C. (2000). Peer Effects in the Classroom: Learning From Gender and Race Variation. Technical report, National Bureau of Economic Research.

Hu, F. (2015). Do Girl Peers Improve Your Academic Performance? *Economics Letters 137*, 54–58.

Hyde, J. S. and J. E. Mertz (2009). Gender, Culture, and Mathematics Performance. *Proceedings of the National Academy of Sciences 106*(22), 8801–8807.

Kara, E., M. Tonin, and M. Vlassopoulos (2021). Class Size Effects in Higher Education: Differences across STEM and non-STEM Fields. *Economics of Education Review 82*, 102104.

Keenan, K. and D. Shaw (1997). Developmental and Social Influences on Young Girls' Early Problem Behavior. *Psychological Bulletin 121*(1), 95.

Krueger, A. B. (1999). Experimental Estimates of Education Production Functions. *The Quarterly Journal of Economics 114*(2), 497–532.

Landaud, F., S. T. Ly, and É. Maurin (2020). Competitive Schools and the Gender Gap in the Choice of Field of Study. *Journal of Human Resources 55*(1), 278–308.

Lavy, V. and R. Megalokonomou (2023a). Longer-Term Effects of Teachers: Evidence From a Random Assignment of Teachers to Students.

Lavy, V. and R. Megalokonomou (2023b). Persistency in Teachers Grading Bias and Effects on Longer-term Outcomes: University Admissions Exams and Choice of Field of Study. *American Economic Journal: Applied Economics, Forthcoming*.

Lavy, V., M. D. Paserman, and A. Schlosser (2012). Inside the Black Box of Ability Peer Effects: Evidence from Variation in the Proportion of Low Achievers in the Classroom. *Economic Journal 122*(559), 208–237.

Lavy, V. and A. Schlosser (2011). Mechanisms and Impacts of Gender Peer Effects at School. *American Economic Journal: Applied Economics 3*(2), 1–33.

Lazear, E. P. (2001). Educational Production. *Quarterly Journal of Economics 116*(3), 777–803.

Manski, C. F. (1993). Identification of Endogenous Social Effects: The Reflection Problem. *Review of Economic Studies 60*(3), 531–542.

McKee, G., K. R. Sims, and S. G. Rivkin (2015). Disruption, Learning, and the Heterogeneous Benefits of Smaller Classes. *Empirical Economics 48*, 1267–1286.

Megalokonomou, R. and Y. Zhang (2022). How Good Am I? Effects and Mechanisms Behind Salient Ranks. *CESifo Working Paper*.

Mouganie, P. and Y. Wang (2020). High-Performing Peers and Female STEM Choices in School. *Journal of Labor Economics 38*(3), 000–000.

OECD (2016). PISA 2015 Results (Volume I): Excellence and Equity in Education.

OECD (2017). OECD Indicators Education at a Glance.

OECD (2018a). *Education for a Bright Future in Greece*. OECD Publishing.

OECD (2018b). OECD Employment Outlook 2018. *OECD Publishing, Paris*.

Papay, J. P. and M. A. Kraft (2015). Productivity Returns to Experience in the Teacher Labor Market: Methodological Challenges and New Evidence on Long-term Career Improvement. *Journal of Public Economics 130*, 105–119.

Rozek, C. S., G. Ramirez, R. D. Fine, and S. L. Beilock (2019). Reducing Socioeconomic Disparities in the STEM Pipeline Through Student Emotion Regulation. *Proceedings of the National Academy of Sciences 116*(5), 1553–1558.

Sacerdote, B. (2001). Peer Effects with Random Assignment: Results for Dartmouth Roommates. *Quarterly Journal of Economics 116*(2), 681–704.

Schneeweis, N. and M. Zweimüller (2012). Girls, Girls, Girls: Gender Composition And Female School Choice. *Economics of Education Review 31*(4), 482–500.

Schøne, P., K. von Simson, and M. Strøm (2016). Girls Helping Girls: The Impact of Female Peers on Grades and Educational Choices. *Available at SSRN 2760782*.

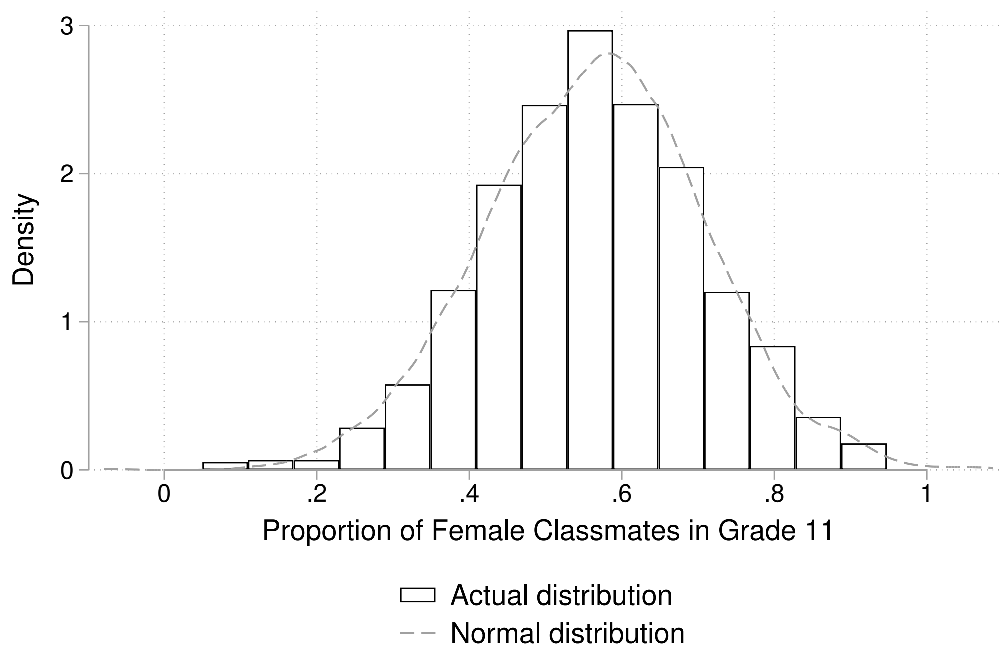
Whitmore, D. (2005). Resource and Peer Impacts on Girls' Academic Achievement: Evidence from a Randomized Experiment. *American Economic Review 95*(2), 199–203.

Xu, D., Q. Zhang, and X. Zhou (2020). The Impact of Low-Ability Peers on Cognitive and Non-Cognitive Outcomes: Random Assignment Evidence on the Effects and Operating Channels. *Journal of Human Resources*, *Issue 6*, *Volume 58*.

Zimmerman, D. J. (2003). Peer Effects in Academic Outcomes: Evidence from a Natural Experiment. *Review of Economics and Statistics 85*(1), 9–23.

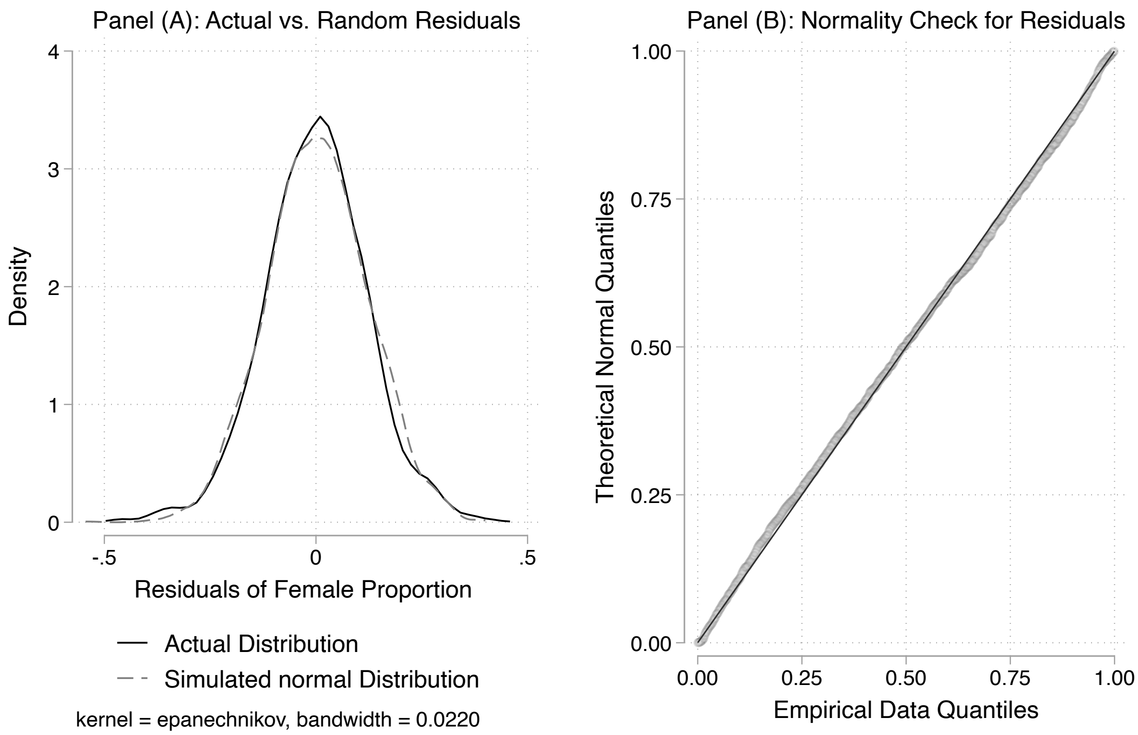
Zölitz, U. and J. Feld (2021). The Effect of Peer Gender on Major Choice in Business School. *Management Science 67*(11), 6963–6979.

Figure 1: Histogram of the Proportion of Females in Classrooms



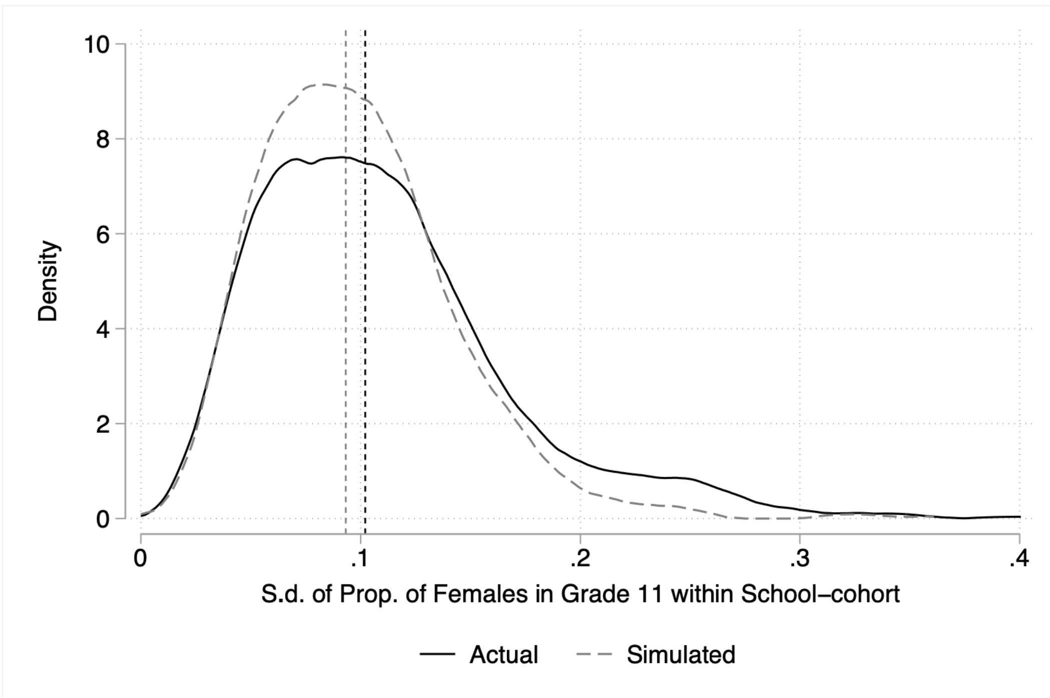
Notes: The figure presents the distribution of the proportion of females across classrooms in 11th grade and the simulated kernel density of the normal distribution using the actual mean (0.563) and standard deviation (0.148).

Figure 2: Residuals of Classroom-level Female Share after Controlling for Unobserved School-Cohort Effects and Test for Normality of Residuals



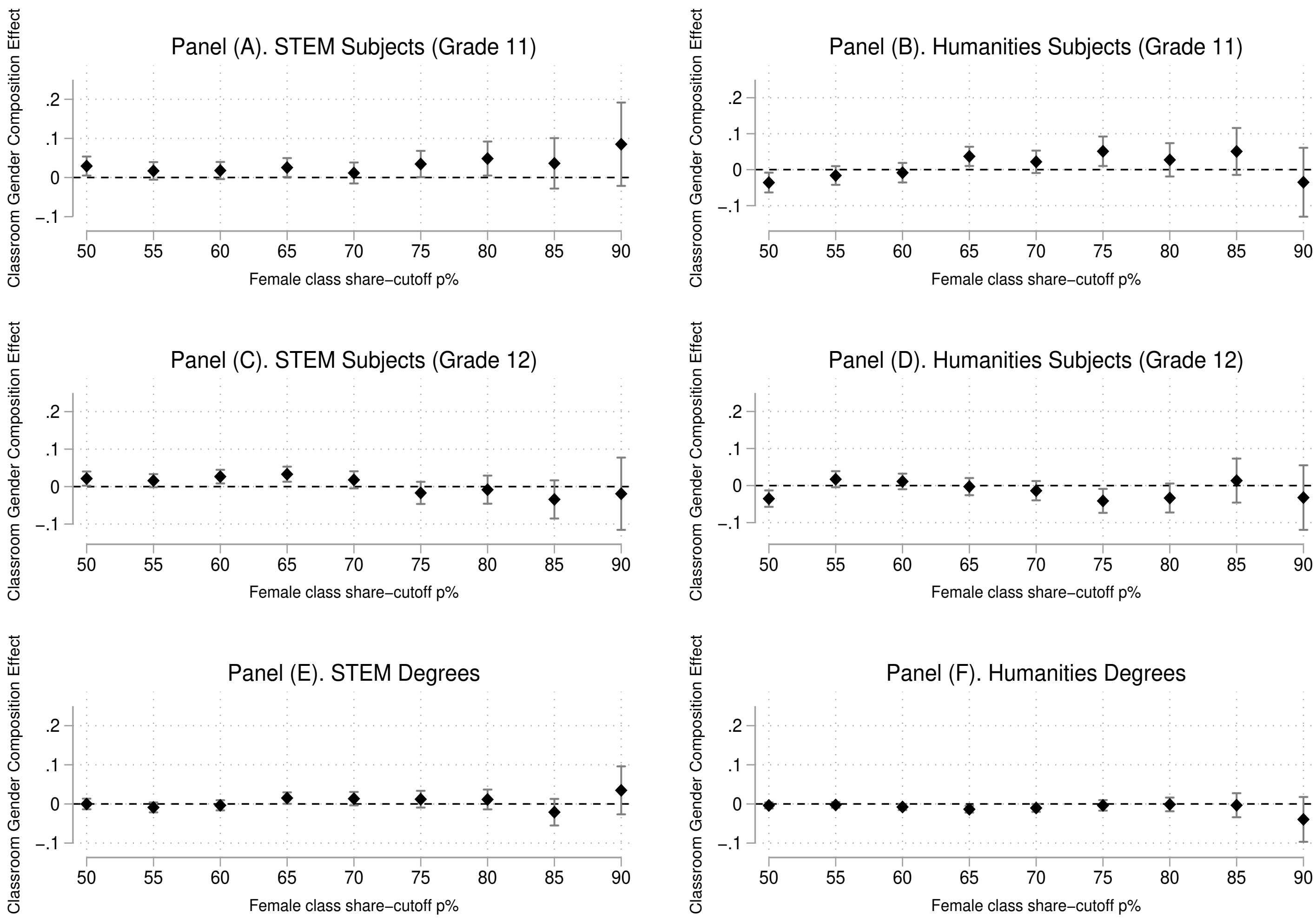
Notes: In Panel (A), the solid curve represents the kernel density of the residuals from a regression of the share of female classmates in grade 11 on school-cohort fixed effects. The dashed curve in Panel (A) represents a simulated normal distribution with the same mean (0), the standard deviation (0.148), and the same number of unique school-by-cohort classrooms (2,517) as the actual residuals' distribution. Panel (B) shows empirical female classmate share residuals which follow a normal distribution (45-degree line), and the residuals obtained by regressing the share of female classmates on school-cohort fixed effects.

