Does Minority Status Drive Women Out of Male-Dominated Fields?

Xiaoyue Shan*

February 6, 2021

click here for the most recent version

Abstract

This paper examines whether women's status as the minority gender causes them to drop out of male-dominated fields. I conduct a field experiment in an introductory economics course where I randomly assign students to small study groups with different gender compositions. Results show that women assigned to female-minority groups become 10 percentage points more likely to drop out of the course than women assigned to other groups. In contrast, minority status does not significantly affect men's dropout behavior. I present suggestive evidence on educational expectations and peer-to-peer interaction as underlying mechanisms through which minority status raises female dropout. Women in the minority form more pessimistic expectations about their future academic achievement and interact less frequently with their peers. The findings of the paper suggest that minority status can reinforce itself and underrepresented groups can be trapped in a vicious cycle of being in the minority.

*I thank my supervisors, Ulf Zölitz and Michel Maréchal, for their great guidance and generous support throughout the project. I appreciate the encouragement and feedback from wonderful female role models in economics: Teodora Boneva, Anne Brenøe, Patricia Funk, Shelly Lundberg, Olga Stoddard, Linda Tesar, and Jane Zhang. I also thank Pietro Biroli, David Dorn, Jan Feld, Alexandra de Gendre, Andrea Hofer, Nicolás Salamanca, Giuseppe Sorrenti, Johannes Ullrich, Joachim Voth, as well as seminar attendants at University of Zurich, IZA, SOFI Stockholm University, Graduate Institute Geneva, Jinan University, GEEZ, AYEW, and the CRC Summer School on Peer Effects for useful comments and suggestions. The study is registered in the AEA RCT Registry (ID: AEARCTR-0004392) and it has received IRB approval from University of Zurich (see Appendix D for details of the ethical considerations in this study). Xiaoyue Shan: Department of Economics, University of Zurich; email: xiaoyue.shan@econ.uzh.ch.

1 Introduction

Women are underrepresented in a lot of majors and professions, such as STEM and economics. When climbing the career ladder, they are also more likely to drop out and the minority status of women in these fields becomes more and more pronounced—a phenomenon widely known as the "leaky pipeline".[1] For example, only about a third of students majoring in economics are women in the United States; the share further drops to 28% and 14% for assistant and full professors respectively (Lundberg, 2018). The leaky pipeline is concerning not only because of the loss of diversity, which may affect the collective creativity and performance (Herring, 2009; Woolley et al., 2010; Hoogendoorn et al., 2013). The lack of women in skill-intensive and high-earning professions also implies a gender pay gap, potential misallocation of talent across sectors, and even barriers to economic growth (Arulampalam et al., 2007; Autor, 2014; Blau & Kahn, 2017; Hsieh et al., 2019). Given the worrying implications of the leaky pipeline, the natural question to ask is why women "leak" out of male-dominated fields. While acknowledging many potential reasons, this paper focuses on a fundamental factor: minority status itself.

To examine whether minority status causes women to leave male-dominated fields, I conduct a field experiment in an introductory economics course at a Swiss university by randomizing the composition of study groups. The setting has three nice features to tackle the research question. First, the course is male-dominated. As is typical in economics, around 35% of the students are female. Second, most students take the course in their first semester at college, a period when they meet new peers and form new friendships. Third, the course is compulsory for students' major or minor study programs related to business, economics, and informatics. Therefore, dropping the course also implies dropping out of the study program.

The intervention of the experiment is to randomly assign students to small study groups of four, thus creating exogenous variation in the group gender composition. More specifically, some women are assigned to female-minority groups where women are in the minority, while other women become members of gender-balanced, male-minority, or female-only groups. I then compare women in female-minority groups to other women to identify the impact of minority status on women.

[1] This paper uses the term *minority status* to refer to the numerical underrepresentation of women among peers. A similar term used in the literature is the *token status* (Laws, 1975; Kanter, 1977). However, the tokenism concept stresses the intention to include the minority group, the tokens, for the appearance of equality.