Figure 3: Simulated and Actual Standard Deviation in the Proportion of Females across Classes within School-Cohorts



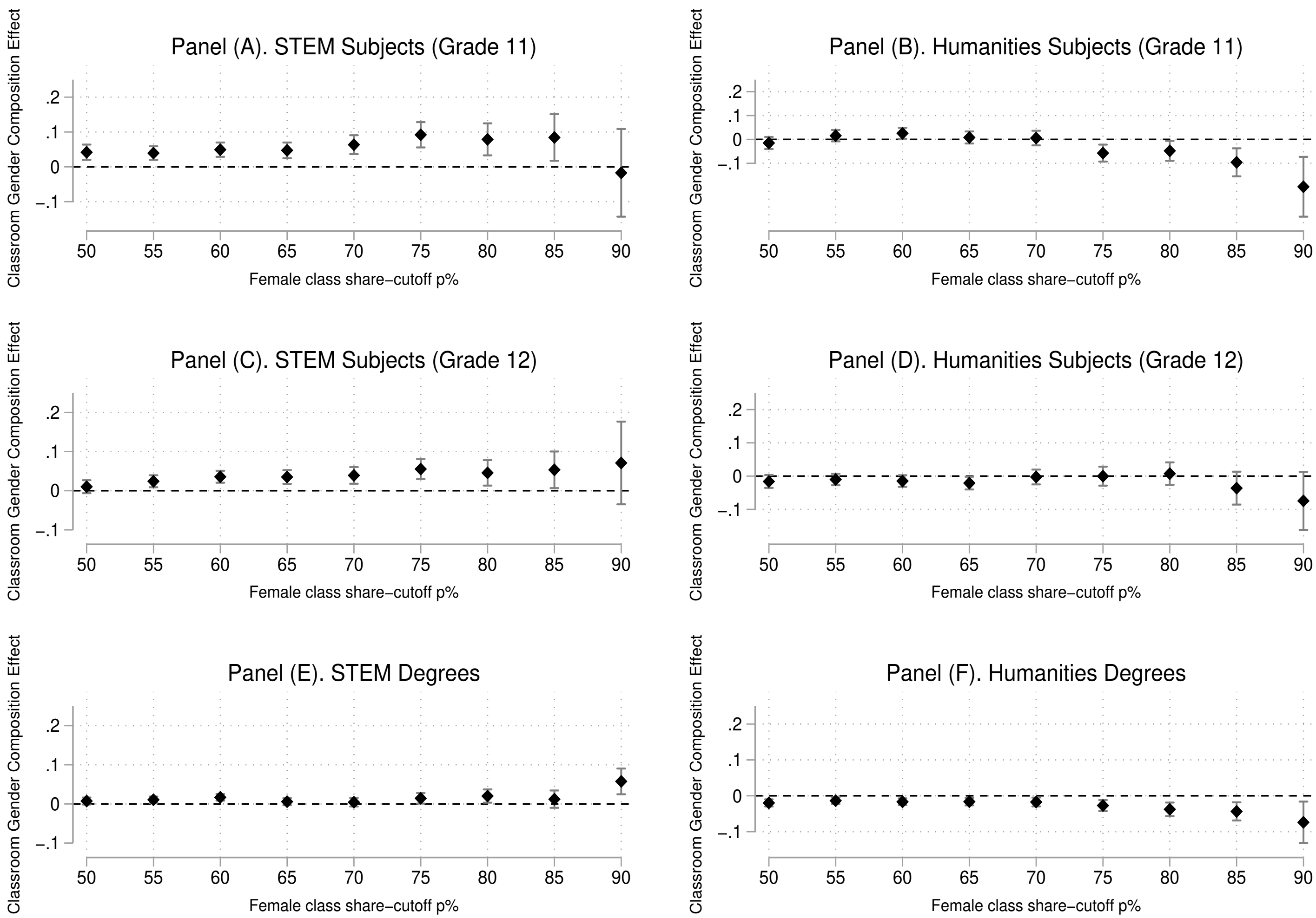
Notes: The figure presents the kernel density of the actual standard deviation in the proportion of females in grade 11 within school-cohort and the simulated standard deviation of females within school-cohort. Vertical lines indicate the median of each distribution. To produce the simulated standard deviation, we randomly generate the gender of students using a binomial distribution with p equal to the actual proportion of females in each school-cohort. We then compute the within-school-cohort standard deviation of the artificially generated proportion of female peers across classrooms and plot it along with the actual one.

Figure 4: Nonlinear Impact of Classroom Gender Composition on Performance and Enrollment in STEM and Humanities for Males



Notes: Scatter points represent estimated coefficients of the female classmates share being larger than p%, where p is a set of integers that includes every fifth integer from 50 to 90. Vertical bars represent the 95% confidence intervals of the coefficients.Test scores in STEM subjects in 11^{th} and 12^{th} grade include the average test scores in Mathematics and Physics. Test scores in Humanities subjects in 11^{th} and 12^{th} grade include the average test scores in Language and History. STEM degrees include degrees in Mathematics, Science, Engineering, and Computer Science. Humanities degrees include degree in Modern Languages, Literature, Philosophy, History, Archaeology, Anthropology, Religion, and Art.

Figure 5: Nonlinear Impact of Classroom Gender Composition on Performance and Enrollment in STEM and Humanities for Females



Notes: Scatter points represent estimated coefficients of the female classmates share being larger than *p*%, where *p* is a set of integers that includes every fifth integer from 50 to 90. Vertical bars represent the 95% confidence intervals of the coefficients. Test scores in STEM subjects in 11^{th} and 12^{th} grade include the average test scores in Mathematics and Physics. Test scores in Humanities subjects in 11^{th} and 12^{th} grade include the average test scores in Language and History. STEM degrees include degrees of Mathematics, Science, Engineering, and Computer Science. Humanities degrees include degree in Modern Languages, Literature, Philosophy, History, Archaeology, Anthropology, Religion, and Art.

Table 1: Descriptive Statistics of the Sample

	Mean	Std. Dev.	Min.	Max.
A. Student-level Characteristics:				
Age (Grade 11)	16.900	0.450	14	35
Female (1=Yes)	0.560	0.496	0	1
Born in 1^{st} Quarter of Birth Year (1=Yes)	0.142	0.349	0	1
GPA (Grade 10)	14.927	2.539	0	20
Track Choice (Grade 11)				
Classics	0.362	0.481	0	1
Science	0.285	0.451	0	1
Exact Science	0.353	0.478	0	1
Absences (Grade 11)				
Total	49.178	26.245	0	164
Excused	19.632	18.969	0	158
Unexcused	29.514	13.459	0	50
Absences (Grade 12)				
Total	74.273	29.735	0	199
Excused	40.066	23.970	0	175
Unexcused	34.127	13.448	0	65
11^{th} *Grade Performance*				
STEM Subjects	15.051	2.942	0	20
Humanities Subjects	15.494	2.525	0	20
12^{th} *Grade Performance*				
STEM Subjects	14.548	3.654	5	20
Humanities Subjects	14.627	2.679	6	20
University Degree Choice				
STEM Degrees	0.230	0.421	0	1
Humanities Degrees	0.185	0.388	0	1
B. Class-level Characteristics (Grade 11):				
Born in 1^{st} Quarter of Calendar Year (1=Yes)	0.136	0.124	0	1
Age	16.908	0.155	16	19
Prop. Female Peers	0.563	0.148	0	1
Classroom Size	16.266	4.728	6	31
C. School-level Characteristics:				
Postcode Income (Euro, 2009)	22894.995	8127.427	13005	66521
No. of Classrooms (Grade 11)	3.558	1.139	2	6
School Enrollment (Grade 11)	59.923	28.136	13	126

Notes: The data span 8 academic years from 2002 to 2009. The sample contains 104 public schools and 2,517 unique classrooms. The number of students is 43,451. "Born in 1st Quarter of Calendar Year" is a binary indicator that takes the value of one if a student is born in the first quarter of the calendar year. Students have to enroll in a track at the beginning of grade 11. They have three track options: Classics, Science, and Exact Science. All schools offer all tracks. Absences are measured in hours.

Table 2: Balancing Tests of the Proportion of Female Classmates

	Prop. Female Peers (Grade 11)
Female (1=Yes)	0.0013
	(0.0021)
Age (Grade 11)	0.0013
	(0.0012)
Born in 1^{st} Quarter of Birth Year (1=Yes)	-0.0016
	(0.0016)
GPA (Grade 10)	0.0002
	(0.0003)
N	43,451
p-value for Joint F Test for Individual Characteristics	0.742
School-cohort FE	✓

Notes: Each cell reports the estimated effect of the related student characteristic (reported vertically) from a separate regression in which the dependent variable is the share of females in the classroom in grade 11. The lower panel shows the P-value for the F-statistic of the joint significance of student characteristics. It comes from the regression of the dependent variable (proportion of female classmates) on all student characteristics, conditional on school-cohort fixed effects. Standard errors clustered at classroom level are reported in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table 3: Summary of Monte Carlo Simulation

Share of the Actual Standard Deviation within 90% Empirical CI	0.865
Share of the Actual Standard Deviation within 95% Empirical CI	0.908
Share of the Actual Standard Deviation within 99% Empirical CI	0.955

Notes: This table summarizes the share of the actual standard deviation of the share of female classmates in grade 11 that falls within the 90%, 95% and 99% of the empirical confidence interval (CI) generated by the simulated standard deviation of the share of females in each school-cohort. To produce the empirical confidence intervals, we randomly generate the gender of students using a binomial distribution with *p* equal to the actual proportion of females in each school-cohort. We then compute the within school-cohorts standard deviation of the proportion of female classmates based on the simulated genders. We repeat this process 1,000 times. For each school-cohort, we then obtain the 95.0% (97.5% or 99.5%) and 5.0% (2.5% or 0.5%) percentile of the simulated standard deviation in the proportion of females as the upper and lower bounds of the 90% (95% or 99%) empirical confidence interval.

Table 4: Main Estimates and Placebo Effects of Female Classmates on Test Scores and Enrollment in STEM and Humanities

	Males			**Females**		
	Main Estimates		**Placebo**	**Main Estimates**		**Placebo**
	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variables:						
Panel A: 11th-grade Performance						
STEM Subjects	0.129***	0.132***	0.010	0.254***	0.257***	0.007
	(0.049)	(0.049)	(0.037)	(0.046)	(0.046)	(0.029)
Mean of Y	-0.015	-0.015	-0.015	0.012	0.012	0.012
N	19,113	19,113	19,113	24,337	24,337	24,337
Humanities Subjects	-0.023	-0.022	-0.035	-0.057	-0.051	0.007
	(0.056)	(0.056)	(0.037)	(0.049)	(0.049)	(0.029)
Mean of Y	-0.015	-0.015	-0.015	0.012	0.012	0.012
N	19,113	19,113	19,113	24,337	24,337	24,337
Panel B: 12th-grade Performance						
STEM Subjects	0.096**	0.093**	0.052	0.119***	0.121***	0.003
	(0.042)	(0.042)	(0.035)	(0.034)	(0.034)	(0.029)
Mean of Y	-0.015	-0.015	-0.015	0.012	0.012	0.012
N	19,113	19,113	19,113	24,337	24,337	24,337
Humanities Subjects	-0.061	-0.059	0.048	-0.075**	-0.074**	-0.002
	(0.041)	(0.041)	(0.038)	(0.036)	(0.036)	(0.029)
Mean of Y	-0.015	-0.015	-0.015	0.012	0.012	0.012
N	19,113	19,113	19,113	24,337	24,337	24,337
Panel C: University Degree Choice						
STEM Degrees	0.037	0.036	-0.033	0.044**	0.043**	-0.009
	(0.028)	(0.028)	(0.030)	(0.019)	(0.019)	(0.019)
Mean of Y	-0.015	-0.015	-0.015	0.012	0.012	0.012
N	19,113	19,113	19,113	24,337	24,337	24,337
Humanities Degrees	-0.036**	-0.036**	0.005	-0.092***	-0.093***	-0.010
	(0.018)	(0.018)	(0.017)	(0.021)	(0.021)	(0.021)
Mean of Y	-0.015	-0.015	-0.015	0.012	0.012	0.012
N	19,113	19,113	19,113	24,337	24,337	24,337
School-cohort FE & Student-level controls	✓	✓	✓	✓	✓	✓
Classroom-level controls		✓	✓		✓	✓

Notes: The table reports the estimated effects of the share of female classmates in 11th grade on students' scholastic outcomes in 11th and 12th grade and students' university degree major choice in columns (1)-(2) and columns (4)-(5). Columns (3) and (6) report the results from placebo tests in which the true share of female classmates is replaced with the placebo female classmates share, which is generated through reshuffling students within school-cohorts. We show the means of the dependent variables below the standard errors. Student-level controls include age, born in the 1st quarter of birth year indicator, prior track choice, previous GPA, and share of female classmates in 10th grade. Classroom-level controls include classroom-level mean age, born in 1st quarter of birth year indicator, and classroom size. Regressions in all columns include school-cohort fixed effects. The regressions in columns in (2)-(3) and (5)-(6) additionally include classroom-level controls. STEM subjects in 11th and 12th grade include Mathematics and Physics. Humanities subjects in 11th and 12th grade include Language and History. STEM degrees include Mathematics, Science, Engineering, and Computer Science. Humanities degrees include Modern Languages, Literature, Philosophy, History, Archaeology, Anthropology, Religion, and Art. Standard errors clustered at classroom level are reported in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table 5: Heterogeneous Effects of Gender Composition Effect by Neighborhood Income and School Quality

	Neighborhood Income				**School Quality**			
	Below Median		Above Median		Below Median		Above Median	
	Males	Females	Males	Females	Males	Females	Males	Females
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent Variables:								
Panel A: 11th-grade Performance								
STEM Subjects	0.184***	0.338***	0.073	0.120*	0.137*	0.260***	0.119*	0.221***
	(0.066)	(0.061)	(0.078)	(0.071)	(0.071)	(0.065)	(0.069)	(0.064)
N	8,671	11,600	10,441	12,736	7,249	9,784	11,863	14,552
Humanities Subjects	0.070	-0.073	-0.154	-0.057	0.139*	-0.008	-0.213**	-0.108
	(0.069)	(0.062)	(0.098)	(0.087)	(0.073)	(0.067)	(0.084)	(0.073)
N	8,671	11,600	10,441	12,736	7,249	9,784	11,863	14,552
Panel B: 12th-grade Performance								
STEM Subjects	0.115**	0.172***	0.062	0.026	0.067	0.165***	0.098*	0.056
	(0.054)	(0.043)	(0.069)	(0.057)	(0.059)	(0.047)	(0.058)	(0.051)
N	8,671	11,601	10,442	12,736	7,249	9,785	11,864	14,552
Humanities Subjects	-0.033	-0.098**	-0.117*	0.006	-0.115**	-0.128***	0.007	0.003
	(0.052)	(0.044)	(0.069)	(0.062)	(0.056)	(0.047)	(0.060)	(0.056)
N	8,671	11,601	10,442	12,736	7,249	9,785	11,864	14,552
Panel C: University Degree Choice								
STEM Degrees	0.018	0.045*	0.050	0.003	0.011	0.051**	0.054	-0.003
	(0.035)	(0.023)	(0.047)	(0.034)	(0.038)	(0.024)	(0.043)	(0.030)
N	8,671	11,601	10,442	12,736	7,249	9,785	11,864	14,552
Humanities Degrees	-0.046*	-0.090***	-0.025	-0.078**	-0.067***	-0.084***	0.009	-0.050*
	(0.024)	(0.027)	(0.027)	(0.031)	(0.024)	(0.027)	(0.026)	(0.030)
N	8,671	11,601	10,442	12,736	7,249	9,785	11,864	14,552
School-cohort FE & Student-level controls	✓	✓	✓	✓	✓	✓	✓	✓
Classroom-level controls	✓	✓	✓	✓	✓	✓	✓	✓

Notes: The table reports heterogeneous effects by postcode income and school percentile rank of the share of females in the classroom in grade 11 on students' test scores in grades 11 and 12 and students' postsecondary degree choice. We stratify the sample by the median of postcode income (median = 20,764 euros) and by the median of percentile rank (median school percentile ranking = 64.57) separately, then obtain the estimated effects of the share of females in the classroom in each restricted sample. Columns (1) and (2) (columns (3) and (4)) report the estimated effect of the female share in the classroom for male and female students who reside in neighborhoods with postcode income below the median (above the median), respectively. Columns (5) and (6) (columns (7) and (8)) report the estimated effect of the female share in the classroom for male and female students who study in schools with percentile ranks below the median (above the median). Student-level controls include age, born in 1st quarter of birth year indicator, prior track choice, GPA in grade 10, and the share of female classmates in grade 10. Classroom-level controls include the classroom-level mean age, share of classmates born in 1st quarter of birth year, and classroom size. Standard errors clustered at classroom level are reported in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table 6: Classroom Gender Composition Effect on Total, Excused and Unexcused Student Absences

Panel A	**Males**			**Females**		
Dependent Variables (Grade 11):	Total Absences	Excused Absences	Unexcused Absences	Total Absences	Excused Absences	Unexcused Absences
	(1)	(2)	(3)	(4)	(5)	(6)
Prop. Female Peers	-0.726	1.429	-2.138*	-4.558**	-3.039*	-1.477
	(2.270)	(1.563)	(1.219)	(2.157)	(1.557)	(1.052)
N	14,401	14,401	14,454	18,443	18,443	18,514
Mean of Y	48.432	18.091	30.312	49.760	20.835	28.891
School-cohort FE & Student-level controls	✓	✓	✓	✓	✓	✓
Class-level controls	✓	✓	✓	✓	✓	✓

Panel B	**Males**			**Females**		
Dependent Variables (Grade 12):	Total Absences	Excused Absences	Unexcused Absences	Total Absences	Excused Absences	Unexcused Absences
	(1)	(2)	(3)	(4)	(5)	(6)
Prop. Female Peers	-1.505	1.245	-2.887***	-3.295*	-1.620	-1.674
	(2.053)	(1.638)	(1.027)	(1.965)	(1.648)	(1.032)
N	14,406	14,407	14,526	18,423	18,428	18,579
Mean of Y	71.922	37.027	34.820	76.110	42.442	33.585
School-cohort FE & Student-level controls	✓	✓	✓	✓	✓	✓
Class-level controls	✓	✓	✓	✓	✓	✓

Notes: The table reports the estimated effects of the female peers classroom share in 11th grade on students' attendance pattern in 11th grade (Panel A) and 12th grade (Panel B). Student-level controls include age, born in 1st quarter of birth year indicator, prior track choice, prior GPA, and female peers classroom share in 10th grade. Classroom-level controls include classroom-level mean age, born in 1st quarter of birth year indicator, and classroom size. Standard errors clustered at classroom level are reported in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table 7: How Much of the Gender Composition Effect is Explained by Changes in Behavior?

Channels	**% of Effect Explained by Channels**			
	Unexcused Absences (1)	Excused Absences (2)	Explained (3)	Unexplained (4)
	Panel A: All Students in the Sample			
STEM Subjects (Grade 11)				
Males	6.791	0.567	7.359	92.641
Females	2.227	0.740	2.967	97.033
STEM Subjects (Grade 12)				
Males	5.32	0.714	6.034	93.966
Females	1.352	0.298	1.65	98.35
	Panel B: Students with Low Household Income			
STEM Subjects (Grade 11)				
Males	8.095	0.346	8.441	91.559
Females	1.513	0.421	1.934	98.066
STEM Subjects (Grade 12)				
Males	6.423	0.320	6.743	93.257
Females	1.81	0.290	2.1	97.9
	Panel C: Students with High Household Income			
STEM Subjects (Grade 11)				
Males	3.639	0.891	4.531	95.469
Females	3.998	0.739	4.737	95.263
STEM Subjects (Grade 12)				
Males	.087	2.528	2.615	97.385
Females	.789	0.172	.96	99.04
	Panel D: Students in Low-quality Schools			
STEM Subjects (Grade 11)				
Males	15.477	0.155	15.632	84.368
Females	4.784	0.481	5.265	94.735
STEM Subjects (Grade 12)				
Males	7.262	0.711	7.973	92.027
Females	.674	0.747	1.421	98.579
	Panel E: Students in High-quality Schools			
STEM Subjects (Grade 11)				
Males	.997	0.823	1.82	98.18
Females	.229	0.677	.906	99.094
STEM Subjects (Grade 12)				
Males	1.133	0.580	1.713	98.287
Females	2.513	0.033	2.546	97.454

Notes: We use a mediation analysis in each panel and quantify the relative importance of each mechanism on explaining test scores in STEM subjects in grades 11 and 12 (Chung and Zou, 2020; Gong, Lu, and Song, 2018, 2021).

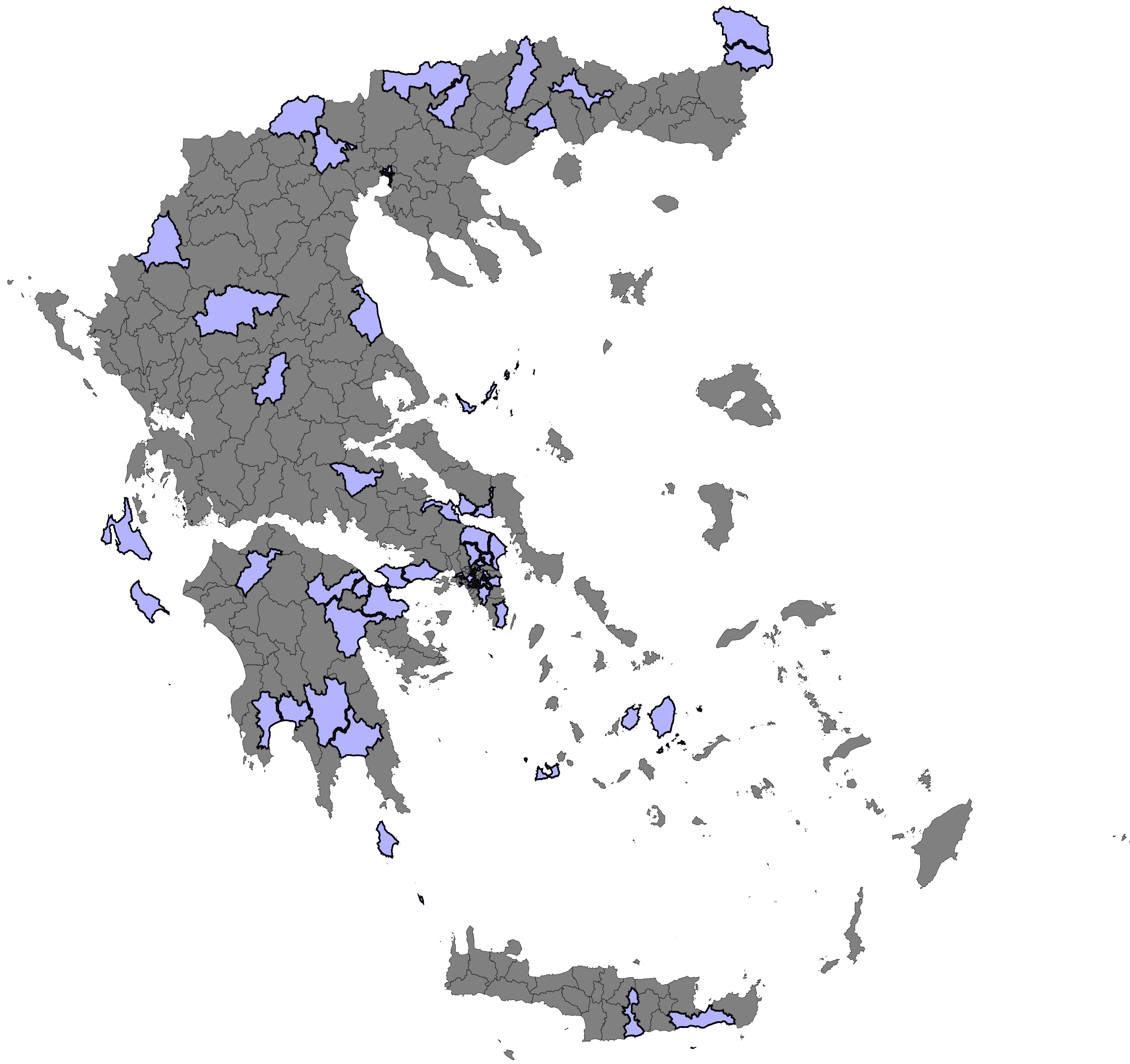
Table 8: Longer-Term Effects of Gender Classroom Composition on Occupation-Related Wages

	All		Low Income		High Income		Low Quality		High Quality	
	Males (1)	Females (2)	Males (3)	Females (4)	Males (5)	Females (6)	Males (7)	Females (8)	Males (9)	Females (10)
Prop. Female Peers	-0.059	0.075	-0.034	0.146*	-0.067	0.092	0.006	0.079	-0.118	0.072
	(0.080)	(0.066)	(0.107)	(0.083)	(0.113)	(0.115)	(0.114)	(0.093)	(0.113)	(0.096)
N	13,948	19,265	6,205	8,970	7,743	10,295	4,977	7,210	8,971	12,055
School-cohort FE & Student-level controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Class-level controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Notes: The table reports the estimated effects of the share of female classmates in grade 11 on students' occupation-related expected wages. The outcome is the standardized occupation-related expected wage and has a mean equal to 0 and a standard deviation equal to 1. We use the 2003 Labour Force Survey to map each college major into the most related occupation and salaries. Columns (1) and (2) show the effects for male and female students, respectively. The estimated effects of gender classroom composition are shown for different subgroups in columns (3)-(10). Columns (3)-(4) show the estimated effects for low-income neighborhoods for males and females. Columns (5)-(6) show the estimated effects for high-income neighborhoods for males and females. Columns (7)-(8) show the estimated effects for low-quality schools for males and females. Columns (9)-(10) show the estimated effects for high-quality schools for males and females. Student-level controls include age, born in 1^{st} quarter of birth year indicator, prior track choice, prior GPA, and female peers classroom share in 10^{th} grade. Classroom-level controls include classroom-level mean age, born in 1st quarter of birth year indicator, and classroom size. Standard errors clustered at classroom level are reported in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

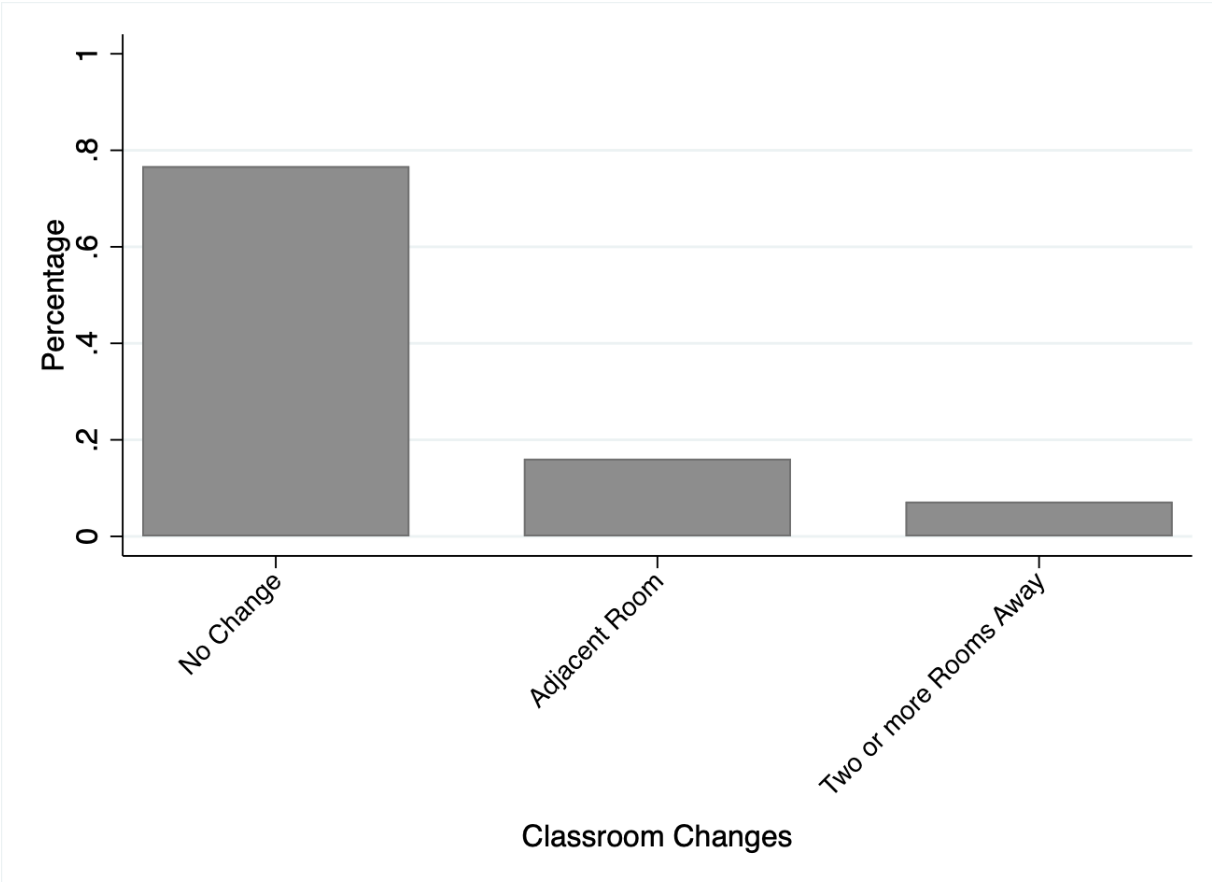
A Figure Appendices

Figure A.1: Map of Schools in the Sample



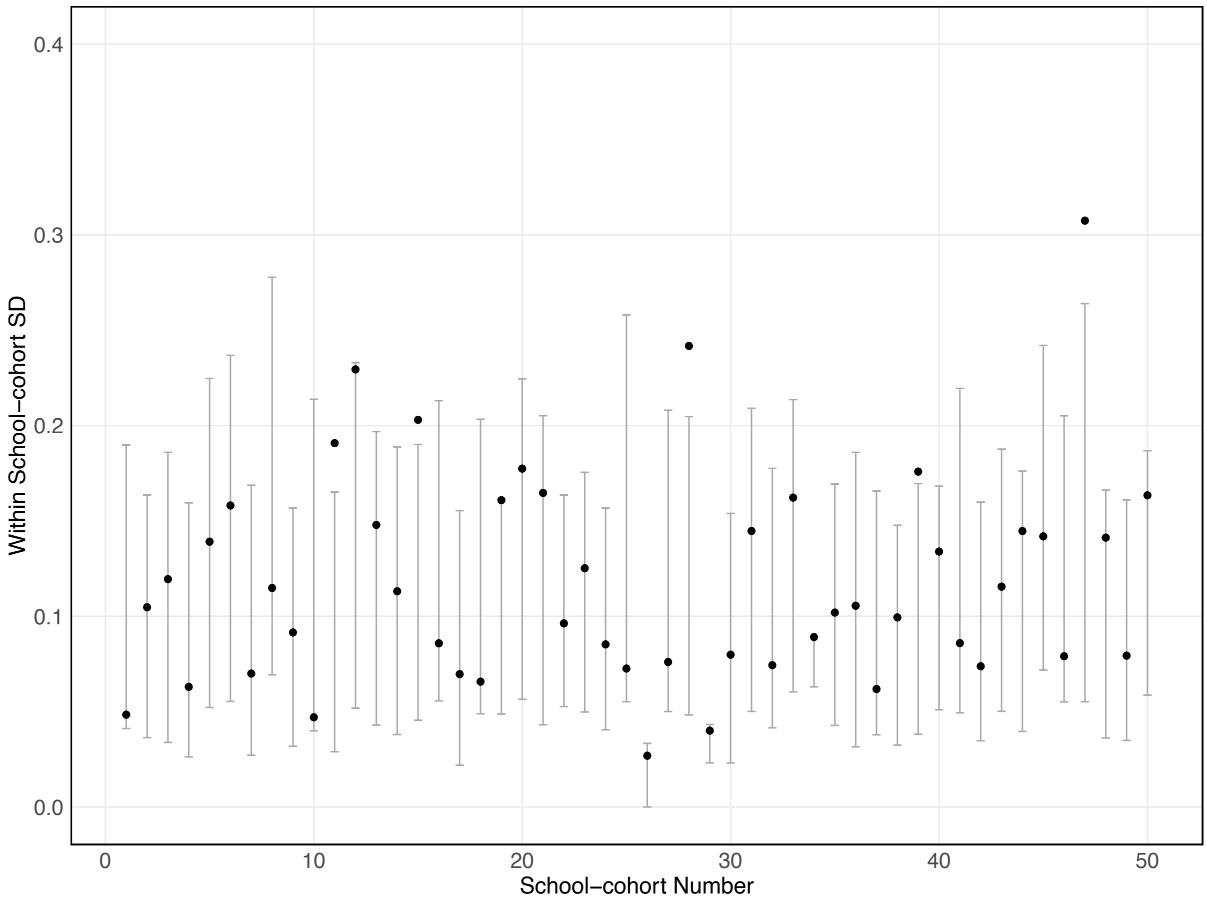
Notes: This figure shows the counties in which high schools in our sample are located.

Figure A.2: How Many Students Change Classrooms between Grade 10 and Grade 11?



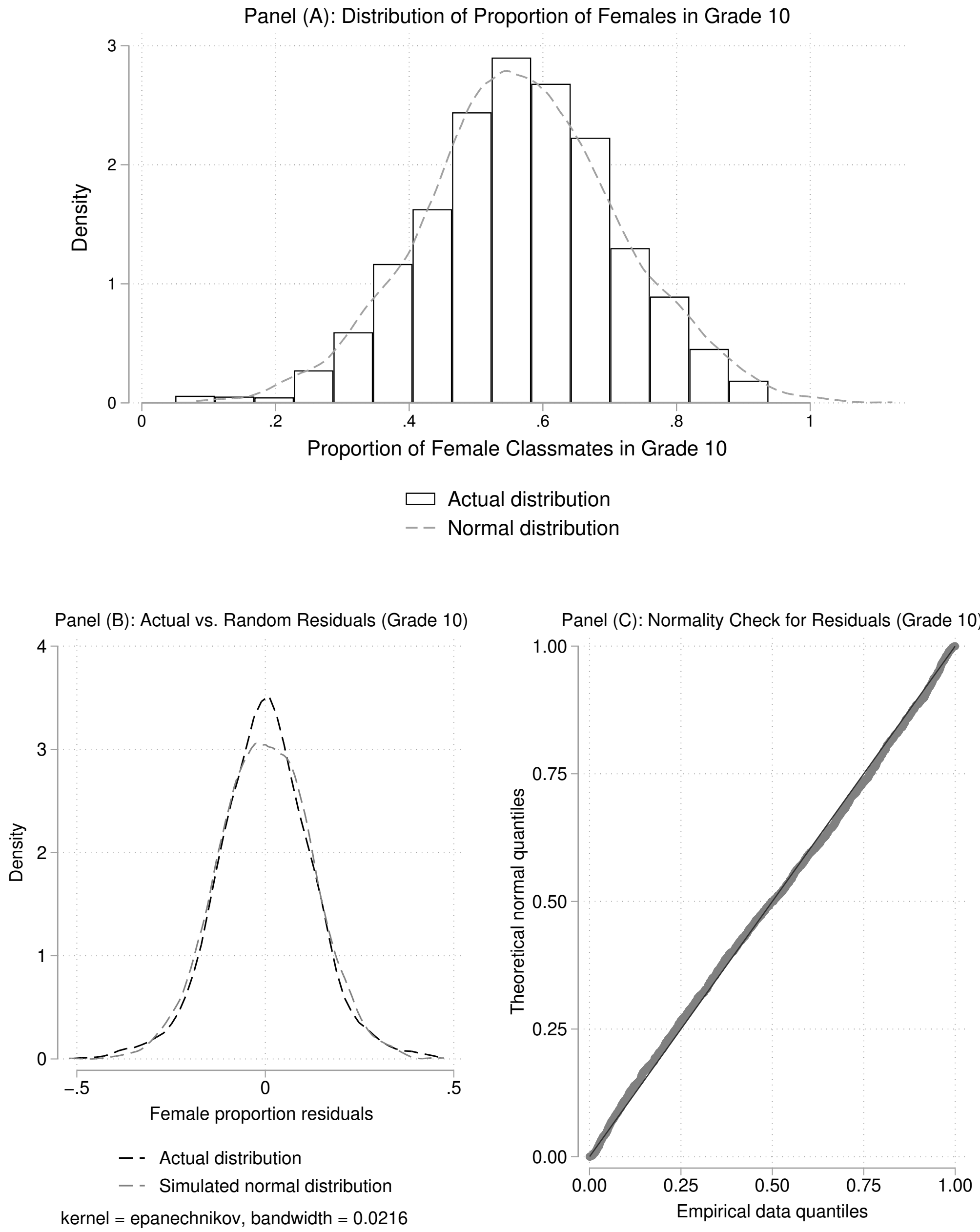
Notes: The figure shows the density of students transferring to a different classroom in grade 11 from the one assigned in grade 10. The vast majority of students remain in the same classroom between grades 10 and 11. Most transitions take place between adjacent classrooms.

Figure A.3: Monte Carlo Simulations of Standard Deviations in the Proportion of Females Classmates within School-Cohort



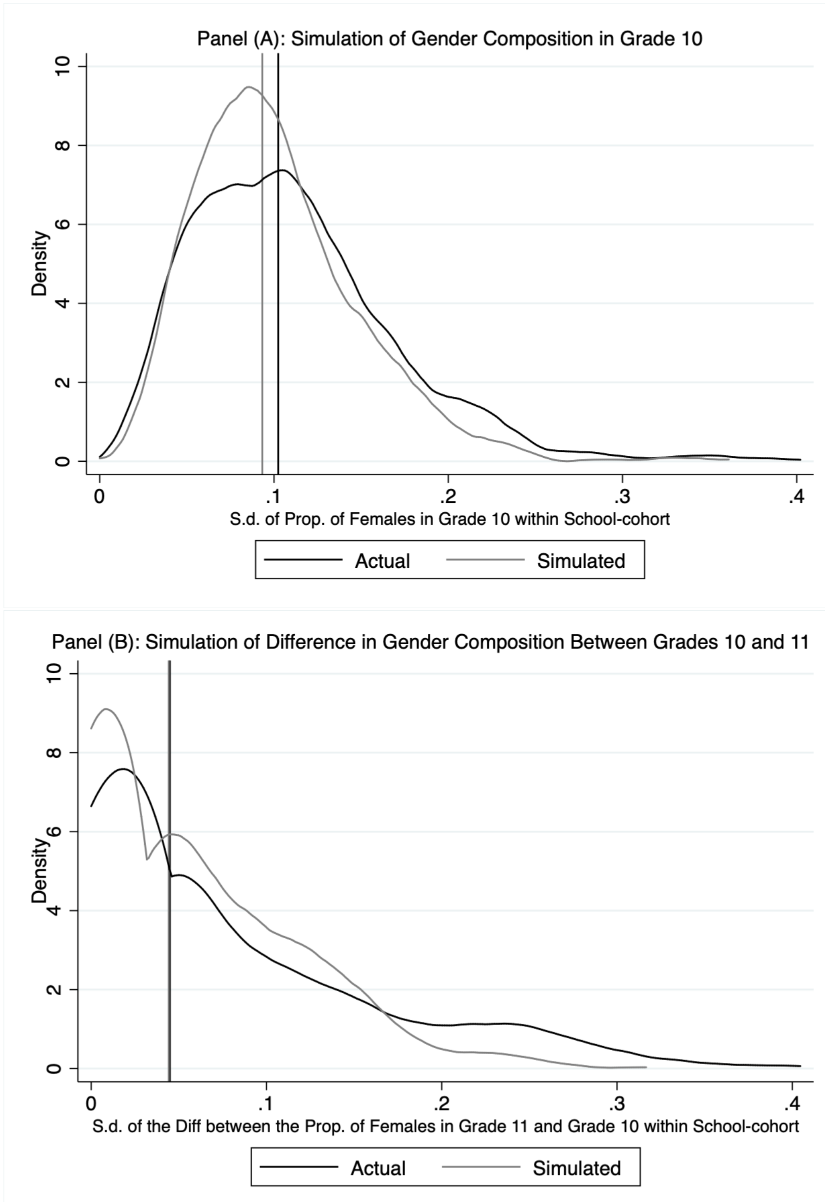
Notes: The figure presents the simulated 95% empirical confidence intervals (shown as vertical bars in gray) of within-school-cohort standard deviations in the proportion of female peers across classrooms. Scatter points (black) represent the actual within school-cohort standard deviations in female peers share across classrooms. To produce the empirical confidence intervals, we randomly generate the gender of students using a binomial distribution with p equal to the actual proportion of females in each school-cohort. We then compute the within-school-cohort standard deviation of the proportion of female peers across classrooms based on the simulated genders. We repeat this process 1,000 times. For each school-cohort, we then obtain the 97.5% percentile of the simulated standard deviation in the proportion of female peers as the upper bound of the confidence interval and the 2.5% percentile of the simulated standard deviation in the proportion of female peers as the lower bound of the confidence interval. To avoid cluttering the paper, we randomly pick 50 school-cohorts (out of 728 school-cohorts). This figure shows that only 5 out of 50 within-school-cohort standard deviations of the proportion of female classmates are not within their corresponding 95% confidence intervals. Simulation results for other school-cohorts are available upon request.

Figure A.4: Histogram of the Proportion of Females in Classrooms in Grade 10



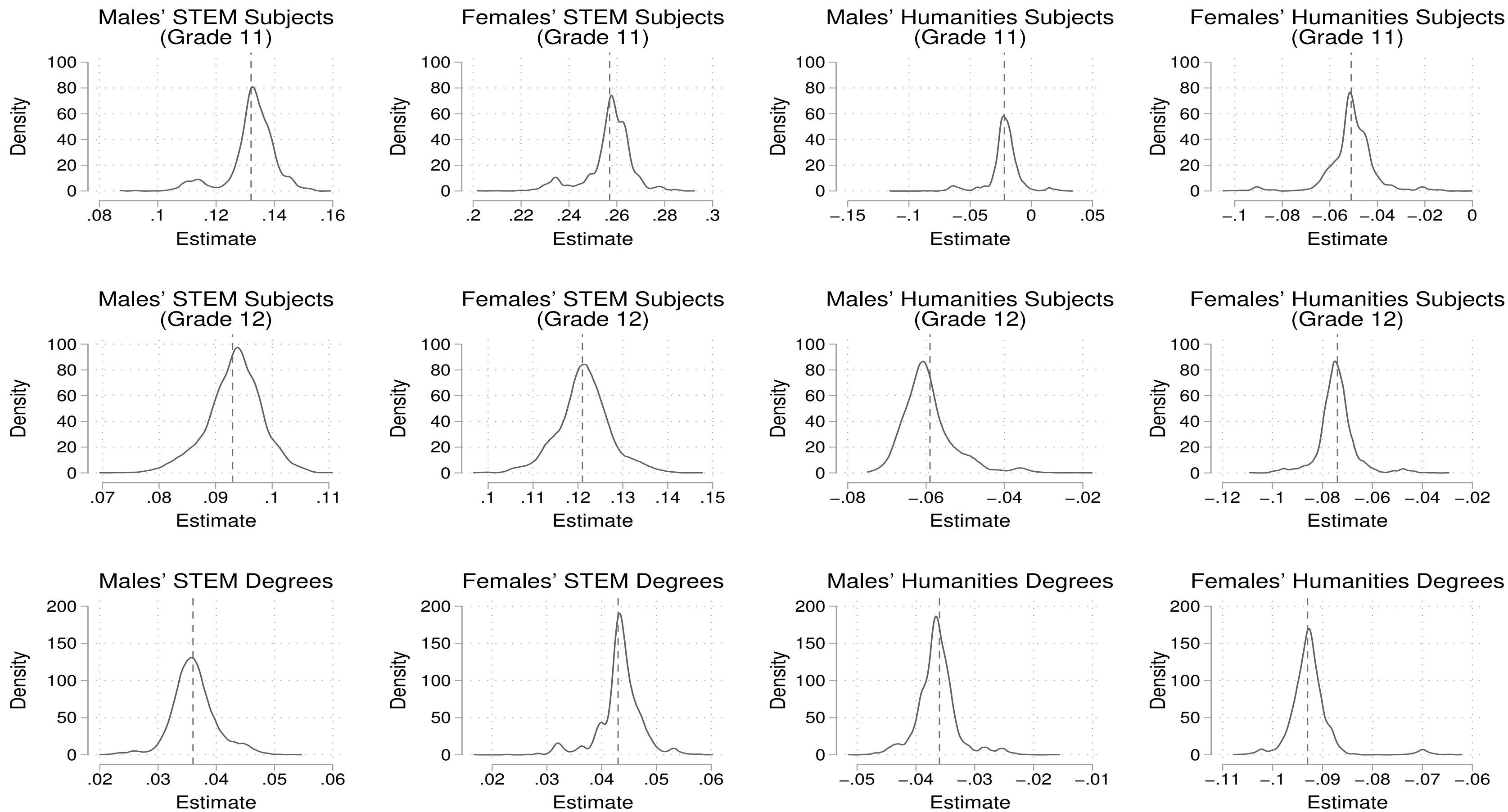
Notes: Panel (A) plots the distribution of the proportion of female classmates and the simulated kernel density of a normal distribution using the sample mean and variation. In Panel (B), the dashed black line represents the kernel density of the residuals from a regression of the share of female classmates on school-cohort fixed effects. The dashed gray line represents a simulated normal distribution with a same mean of 0 and standard deviation of 0.124 and the number of unique classrooms (2,517) of the actual residual distribution. Panel (C) shows the standardized normal probability plot of the residuals obtained by regressing the share of female classmates on school-cohort fixed effects.

Figure A.5: Simulated and Actual Standard Deviations in the Proportion of Females Across Classes within School-Cohorts



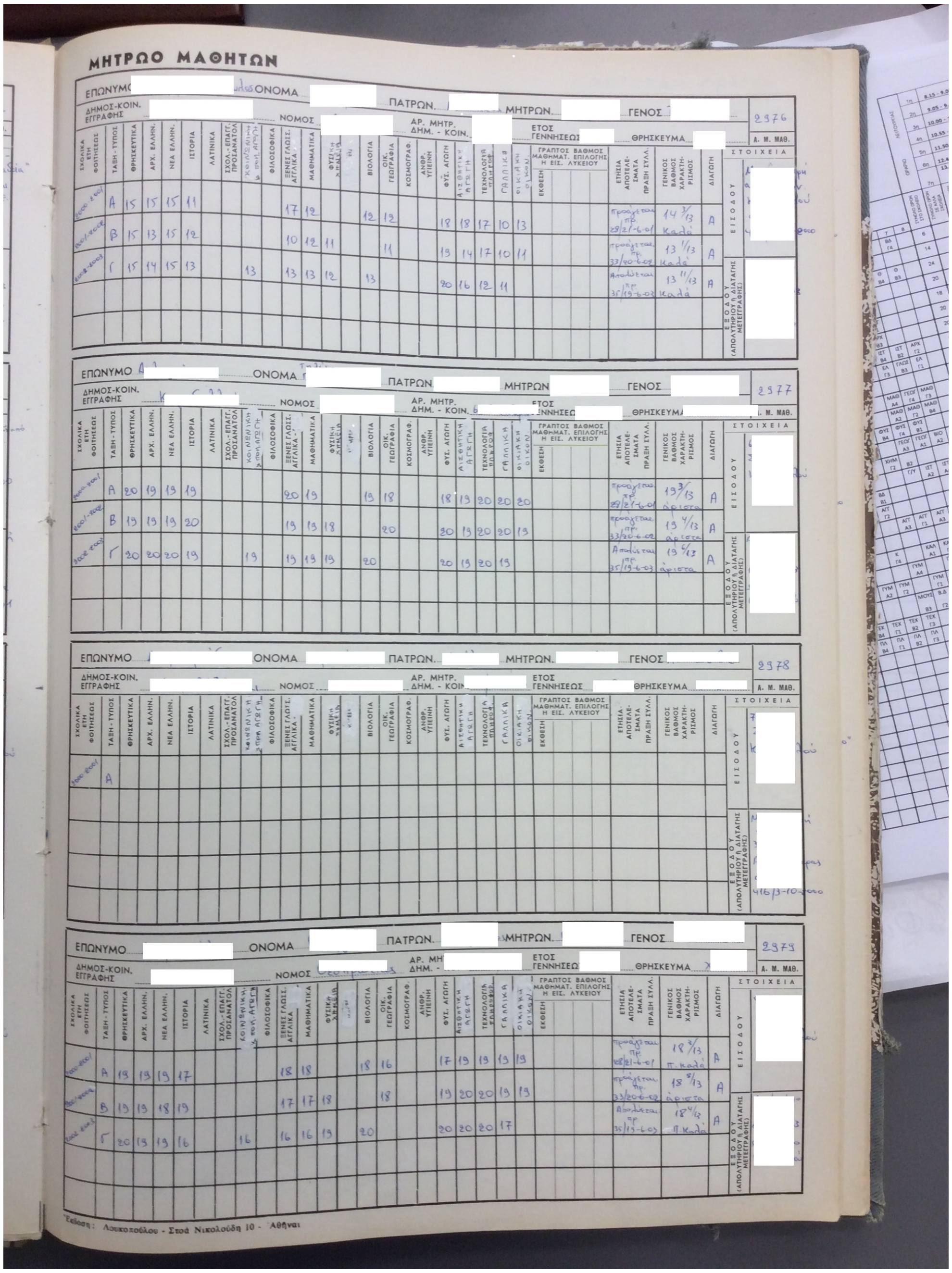
Notes: The figure presents the kernel density of the actual standard deviation in the proportion of females within school-cohort (black line) and the simulated standard deviation of females within school-cohort (dashed line). Vertical lines indicate the median of each distribution. Panel (A) produces the simulated standard deviation for the 10^{th}-grade proportion of female classmates. We randomly generate the gender of students using a binomial distribution with p equal to the actual proportion of females in 10^{th} grade for each school-cohort. We then compute the within-school-cohort standard deviation of the artificially generated proportion female peers across classrooms and plot it with the actual one. Panel (B) produces the simulated standard deviation for the difference between 11^{th} and 10^{th} grade proportion of female classmates. We randomly generate the gender of students using a binomial distribution with p equal to the actual proportion of females in 11^{th} and 10^{th} grade for each school-cohort, respectively, while keeping the same class size. We compute the difference in the proportion of females across grades and then compute the within-school-cohort standard deviation of the difference in the artificially generated proportion of female peers across classrooms and plot it with the actual one.

Figure A.6: Robustness Check: Distribution of Estimates on Scholastic Outcomes and University Degree Choices



Notes: These figures present the distribution of estimated effects of the female peers classroom share in 11^{th} grade. We randomly drop schools from two selected postcode areas out of the total 83 postcode areas each time without replacement. This process is repeated exhaustively, and we conduct a total of $C_{83}^{2} = 3,403$ regressions for each outcome variable. Black lines represent the density of estimates. Vertical dashed lines represent the baseline estimates reproduced from columns (2) and (5) of Table 4.

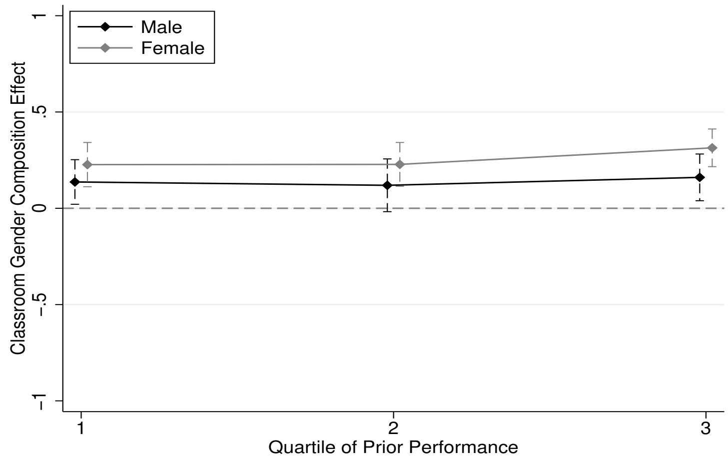
Figure A.7: Student Records from School Archives



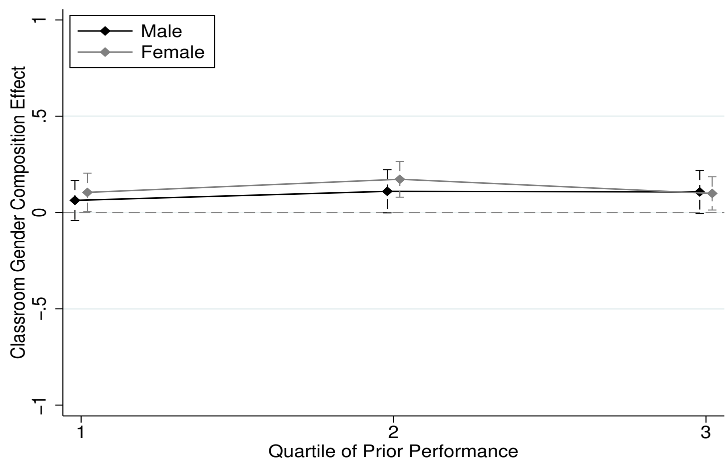
Notes: This figure shows a student record from the school archive of the few schools that kept the data in books for the earlier years in our sample. Each panel corresponds to a different student, with a total of 4 panels appearing vertically on every page. At the top of each panel the following information is provided: student surname, student first name, mother's surname, father's surname, unique school identifier, municipality unique identifier, district unique identifier, year of birth, and religion. Below that, subjects' names are reported. Then, there are three rows of data in each student record; each of the rows corresponds to a different grade (10, 11, and 12). Information about students' classroom identifiers and test scores in each subject, grade, and year are recorded. Raw test scores are out of 20 and are increasing in performance. On the right, there is information on students' GPA in each grade. We have covered all students' personal information to protect the confidentiality of the student and the school.

Figure A.8: Heterogeneous Effects of Female Classmates by Student Baseline Performance

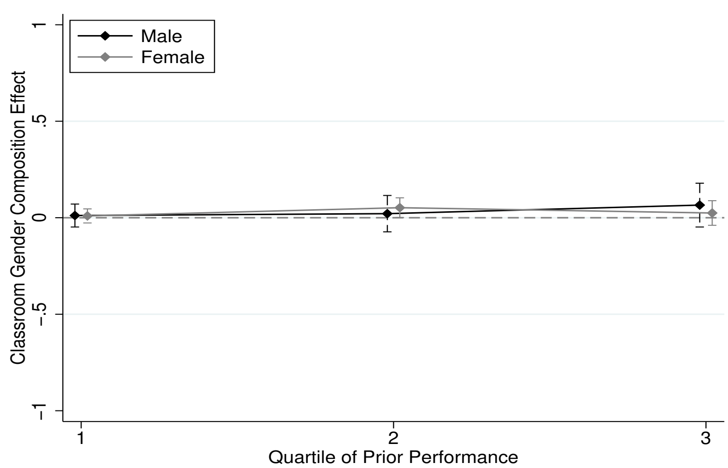
Panel (A). STEM Subjects (Grade 11)



Panel (B). STEM Subjects (Grade 12)



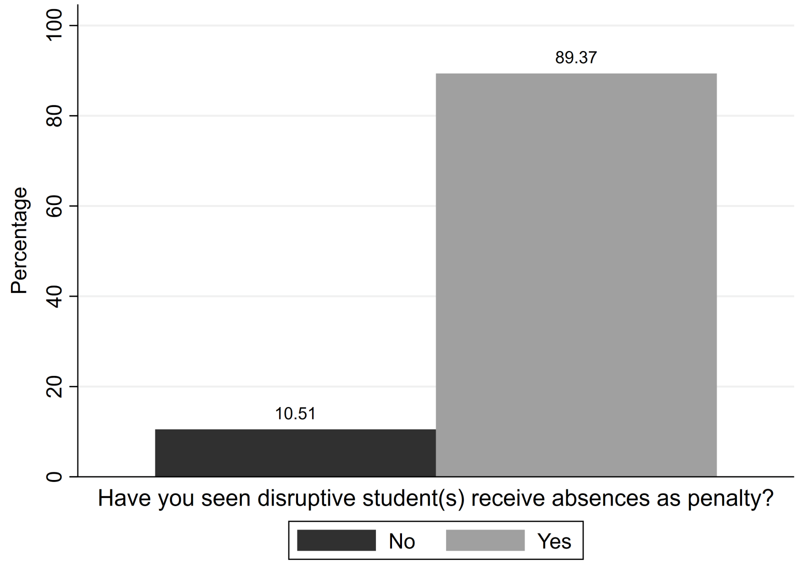
Panel (C). STEM Degrees



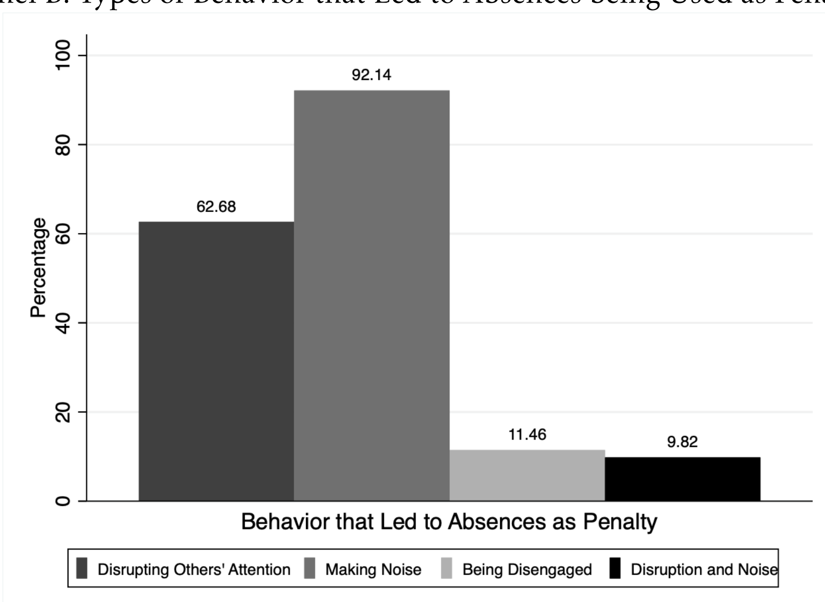
Notes: This figure shows the estimated heterogeneous effects of share of female classmates by student baseline (prior) performance in grade 10. Specifically, we estimate the effects of the female classmates share on males' and for each quartile of students' baseline performance in grade 10. Three outcomes are reported here: 1) Test scores in STEM subjects in grade 11 (Panel A), 2) Test scores in STEM subjects in grade 12 (Panel B), and 3) Enrollment in STEM degrees at the university level (Panel C). The coefficient bars represent the 90% confidence interval.

Figure A.9: Survey Evidence that Unexcused Absences are Used as a Punishment for Disruption and Noise

Panel A: Frequency of Disruptive Students Receiving Absences as Penalty



Panel B: Types of Behavior that Led to Absences being Used as Penalty



Notes: These figures use data from a survey that we conducted in 2022 as part of a field experiment in 6 of the sampled schools. Panel A relies on the following questionnaire item “Have you witnessed hourly unexcused absences as a penalty to disruptive students?". Students can respond to the questionnaire question by “Yes" or “No". 89.37% of students respond that they have seen disruptive students receive absences as a penalty. Panel B uses student responses to the following questionnaire item “In which way can a student in your classroom behave and receive unexcused absences as a penalty". Students have the following options to choose from “Disrupting Others’ Attention", “Making Noise", “Being Disengaged", and “Disruption and Noise". This is a multiple choice question. 92.14% of students report that making noise, 62.68% report that disrupting others’ attention, 11.48% report disengagement, and 9.82% of students report both disruption and noise as the most common reasons for a student to receive an unexcused absence.

B Table Appendices

Table B.1: How Representative is the Sample?

	Sample (104 Schools) /S.D. (1)	Remaining Population (1,024 Schools) /S.D. (2)	(1)-(2) Difference /S.E. (3)
Student Characteristics			
Share of Female Students (%)	0.548	0.562	-0.014
	0.076	0.107	0.011
P-value			0.188
Average Student Age	17.952	17.923	0.029
	0.133	0.185	0.019
P-value			0.124
Share of Students Being Born in 1st Quarter (1=Yes)	0.188	0.201	-0.013
	0.053	0.083	0.008
P-value			0.120
University Admission Score	13566.351	13383.786	182.564
	1065.227	1279.232	131.021
P-value			0.164
Share of Students in Each Track			
Classics	0.371	0.409	-0.037
	0.094	0.167	0.017
P-value			0.027
Science	0.180	0.179	0.001
	0.085	0.138	0.014
P-value			0.956
Exact Science	0.449	0.412	0.036
	0.106	0.190	0.019
P-value			0.057
District Unemployment	9.545	9.826	-0.281
	1.679	3.235	0.325
P-value			0.387

Notes: This table examines the representativeness of schools in our sample. We compare our sample to the remaining public coeducational schools in Greece in terms of students' characteristics (gender, age, being born in 1^{st} quarter of calendar year, university admission score, and high school track choices) at school level and unemployment at the district level. Unemployment is measured at the district level in 2003. Column (1) presents the means of variables in our study sample (104 schools) and column (2) presents the means of variables in the remaining public coeducational population of schools in Greece (containing 1,024 schools). Column (3) presents the differences between sample and population means, the standard error of the difference, and p-values. The comparisons are made using data from the first year for which the dataset is available.

Table B.2: Descriptive Statistics by Cohort

Cohort	No. of Schools	Mean No. of Classrooms	Mean School Size	Mean Classroom Size	School-level Prop. Females		Classroom-level Prop. Females	
					Mean	SD	Mean	SD
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
2002	84	3.583	65.202	18.757	0.557	0.078	0.554	0.162
2003	87	3.414	66.276	19.612	0.553	0.087	0.551	0.148
2004	90	3.544	69.089	19.553	0.557	0.073	0.554	0.139
2005	95	3.600	63.253	17.674	0.566	0.084	0.566	0.139
2006	91	3.440	51.527	15.475	0.577	0.074	0.576	0.152
2007	93	3.484	47.828	14.584	0.580	0.095	0.573	0.159
2008	93	3.516	48.032	14.410	0.593	0.099	0.585	0.150
2009	95	3.768	67.126	17.963	0.554	0.076	0.554	0.132

Notes: The table shows the summary statistics of the number of schools, number of classrooms, school size, classroom size, and proportion of females in school and classroom by each cohort.

Table B.3: Number of Schools by Cohort with Different Number of Classrooms

	Number of Schools with					
Cohort	2 classes	3 classes	4 classes	5 classes	6 classes	7 classes
(1)	(2)	(3)	(4)	(5)	(6)	(7)
2002	16	25	28	9	5	1
2003	19	30	24	11	3	0
2004	21	19	36	9	4	1
2005	15	29	34	14	2	1
2006	19	28	32	9	3	0
2007	19	30	28	12	4	0
2008	17	31	30	11	3	1
2009	15	24	31	18	7	0

Notes: The table shows the number of schools with different numbers of classrooms by cohort.

Table B.4: Variance Decomposition

Panel A	Variation of Female Classmate Share in Grade 11		
	Sum of Squares	Share of Total	Degrees of Freedom
	(1)	(2)	(3)
Within School-cohorts	39.512	0.720	1,789
Between School-cohorts	15.401	0.280	727
Total	54.912		2,516
Panel B	**Variation of Female Classmate Share in Grade 10**		
Within School-cohorts	38.761	0.705	1,807
Between School-cohorts	16.207	0.295	727
Total	54.968		2,534
Panel C	**Variation of Change in Female Classmate Share from Grade 10 to Grade 11**		
Within School-cohorts	27.332	0.872	1,789
Between School-cohorts	3.999	0.128	727
Total	31.331		2,516

Notes: The table presents the decomposition of variance in the proportion of females in 11th grade (Panel A), the decomposition of variance in the proportion of females in 10th grade (Panel B), and the decomposition of changes in the proportion of females from grade 10 to grade 11 (Panel C) into within-school-cohort variation and between-school-cohorts variation. Columns (1)-(3) present the sum of squares, share of total variation, and degrees of freedom.

Table B.5: Random Assignment of Students to Classrooms

Dependent Variables:	Class Average Gender	Class Average Age	Class Average Born in 1^{st} Quarter	Female Share in Dropouts
	(1)	(2)	(3)	(4)
Classroom Number=1	0.025	-0.002	-0.002	0.031**
	(0.016)	(0.006)	(0.003)	(0.013)
Classroom Number=2	-0.000	0.006	-0.005	0.018
	(0.018)	(0.005)	(0.004)	(0.011)
Classroom Number=3	0.008	0.004	-0.004	0.016*
	(0.010)	(0.009)	(0.004)	(0.007)
Classroom Number=5	-0.059*	0.003	-0.001	-0.035
	(0.026)	(0.013)	(0.011)	(0.019)
Classroom Number=6	-0.048	0.024	-0.017	-0.045
	(0.031)	(0.022)	(0.015)	(0.026)
Classroom Number=7	0.007	0.069	-0.077	0.009
	(0.048)	(0.047)	(0.042)	(0.031)
Mean of Dep.	0.564	16.909	0.136	-0.001
F statistic	2.811	0.937	0.825	2.153
P-value of F-stat.	0.101	0.511	0.549	0.169
School-cohort FE	✓	✓	✓	✓

Notes: We regress each classroom-level mean variable in 11^{th} grade and female share of dropouts from 10^{th} to 11^{th} grade (listed in the top row of the table) on classroom number indicators. The omitted category is classroom number 4. Each cell presents the estimates for the listed classroom number indicators from the regression. All regressions control for school-cohort fixed effects. We report the p-value of the joint F-test for testing the null hypothesis that all the coefficients of listed independent variables are equal to null. Standard errors clustered at school-cohort level are reported in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table B.6: Balancing Tests of the Proportion of Female Classmates, By Gender

Dependent Variables:	Prop. Female Peers (Grade 11)
	Panel A: Males
Age (Grade 11)	-0.0004
	(0.0018)
Being Born in 1^{st} Quarter of Calendar Year (1=Yes)	-0.0044
	(0.0048)
GPA (Grade 10)	0.0002
	(0.0004)
N	19,114
p-value for Joint F Test for Individual Characteristics	0.6077
School-cohort FE	✓
	Panel B: Females
Age (Grade 11)	0.0023
	(0.0016)
Being Born in 1^{st} Quarter of Calendar Year (1=Yes)	-0.0016
	(0.0037)
GPA (Grade 10)	0.0002
	(0.0003)
N	24,337
p-value for Joint F Test for Individual Characteristics	0.4455
School-cohort FE	✓

Notes: The dependent variable is the share of females in the classroom in grade 11. Each cell reports the estimated effect of the related student characteristic (reported vertically) from a separate regression in which the dependent variable is the share of female classmates in grade 11. We run these regressions for males and females, separately. Panel A reports the estimated effects for males, and Panel B shows the estimated effects for females. We also show the F-statistic of the joint significance of student characteristics from the regression of the dependent variable on all student characteristics, conditional on school-cohort fixed effects for males (Panel A) and females (Panel B). Standard errors clustered at classroom level are reported in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table B.7: Non-systematic Transfer and Change in Proportion of Females

Dependent Variables:	Change in Prop. Female Peers	Indicator Classroom Transfer (1=Yes)	Prop. Female Peers (Grade 10)
	(1)	(2)	(3)
Female (1=Yes)	0.0011	0.0003	-0.0004
	(0.0020)	(0.0034)	(0.0018)
Age (Grade 11)	0.0008	0.0004	0.0005
	(0.0011)	(0.0033)	(0.0013)
Born in 1^{st} Quarter of Birth Year (1=Yes)	-0.0005	-0.0007	-0.0010
	(0.0015)	(0.0042)	(0.0017)
GPA (Grade 10)	0.0001	-0.0010	0.0001
	(0.0003)	(0.0007)	(0.0002)
N	43,451	43,451	43,451
Joint F Test for Individual Characteristics	0.373	0.971	0.139
School-cohort FE	✓	✓	✓

Notes: The dependent variables in column 1-3 are the change in the proportion of female classmates from grade 10 to grade 11, indicator for classroom transfer, and the proportion of female classmates in grade 10. Each cell reports the estimate on a student characteristic from a separate regression in which the dependent variable is the one listed in the corresponding column. The lower panel shows the F-statistic of joint significance of student characteristics from the regression of the dependent variable on all students' characteristics, conditional on school-cohort fixed effects. Standard errors clustered at classroom level are reported in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table B.8: Summary of Monte Carlo Simulation for the Proportion of Females in Grade 10 and the Difference in the Proportion of Females between Grade 10 and Grade 11

Proportion of Female Classmates in Grade 10	
Share of the Actual Standard Deviation within 90% Empirical CI	0.843
Share of the Actual Standard Deviation within 95% Empirical CI	0.906
Share of the Actual Standard Deviation within 99% Empirical CI	0.948
Difference between the Proportion of Females in Grade 10 and Grade 11	
Share of the Actual Standard Deviation within 90% Empirical CI	0.843
Share of the Actual Standard Deviation within 95% Empirical CI	0.868
Share of the Actual Standard Deviation within 99% Empirical CI	0.913

Notes: This table summarizes the share of the actual standard deviation of proportion of females in 10^{th} grade and the share of the actual standard deviation of the difference between the proportions of females in 11^{th} grade and 10^{th} grade, which falls within the 90%, 95% and 99% of the empirical confidence interval generated by the simulated standard deviation of gender composition for each school-cohort. To produce the empirical confidence intervals, we repeat the process of simulating the standard deviations described in Figure A.5 1,000 times. For each school-cohort, we then obtain the 95.0% (97.5% or 99.5%) and 5.0% (2.5% or 0.5%) percentile of the simulated standard deviation in the proportion of females as the upper and lower bounds of the 90% (95% or 99%) empirical confidence interval.

Table B.9: Estimates of the Share of Female Classmates in Grade 10 on End of Grade 10 Test Scores

	Males	Females
	(1)	(2)
10th-grade Performance in		
STEM Subjects	-0.059	0.163
	(0.062)	(0.058)***
N	19,215	24,434
Humanities Subjects	-0.085	0.036
	(0.060)	(0.057)
N	19,215	24,434
School-cohort FE & Student-level controls	✓	✓
Classroom-level controls	✓	✓

Notes: The table reports the estimated effects of the share of female classmates in 10th grade on students' scholastic outcomes in the school exams at the end 10thgrade. Student-level controls include age, and born in the 1stst quarter of birth year indicator. Classroom-level controls include classroom-level mean age, born in 1st quarter of birth year indicator, and classroom size. Regressions in all columns include school-cohort fixed effects and classroom-level controls. STEM subjects in 10th grade include Mathematics and Physics. Humanities subjects in 10th grade include Language and History. Standard errors clustered at classroom level are reported in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table B.10: Impact of Classroom Gender Composition on STEM Degree Choice (Alternative Definition of STEM)

	Males		Females	
	(1)	(2)	(3)	(4)
Dependent Variables:				
STEM Degrees (Alternative)	0.046	0.045	0.068***	0.064***
	(0.029)	(0.029)	(0.023)	(0.023)
Mean of Y	-0.015	-0.015	0.012	0.012
N	19,113	19,113	24,337	24,337
School-cohort FE & Student-level controls	✓	✓	✓	✓
Classroom-level controls		✓		✓

Notes: The alternative definition of STEM degrees additionally includes Economics and Health Sciences degrees. Clearly, the pattern of results consistently follows the baseline pattern, which shows that our results are robust to the change in definition of STEM. Standard errors clustered at classroom level are reported in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table B.11: Impact of Classroom Gender Composition on Test Scores by Subject

	Males	Females
	(1)	(2)
Dependent Variables:		
Panel A: 11th-grade STEM Subjects		
Mathematics	0.117**	0.189***
	(0.059)	(0.054)
Mean of Y	-0.018	0.014
N	19,112	24,336
Physics	0.128**	0.291***
	(0.057)	(0.054)
Mean of Y	-0.010	0.008
N	19,112	24,336
Panel B: 11th-grade Humanities Subjects		
Language	0.057	-0.064
	(0.066)	(0.061)
Mean of Y	-0.247	0.194
N	19,112	24,336
History	-0.079	-0.033
	(0.063)	(0.054)
Mean of Y	-0.159	0.125
N	19,112	24,336
Panel C: 12th-grade STEM Subjects		
Mathematics	0.116**	0.145***
	(0.046)	(0.038)
Mean of Y	0.112	-0.088
N	19,113	24,337
Physics	0.051	0.056
	(0.048)	(0.040)
Mean of Y	0.025	-0.020
N	19,113	24,337
Panel D: 12th-grade Humanities Subjects		
Language	0.030	-0.079**
	(0.047)	(0.039)
Mean of Y	-0.254	0.200
N	19,113	24,337
History	-0.097**	-0.056
	(0.048)	(0.044)
Mean of Y	-0.167	0.131
N	19,113	24,337
School-cohort FE & Student-level controls	✓	✓
Classroom-level controls	✓	✓

Notes: We decompose the dependent variables of STEM subjects and humanities subjects in the main results of Table 4 into their components. Each cell reports the estimates on female peers classroom share in 11th grade. Student-level controls include age, born in 1st quarter of birth year indicator, concurrent track choice, previous GPA, and the share of female classmates in grade 10. Classroom-level controls include classroom-level mean age, born in 1st quarter of birth year indicator, and classroom size. Standard errors clustered at classroom level are reported in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table B.12: Dispersion of Female Class Share by Cohort

	Number of Schools with Female Share of Classmates in					
Cohort	1 Quantile	2 Quantiles	3 Quantiles	4 Quantiles	Total N. of Schools	Share of Schools with Female Share in at Least 2 Quantiles
(1)	(2)	(3)	(4)	(5)	(6)	(7)
2002	2	16	43	23	84	.976
2003	3	21	43	20	87	.966
2004	4	22	39	25	90	.956
2005	2	21	53	19	95	.979
2006	1	25	39	26	91	.989
2007	5	24	36	28	93	.946
2008	8	18	38	29	93	.914
2009	2	21	43	29	95	.979

Notes: The table shows the number of schools with female classmate shares at different quantiles of the distribution by cohort.

Table B.13: Estimated Effects of Peer Gender Composition on Student Outcomes using a Within-subject Estimation

Dependent Variable:	STEM Subjects (Grade 11)		Humanities Subjects (Grade 11)		STEM Subjects (Grade 12)		Humanities Subjects (Grade 12)	
	Male	Female	Male	Female	Male	Female	Male	Female
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Prop. Female Peers	0.122***	0.240***	-0.011	-0.049*	0.083***	0.100***	-0.033	-0.068**
	(0.032)	(0.026)	(0.034)	(0.027)	(0.031)	(0.025)	(0.033)	(0.026)
N	38,224	48,672	38,224	48,672	38,226	48,674	38,226	48,674
School-subject-cohort FE & Student-level controls	✓	✓	✓	✓	✓	✓	✓	✓
Class-level controls	✓	✓	✓	✓	✓	✓	✓	✓

Notes: The table shows the effects from within STEM subjects-by-school-cohort estimation (columns 1-2 and columns 5-6) and within Humanities subjects-by-school-cohort estimation (columns 3-4 and columns 7-8). To control for the unobserved school-cohort-by-STEM subjects shock or the unobserved school-cohort-by-humanities subjects shock, we stack observations across STEM subjects or humanities subjects and additionally include STEM subject-school-cohort or humanities subject-school-cohort fixed effects. Standard errors clustered at classroom level are reported in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table B.14: Robustness Check: Spillovers from Female Peers in Other Classrooms and from Female Peers' Ability in Own Classroom

	Baseline		Robustness Checks Control for Female Share Spillover		Robustness Checks Control for Females' Ability Spillover	
	Males	Females	Males	Females	Males	Females
	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variables:						
Panel A: 11th-grade Performance						
STEM Subjects	0.132***	0.257***	0.085	0.230***	0.173***	0.272***
	(0.049)	(0.046)	(0.092)	(0.081)	(0.049)	(0.046)
N	19,112	24,336	19,112	24,336	19,046	24,215
Humanities Subjects	-0.022	-0.051	-0.072	-0.060	-0.002	-0.037
	(0.056)	(0.049)	(0.102)	(0.086)	(0.055)	(0.049)
N	19,112	24,336	19,112	24,336	19,046	24,215
Panel B: 12th-grade Performance						
STEM Subjects	0.093**	0.121***	0.149**	0.184***	0.094**	0.122***
	(0.042)	(0.034)	(0.075)	(0.065)	(0.042)	(0.034)
N	19,113	24,337	19,113	24,337	19,047	24,216
Humanities Subjects	-0.059	-0.074**	0.023	-0.072	-0.056	-0.068*
	(0.041)	(0.036)	(0.079)	(0.065)	(0.042)	(0.036)
N	19,113	24,337	19,113	24,337	19,047	24,216
Panel C: University Degree Choice						
STEM Subjects	0.036	0.043**	0.014	0.001	0.033	0.042**
	(0.028)	(0.019)	(0.050)	(0.033)	(0.028)	(0.019)
N	19,113	24,337	19,113	24,337	19,047	24,216
Humanities Subjects	-0.036**	-0.093***	-0.032	-0.132***	-0.042**	-0.092***
	(0.018)	(0.021)	(0.031)	(0.037)	(0.018)	(0.021)
N	19,113	24,337	19,113	24,337	19,047	24,216
School-cohort FE & Student-level controls	✓	✓	✓	✓	✓	✓
Classroom-level Controls	✓	✓	✓	✓	✓	✓
Prop. Females in Other Classrooms			✓	✓		
Female Classmates' Prior GPA					✓	✓

Notes: Columns (1) and (2) reproduce the baseline results from columns (2) and (5) of Table 4. Results for males (females) from columns (3) and (5) (columns 4 and 6) are the estimated effects of own classroom female peers share and average female peers share of other classrooms, which are obtained from the same regression in which both own class female peers share and average female peers share in other classrooms within the same school and cohort are included. Student-level controls include age, born in 1st quarter of birth year indicator, prior track choice, previous GPA, and the share of female classmates in grade 10. Classroom-level controls include classroom-level mean age, born in 1st quarter of birth year indicator, and classroom size. Standard errors clustered at classroom level are reported in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table B.15: Nonlinear Effects of Female Classmates in High School

	Males		Females	
	Est.	SE.	Est.	SE.
	(1)	(2)	(3)	(4)
Panel A: 11^{th}-grade Performance				
STEM Subjects				
Prop.Females in Quantile 2	0.019	(0.018)	0.031*	(0.016)
Prop.Females in Quantile 3	0.017	(0.019)	0.047***	(0.015)
Prop.Females in Quantile 4	0.040**	(0.019)	0.080***	(0.017)
Humanities Subjects				
Prop.Females in Quantile 2	-0.014	(0.021)	-0.036**	(0.018)
Prop.Females in Quantile 3	-0.059***	(0.022)	0.007	(0.019)
Prop.Females in Quantile 4	0.003	(0.021)	0.000	(0.019)
Panel B: 12^{th}-grade Performance				
STEM Subjects				
Prop.Females in Quantile 2	0.018	(0.013)	-0.010	(0.012)
Prop.Females in Quantile 3	0.009	(0.014)	0.003	(0.012)
Prop.Females in Quantile 4	0.043***	(0.016)	0.033**	(0.014)
Humanities Subjects				
Prop.Females in Quantile 2	-0.045***	(0.017)	-0.010	(0.015)
Prop.Females in Quantile 3	-0.007	(0.017)	-0.014	(0.014)
Prop.Females in Quantile 4	-0.019	(0.017)	-0.031**	(0.015)
Panel C: University Degree Choice				
STEM Degrees				
Prop.Females in Quantile 2	0.012	(0.010)	0.002	(0.006)
Prop.Females in Quantile 3	-0.011	(0.010)	0.011	(0.007)
Prop.Females in Quantile 4	0.013	(0.011)	0.012	(0.008)
Humanities Degrees				
Prop.Females in Quantile 2	-0.002	(0.005)	-0.013*	(0.007)
Prop.Females in Quantile 3	-0.005	(0.006)	-0.018**	(0.007)
Prop.Females in Quantile 4	-0.017**	(0.007)	-0.029***	(0.008)
School-cohort FE & Student-level controls	✓	✓	✓	✓
Classroom-level controls	✓	✓	✓	✓

Notes: The table reports nonlinear effects of the share of female classmates in 11^{th} grade on males' and females' outcomes. The model replaces the single treatment variable with a set of quantile indicators for the different quantiles of the proportion of female peers in the classroom. The omitted category is the first quantile (average share of females is 38% for males and 39% for females), which is the quantile with the lowest share of female classmates. For male students the range for the proportion of female classmates is 47%-56% and the mean is 52% for Quantile 2, 56%-65% and 60% for Quantile 3, and 65%-95% and 73% for Quantile 4. For female students the range for the proportion of female classmates is 47%-56% and the mean is 52% for Quantile 2, 56%-65% and 61% for Quantile 3, and 65-95% and 74% for Quantile 4. Estimates in each row by gender are generated from the same regression. Student-level controls include age, born in 1^{st} quarter of birth year indicator, prior track choice, previous GPA, and the share of female classmates in grade 10. Classroom-level controls include classroom-level mean age, born in 1^{st} quarter of birth year indicator, and classroom size. Standard errors clustered at classroom level are reported in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table B.16: Randomness in Sample Attrition

Dependent Variables:	Excused Absence (Grade 11)	Unexcused Absence (Grade 11)	Excused Absence (Grade 12)	Unexcused Absence (Grade 12)
	(1)	(2)	(3)	(4)
Prop. Female Peers	0.003	0.005	0.004	0.001
	(0.006)	(0.006)	(0.006)	(0.005)
N	43,451	43,451	43,451	43,451
Mean of Y	0.244	0.241	0.244	0.238
School-cohort FE	✓	✓	✓	✓

Notes: Each dependent variable listed in the top row is an indicator that equals one if the corresponding measurement is missing. All regressions control for school-cohort fixed effects. Standard errors clustered at classroom level are reported in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table B.17: Heterogeneous Gender Composition Effects on Absenteeism by Neighborhood Income

Panel A - Dependent Variables (Grade 11)

	Total Absences				Excused Absences				Unexcused Absences			
	Below Median		Above Median		Below Median		Above Median		Below Median		Above Median	
	Males	Females	Males	Females	Males	Females	Males	Females	Males	Females	Males	Females
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Prop. Female Peers	-2.989	-4.763	2.138	-3.668	1.285	-3.535*	1.458	-1.663	-4.286***	-1.169	0.709	-1.986
	(3.152)	(2.906)	(3.232)	(3.233)	(2.178)	(2.145)	(2.205)	(2.223)	(1.644)	(1.404)	(1.830)	(1.626)
N	5,835	7,803	8,566	10,640	5,835	7,803	8,566	10,640	5,872	7,861	8,582	10,653
Y Mean	47.147	48.595	49.307	50.615	17.219	20.025	18.686	21.430	29.906	28.523	30.590	29.162
School-cohort FE & Student-level controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Class-level controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Panel B - Dependent Variables (Grade 12)

	Total Absences				Excused Absences				Unexcused Absences			
	Below Median		Above Median		Below Median		Above Median		Below Median		Above Median	
	Males	Females	Males	Females	Males	Females	Males	Females	Males	Females	Males	Females
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Prop. Female Peers	-3.534	-3.141	1.136	-2.555	0.706	-1.077	1.990	-1.669	-4.203***	-2.202	-1.323	-0.655
	(2.630)	(2.509)	(3.380)	(3.167)	(2.055)	(2.122)	(2.750)	(2.640)	(1.366)	(1.411)	(1.639)	(1.525)
N	5,940	7,899	8,466	10,524	5,941	7,900	8,466	10,528	5,994	7,986	8,532	10,593
Y Mean	72.958	76.503	71.196	75.816	37.352	42.086	36.799	42.709	35.560	34.329	34.301	33.024
School-cohort FE & Student-level controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Class-level controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Notes: The table reports heterogeneous effects by neighborhood income of the female peers classroom share in 11^{th} grade on students' attendance pattern at the end of grade 11 (Panel A) and grade 12 (Panel B). Student-level controls include age, born in 1^{st} quarter of birth year indicator, prior track choice, previous GPA, and the share of female classmates in grade 10. Classroom-level controls include classroom-level mean age, share of classmates who are born in 1^{st} quarter of birth year, and classroom size. Standard errors clustered at classroom level are reported in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table B.18: Heterogeneous Gender Composition Effect on Absenteeism by School Quality

Panel A - Dependent Variables (Grade 11)

	Total Absences				Excused Absences				Unexcused Absences			
	Low-quality		High-quality		Low-quality		High-quality		Low-quality		High-quality	
	Males	Females	Males	Females	Males	Females	Males	Females	Males	Females	Males	Females
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Prop. Female Peers	-1.922	-5.989**	1.591	-2.541	0.876	-3.182	2.022	-2.373	-2.826*	-2.739*	-0.394	-0.151
	(3.119)	(3.039)	(3.367)	(3.068)	(2.140)	(2.172)	(2.356)	(2.244)	(1.699)	(1.518)	(1.761)	(1.450)
N	4,930	6,707	9,471	11,736	4,930	6,707	9,471	11,736	4,967	6,765	9,487	11,749
Y Mean	46.292	48.128	50.680	51.739	16.757	19.818	19.493	22.069	29.519	28.276	31.148	29.640
School-cohort FE & Student-level controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Class-level controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Panel B - Dependent Variables (Grade 12)

	Total Absences				Excused Absences				Unexcused Absences			
	Low-quality		High-quality		Low-quality		High-quality		Low-quality		High-quality	
	Males	Females	Males	Females	Males	Females	Males	Females	Males	Females	Males	Females
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Prop. Female Peers	-3.024	-0.938	-0.603	-3.562	0.854	-0.313	-0.164	0.732	-4.150***	-0.860	-0.424	-4.280***
	(2.840)	(2.315)	(2.283)	(2.168)	(2.201)	(1.838)	(1.943)	(1.888)	(1.405)	(1.254)	(1.241)	(1.200)
N	5,078	6,925	9,328	11,498	5,079	6,930	9,328	11,498	5,150	7,010	9,376	11,569
Y Mean	68.559	71.681	73.753	78.779	33.954	38.693	38.700	44.702	34.466	32.866	35.015	34.020
School-cohort FE & Student-level controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Class-level controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Notes: The table reports heterogeneous effects by school quality of the female peers classroom share in 11th grade on students' attendance pattern at the end of grade 11 (Panel A) and grade 12 (Panel B). Student-level controls include age, born in 1st quarter of birth year indicator, prior track choice, previous GPA, and the share of female classmates in grade 10. Classroom-level controls include classroom-level mean age, share of classmates who are born in 1st quarter of birth year, and classroom size. Standard errors clustered at classroom level are reported in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table B.19: Impact of Classroom Gender Composition on Test Scores and Degree Choice when Controlling for Attendance

Panel A	**Males**			**Females**		
Dependent Variables: (Grade 11)	STEM Subjects (Full Sample)	STEM Subjects (Reduced Sample)	STEM Subjects (Reduced Sample)	STEM Subjects (Full Sample)	STEM Subjects (Reduced Sample)	STEM Subjects (Reduced Sample)
	(1)	(2)	(3)	(4)	(5)	(6)
Prop. Female Peers	0.132***	0.091	0.085	0.257***	0.257***	0.249***
	(0.049)	(0.059)	(0.059)	(0.046)	(0.055)	(0.054)
N	19,112	14,400	14,400	24,336	18,442	18,442
Mean of Y	-0.015	-0.014	-0.014	0.012	0.006	0.006
Panel B	**Males**			**Females**		
Dependent Variables: (Grade 12)	STEM Subjects (Full Sample)	STEM Subjects (Reduced Sample)	STEM Subjects (Reduced Sample)	STEM Subjects (Full Sample)	STEM Subjects (Reduced Sample)	STEM Subjects (Reduced Sample)
	(1)	(2)	(3)	(4)	(5)	(6)
Prop. Female Peers	0.093**	0.094*	0.084*	0.121***	0.128***	0.123***
	(0.042)	(0.049)	(0.049)	(0.034)	(0.042)	(0.042)
N	19,113	14,406	14,406	24,337	18,423	18,423
Mean of Y	0.078	0.067	-0.031	-0.062	-0.080	-0.011
Panel C	**Males**			**Females**		
Dependent Variables: (Major Choice)	STEM Degrees (Full Sample)	STEM Degrees (Reduced Sample)	STEM Degrees (Reduced Sample)	STEM Degrees (Full Sample)	STEM Degrees (Reduced Sample)	STEM Degrees (Reduced Sample)
	(1)	(2)	(3)	(4)	(5)	(6)
Prop. Female Peers	0.036	0.042	0.043	0.043**	0.055**	0.056**
	(0.028)	(0.034)	(0.034)	(0.019)	(0.022)	(0.022)
N	19,113	14,406	14,406	24,337	18,423	18,423
Mean of Y	0.352	0.351	0.351	0.133	0.131	0.131
School-cohort FE & Student-level controls	✓	✓	✓	✓	✓	✓
Class-level controls	✓	✓	✓	✓	✓	✓
Absenteeism Channel			✓			✓

Notes: The table reports the estimated effects of the share of female classmates in grade 11 on students' academic performance in 11th grade (Panel A), 12th grade (Panel B), and students' decision to study STEM at the university level (Panel C) under different specifications. Columns (1) and (4) reproduce the baseline estimates for males and females. Column (2) reports the estimated effect from a smaller sample for which we have data on student absenteeism. Columns (3) and (4) report the estimated effects from a specification that includes controls for student excused and unexcused absenteeism in 11th grade (Panel A), or 12th grade (Panels B and C). Student-level controls include age, born in 1st quarter of birth year indicator, prior track choice, prior GPA, and female peers classroom share in 10th grade. Classroom-level controls include classroom-level mean age, born in 1st quarter of birth year indicator, and classroom size. Standard errors clustered at classroom level are reported in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

C Appendix

We closely follow the notation and methodology of Gelbach (2016). Let $\boldsymbol{X_1} = [\boldsymbol{1}\ PropFemalePeers\ \boldsymbol{X}\ \boldsymbol{W}]$, which combines all regressors in main specification (1), and let $\boldsymbol{X_2} = [m^1\ m^2]$ be the measured mechanisms covariates, where m^1 denotes excused absences and m^2 denotes unexcused absences. We rewrite equation (1) in the form of a base model (C.1), where $\boldsymbol{\beta_1^{Base}} = [\lambda\ \beta\ \gamma\ \delta]$ nests all model parameters in equation (1). We rewrite equation (4) in the representation of equation (C.2) and use it as a full model in our analysis, where $\boldsymbol{\beta_1^{Full}} = [\underline{\lambda}\ \underline{\beta}\ \underline{\gamma}\ \underline{\delta}]$, and $\boldsymbol{\beta_2^{Full}} = [\zeta^1\ \zeta^2]$.

Base model:

$$\boldsymbol{Y} = \boldsymbol{X}_1\boldsymbol{\beta}_1^{\boldsymbol{Base}} + \boldsymbol{\epsilon}^{\boldsymbol{Base}} \tag{C.1}$$

Full model:

$$\boldsymbol{Y} = \boldsymbol{X}_1\boldsymbol{\beta}_1^{\boldsymbol{Full}} + \boldsymbol{X}_2\boldsymbol{\beta}_2^{\boldsymbol{Full}} + \boldsymbol{\epsilon}^{\boldsymbol{Full}} \tag{C.2}$$

Under the ideal assumption that $\boldsymbol{E}[\boldsymbol{\epsilon^{Base}}|\boldsymbol{X_1}] = \boldsymbol{E}[\boldsymbol{\epsilon^{Full}}|\boldsymbol{X_1}, \boldsymbol{X_2}] = \boldsymbol{0}$, both $\boldsymbol{\beta_1^{Base}}$ and $\boldsymbol{\beta_1^{Full}}$ are consistently estimated.[46] Now consider the coefficients on $\boldsymbol{X_1}$ from the base model that ignores $\boldsymbol{X_2}$. Derivation of OLS estimator for $\boldsymbol{\beta_1^{Base}}$ is as follows:

$$\begin{aligned}
\hat{\boldsymbol{\beta}}_1^{\boldsymbol{Base}} &= (\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\boldsymbol{Y} \\
&= (\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'(\boldsymbol{X}_1\boldsymbol{\beta}_1^{\boldsymbol{Full}} + \boldsymbol{X}_2\boldsymbol{\beta}_2^{\boldsymbol{Full}} + \boldsymbol{\epsilon}^{\boldsymbol{Full}}) \\
&= \boldsymbol{\beta}_1^{\boldsymbol{Full}} + (\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1\boldsymbol{X}_2\boldsymbol{\beta}_2^{\boldsymbol{Full}} + (\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\boldsymbol{\epsilon}^{\boldsymbol{Full}}
\end{aligned} \tag{C.3}$$

By equation (C.3), the probability limit of $\hat{\boldsymbol{\beta}}_1^{\boldsymbol{Base}}$ is

$$\begin{aligned}
\boldsymbol{plim}\ \hat{\boldsymbol{\beta}}_1^{\boldsymbol{Base}} &= \boldsymbol{\beta}_1^{\boldsymbol{Base}} \\
&= \boldsymbol{\beta}_1^{\boldsymbol{Full}} + \boldsymbol{plim}\ (\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\boldsymbol{X}_2\boldsymbol{\beta}_2^{\boldsymbol{Full}} + \boldsymbol{plim}\ (\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\boldsymbol{\epsilon}^{\boldsymbol{Full}} \\
&= \boldsymbol{\beta}_1^{\boldsymbol{Full}} + \boldsymbol{\Gamma}\boldsymbol{\beta}_2^{\boldsymbol{Full}} \qquad (Weak\ Law\ of\ Large\ Number) \\
&= \boldsymbol{\beta}_1^{\boldsymbol{Full}} + \boldsymbol{\xi}
\end{aligned} \tag{C.4}$$

where $\boldsymbol{\Gamma} = (\boldsymbol{X_1'X_1})^{-1}\boldsymbol{X_1'X_2}$ is the matrix of coefficients obtained by projecting the plane spanned by the columns of $\boldsymbol{X_2}$ on the plane spanned by the columns of $\boldsymbol{X_1}$, namely:

$$\boldsymbol{X}_2 = \boldsymbol{X}_1\boldsymbol{\Gamma} + \boldsymbol{M} \tag{C.5}$$

[46]In many settings, the error term may contain components that influence both $\boldsymbol{X_1}$ and $\boldsymbol{X_2}$. For example, in survey experiments in which the manipulation variable is gender, the outcome is job promotion, and the moderator is masculinity, attention errors by respondents may influence both the (perceived) value of the manipulation variable and the assessment of masculinity. In such cases of endogeneity, $\boldsymbol{\beta_1^{Full}}$ and $\boldsymbol{\beta_2^{Full}}$ may be less than consistently estimated.

where $\boldsymbol{M}$ is the projection residuals on the plane orthogonal to the plane spanned by columns of $\boldsymbol{X_1}$. Equation (C.5) is essentially a matrix operation representation of the OLS estimation of equation (3), where the j-th column of $\boldsymbol{\Gamma}$ is $\boldsymbol{\Gamma^j} = [\hat{\lambda}^j \ \hat{\alpha}^j \ \hat{\delta}^j \ \hat{\theta}^j]'$ $(j = 1, 2)$, which is the vector of estimated coefficients in equation (3).

Conventionally, the derivation of equation (C.4) is used to illustrate that $\boldsymbol{\xi}$ is the omitted variable bias that results when we exclude $\boldsymbol{X_2}$ when estimating $\boldsymbol{\beta_1^{Base}}$ through using OLS estimation. In our case, we assume that both OLS estimators $\hat{\boldsymbol{\beta_1}}^{\boldsymbol{Base}}$ and $\hat{\boldsymbol{\beta_1}}^{\boldsymbol{Full}}$ are consistent estimators of $\boldsymbol{\beta_1^{Base}}$ and $\boldsymbol{\beta_1^{Full}}$, since $\boldsymbol{E[\epsilon^{Base}|X_1] = E[\epsilon^{Full}|X_1, X_2] = 0}$. We can use this omitted-variable bias formula to decompose the difference in estimated peer effects between the base and full models, $(\hat{\boldsymbol{\beta_1}}^{\boldsymbol{Base}} - \hat{\boldsymbol{\beta_1}}^{\boldsymbol{Full}})$, into meaningful components. We derive the difference as follows:

$$
\begin{aligned}
&\boldsymbol{plim}\ \hat{\boldsymbol{\beta_1}}^{\boldsymbol{Base}} = \boldsymbol{\beta_1^{Full}} + \boldsymbol{\xi} \\
&\Rightarrow \boldsymbol{plim}\ \hat{\boldsymbol{\beta_1}}^{\boldsymbol{Base}} = \boldsymbol{plim}\ \hat{\boldsymbol{\beta_1}}^{\boldsymbol{Full}} + \boldsymbol{plim}\ \hat{\boldsymbol{\xi}} \quad (\hat{\beta_1}^{Full} \ is\ a\ consistent\ estimator\ for\ \beta_1^{Full}) \\
&\Rightarrow \hat{\boldsymbol{\beta_1}}^{\boldsymbol{Base}} = \hat{\boldsymbol{\beta_1}}^{\boldsymbol{Full}} + \hat{\boldsymbol{\xi}} \\
&\Rightarrow \hat{\boldsymbol{\xi}} = \hat{\boldsymbol{\beta_1}}^{\boldsymbol{Base}} - \hat{\boldsymbol{\beta_1}}^{\boldsymbol{Full}}
\end{aligned} \tag{C.6}
$$

We know that estimating (C.1) and (C.2) using OLS yields the following equality:

$$\boldsymbol{X_1}\hat{\boldsymbol{\beta_1}}^{\boldsymbol{Base}} + \hat{\boldsymbol{\epsilon}}^{\boldsymbol{Base}} = \boldsymbol{Y} = \boldsymbol{X_1}\hat{\boldsymbol{\beta_1}}^{\boldsymbol{Full}} + \boldsymbol{X_2}\hat{\boldsymbol{\beta_2}}^{\boldsymbol{Full}} + \hat{\boldsymbol{\epsilon}}^{\boldsymbol{Full}}$$

Multiplying both sides by $\boldsymbol{(X_1'X_1)^{-1}X_1'}$ yields:

$$\hat{\boldsymbol{\beta_1}}^{\boldsymbol{Base}} = \hat{\boldsymbol{\beta_1}}^{\boldsymbol{Full}} + \boldsymbol{(X_1'X_1)^{-1}X_1'X_2}\hat{\boldsymbol{\beta_2}}^{\boldsymbol{Full}} \tag{C.7}$$

From equation (C.6), we know:

$$
\begin{aligned}
\hat{\boldsymbol{\xi}} &= \boldsymbol{(X_1'X_1)^{-1}X_1'X_2}\hat{\boldsymbol{\beta_2}}^{\boldsymbol{Full}} \\
&= \boldsymbol{\Gamma}\hat{\boldsymbol{\beta_2}}^{\boldsymbol{Full}} = \sum_{\boldsymbol{j=1}}^{2} \boldsymbol{\Gamma^j}\hat{\boldsymbol{\beta_2}}^{\boldsymbol{Full,\ j}}
\end{aligned} \tag{C.8}
$$

where $\hat{\boldsymbol{\beta_2}}^{\boldsymbol{Full,\ j}}$ is the j-th row of $\hat{\boldsymbol{\beta_2}}^{\boldsymbol{Full}}$. Equation (C.8) provides the decomposition of the difference in estimated peer effects into components that are caused by the explanatory power of $\boldsymbol{X_2}$, the mechanisms of measured absenteeism behaviors:

$$
\begin{aligned}
\hat{\boldsymbol{\xi}}_{PropFemalePeers} &= \hat{\boldsymbol{\beta}}^{\boldsymbol{Base}}_{PropFemalePeers} - \hat{\boldsymbol{\beta}}^{\boldsymbol{Full}}_{PropFemalePeers} \\
&= \underbrace{\boldsymbol{\Gamma}^{\boldsymbol{1}}_{PropFemalePeers}\hat{\boldsymbol{\beta_2}}^{\boldsymbol{Full,\ 1}}}_{\text{excused absences component}} + \underbrace{\boldsymbol{\Gamma}^{2}_{PropFemalePeers}\hat{\boldsymbol{\beta_2}}^{\boldsymbol{Full,\ 2}}}_{\text{unexcused absences component}}
\end{aligned} \tag{C.9}
$$

Rearranging equation (C.9), we have:

$$\underbrace{\hat{\beta}^{\boldsymbol{Base}}_{PropFemalePeers}}_{\substack{\text{overall effects}\\(\hat{\beta})}} = \underbrace{\hat{\beta}^{\boldsymbol{Full}}_{PropFemalePeers}}_{\substack{\text{unexplained}\\\text{component }(\underline{\hat{\beta}})}} + \underbrace{\boldsymbol{\Gamma}^{\boldsymbol{1}}_{PropFemalePeers}\hat{\boldsymbol{\beta}}_2^{\boldsymbol{Full},\,1}}_{\substack{\text{excused absences}\\\text{component }(\hat{\zeta^1}\hat{\alpha^1})}} + \underbrace{\boldsymbol{\Gamma}^{2}_{PropFemalePeers}\hat{\boldsymbol{\beta}}_2^{\boldsymbol{Full},\,2}}_{\substack{\text{unexcused absences}\\\text{component }(\hat{\zeta^2}\hat{\alpha^2})}} \tag{C.10}$$

Therefore, the explanatory power of excused absences in explaining the female classroom composition effect is $\hat{\zeta^1}\hat{\alpha^1}/\hat{\beta}$, and for unexcused absences is $\hat{\zeta^2}\hat{\alpha^2}/\hat{\beta}$.

D Appendix

We closely follow the notation and method of Heckman, Pinto, and Savelyev (2013) to establish equation (4). We suppress the notation for individual i, classroom c, school s, and cohort t for simplicity. Let D denote treatment assignment. $D = 1$ if a male or female student is treated (i.e., has more female peers in the classroom) and $D = 0$ otherwise. The subscript $d \in \{0,1\}$ of variables represents the treatment status when treatment is fixed at d. Let Y_0 and Y_1 be the counterfactual outcomes when d is fixed at "1" (treated) and "0" (non-treated), respectively. Therefore, the observed outcome of each individual is

$$Y = DY_1 + (1-D)Y_0 \tag{D.1}$$

Let the vector of absenteeism behaviors that can be changed by gender composition and partially produce the treatment effect be $\mathbf{m}$. The vector of absenteeism behaviors when treatment is fixed at d is $\mathbf{m}_d = (m_d^j : j \in \mathcal{J})$, where $\mathcal{J}$ is a set for all mechanisms. Therefore $\mathbf{m}_d$ can written as: $\mathbf{m}_d = D\mathbf{m}_1 + (1-D)\mathbf{m}_0$. For simplicity, we combine vectors of controls $\mathbf{X}'$ and $\mathbf{W}'$ in equation (1) in $\mathbf{Z}' = [\mathbf{X}'\ \mathbf{W}']$, and combine coefficients γ and δ in $\eta = [\gamma\ \delta]$. Outcome equation of treatment status d for a male or female student can be written as follows:

$$Y_d = \kappa_d + \mathbf{m}_d'\zeta + \mathbf{Z}'\eta + \tilde{\epsilon_d}, \qquad d \in \{0,1\} \tag{D.2}$$

where $\tilde{\epsilon_d}$ is assumed to be a zero-mean error term and independent of regressors $\mathbf{m}_d'$ and $\mathbf{Z}'$. Treatment is assumed to affect mediating factors (so $\mathbf{m}_d'$ varies with d), but not to affect the impact of mediating factors and controls on outcomes (so ζ and η are independent of d). Because we cannot measure all relevant mechanisms, we let $\mathcal{J}_p \subseteq \mathcal{J}$ be the set of mechanisms on which we have measurements so that $\mathcal{J} \setminus \mathcal{J}_p \subseteq \mathcal{J}$ is the set of mechanisms on which we have no measurements. We decompose the term $\mathbf{m}_d'\zeta$ in equation (D.2) into measured components and unmeasured components:

$$\begin{aligned}
Y_d &= \kappa_d + \sum_{j\in\mathcal{J}} \zeta^j m_d^j + \mathbf{Z}'\eta + \tilde{\epsilon_d} \\
&= \kappa_d + \underbrace{\sum_{j\in\mathcal{J}_p} \zeta^j m_d^j}_{\substack{\text{measured}\\ \text{mechanisms}}} + \underbrace{\sum_{j\in\mathcal{J}\setminus\mathcal{J}_p} \zeta^j m_d^j}_{\substack{\text{umeasured}\\ \text{mechanisms}}} + \mathbf{Z}'\eta + \tilde{\epsilon_d} \\
&= \kappa_d + \sum_{j\in\mathcal{J}_p} \zeta^j m_d^j + \sum_{j\in\mathcal{J}\setminus\mathcal{J}_p} \zeta^j [m_d^j - E(m_d^j)] + \sum_{j\in\mathcal{J}\setminus\mathcal{J}_p} \zeta^j E(m_d^j) + \mathbf{Z}'\eta + \tilde{\epsilon_d} \\
&= \underbrace{\{\kappa_d + \sum_{j\in\mathcal{J}\setminus\mathcal{J}_p} \zeta^j E(m_d^j)\}}_{\tau_d} + \sum_{j\in\mathcal{J}_p} \zeta^j m_d^j + \mathbf{Z}'\eta + \underbrace{\{\tilde{\epsilon_d} + \sum_{j\in\mathcal{J}\setminus\mathcal{J}_p} \zeta^j [m_d^j - E(m_d^j)]\}}_{\epsilon_d} \\
&= \tau_d + \sum_{j\in\mathcal{J}_p} \zeta^j m_d^j + \mathbf{Z}'\eta + \epsilon_d, \qquad d \in \{0,1\}
\end{aligned} \tag{D.3}$$

Here, $\tau_d = \kappa_d + \sum_{j \in \mathcal{J} \setminus \mathcal{J}_p} \zeta^j E(m_d^j)$, and $\epsilon_d = \tilde{\epsilon_d} + \sum_{j \in \mathcal{J} \setminus \mathcal{J}_p} \zeta^j [m_d^j - E(m_d^j)]$, which is a zero-mean error term. For ζ^j to be consistently estimated, we assume that m_d^j is measured without error and independent of ϵ_d, $d \in \{0, 1\}$, which implies that measured mechanisms are uncorrelated with unmeasured mechanisms for both the treated and non-treated group.

Feeding equation (D.3) into equation (D.1) , we have:

$$\begin{aligned}
Y &= D\tau_1 + D \sum_{j \in \mathcal{J}_p} \zeta^j m_1^j + D\mathbf{Z}'\eta + D\epsilon_1 + \tau_0 + \sum_{j \in \mathcal{J}_p} \zeta^j m_0^j + \mathbf{Z}'\eta + \epsilon_0 \\
&\quad - D\tau_0 - D \sum_{j \in \mathcal{J}_p} \zeta^j m_0^j - D\mathbf{Z}'\eta - D\epsilon_0 \\
&= \tau_0 + \underbrace{(\tau_1 - \tau_0)}_{\tau} D + \sum_{j \in \mathcal{J}_p} \zeta^j \underbrace{[Dm_1^j + (1-D)m_0^j]}_{m^j} + \mathbf{Z}'\eta + \underbrace{[D\epsilon_1 + (1-D)\epsilon_0]}_{\epsilon} \\
&= \tau_0 + \tau D + \sum_{j \in \mathcal{J}_p} \zeta^j m^j + \mathbf{Z}'\eta + \epsilon \qquad \text{(D.4)}
\end{aligned}$$

Equation (D.4) is the alternative representation of equation (4), where $\epsilon = D\epsilon_1 + (1-D)\epsilon_0$ is a zero-mean error, and $m^j = Dm_1^j + (1-D)m_0^j$ represents the mechanism we can measure. $\tau = \tau_1 - \tau_0$ captures the treatment effect that is explained by unmeasured mechanisms, or unexplained by our measured mechanisms, which is essentially the $\underline{\beta}$ in equation (4). Replacing τ_0 with $\underline{\lambda_{s,t}}$ and D with $PropFemalePeers_{-i,c}$ would not change the interpretation of estimates and restore equation (4) completely. ζ^j captures the effect of measured mechanism m^j on outcomes. For ζ^j to be consistently estimated, we need to assume that measured mechanisms $m^j, j \in \mathcal{J}_p$, are independent of the unmeasured mechanisms captured in ϵ.