Results show that minority status in the study group significantly increases the female dropout rate. Women assigned to female-minority groups become 10 percentage points more likely to drop out of the course than other women—an effect both economically and statistically significant.[2] In contrast, assignment to male-minority groups does not significantly affect men's dropout behavior. I further show that the dropout effect of minority status is a product of the unbalanced gender ratio rather than other peer characteristics, like academic ability and personality traits. The impact of minority status on women's dropout remains strong and significant even after controlling for a rich set of peer characteristics, including high school achievement, the Big-Five personality traits, competitiveness, and gender attitudes.

To explain why women in the minority are more likely to drop out, I focus on two underlying mechanisms capturing both the academic and social domain of college life: educational expectations and peer-to-peer interaction.[3] Educational expectations refer to expectations about future academic achievement and capture students' confidence about their own academic ability. Peer-to-peer interaction is the interaction with study group peers and serves as a proxy for students' integration to the social environment. A student who expects lower academic achievement or perceives lower social integration in a study program is more likely to withdraw from the program. I present a simple framework in Appendix A to help clarify how minority status may influence dropout through the two channels.

To test the first mechanism, I examine how minority status affects students' educational expectations at the end of the semester. I find that women assigned to female-minority groups form significantly lower expectations: they become more pessimistic about their future academic achievement. The finding is consistent with studies showing that women have lower self-confidence and the lack of confidence can induce them to shy away from competitive tasks or male-dominated fields (Bengtsson et al., 2005; Kamas & Preston, 2012; Niederle & Vesterlund, 2007; Reuben et al., 2017). The finding also echoes the "stereotype threat" effect: people perform worse or expect lower performance when faced with negative stereotypes about their group (Cadinu et al., 2003; Nguyen & Ryan, 2008; Spencer et al., 2016). Being in the minority

[2]The average dropout rate is 13% among female participants in the experiment and 9% among male participants. When using the traditional linear model, I find that one additional woman in a study group lowers the female dropout rate by 4 percentage points.

[3]The two mechanisms correspond to the two factors of dropout behavior in the well-known *student integration model* (Tinto, 1975) in the education literature. According to the model, educational expectations and perceptions of social integration are the key determinants for voluntary withdrawal from study programs in college.

in peer groups can activate and exacerbate the stereotype threat, which may further trigger women's self-doubt about ability and cause them to drop out of male-typed fields.

To test the second mechanism, I investigate the impact of minority status on students' self-reported interaction with study group peers. I find that women assigned to female-minority groups interact significantly less with their peers—both in academic and social activities. Due to their social exclusion, women may feel that they belong less to the program and consequently drop out. I also find suggestive evidence that women's exclusion in male-dominated groups is driven by men's preference for male peers rather than women's preference for female peers. Consistent with evidence on gender homophily in social networks (Shrum et al., 1988; Ibarra, 1992; Mengel, 2020), I find that men display more homophily than women. Given three randomly assigned peers of different genders, men interact more frequently with male peers and value male peers more than female peers. In contrast, women's preference for peers depends less on peers' gender.

I continue by asking which women drop out as a result of the unbalanced gender ratio. Heterogeneous analyses show that the effects of minority status on women are driven by high-ability women whose high school math grades are above the median and women who exhibit typically-female characteristics, such as personality traits, competitiveness, and gender attitudes. In contrast, women with low math achievement or typically-male characteristics are not significantly affected by minority status. The findings suggest that economics as a male-dominated field is losing the highly-skilled and typically-female women, thus facing a loss of talent and diversity.

While many studies have explored why women are underrepresented in competitive, male-typed, or high-status fields, little attention has been paid to the impact of minority status on women's persistence in these fields. Previous studies show that both individual and contextual factors can discourage women from choosing male-dominated majors or professions. Individual factors include preferences for competition (Croson & Gneezy, 2009; Niederle & Vesterlund, 2011; Buser et al., 2014), preferences for workplace attributes like flexibility and stability (Wiswall & Zafar, 2018), beliefs about earnings and ability (Stinebrickner & Stinebrickner, 2012; Wiswall & Zafar, 2015; Bian et al., 2017), and so on. Contextual factors include teachers (Bettinger & Long, 2005; Lavy, 2008), family background (Dryler, 1998; Cheng et al.,

2017), and peers (Anelli & Peri, 2019; Brenøe & Zölitz, 2020; Zölitz & Feld, 2020).

Only a few studies find suggestive evidence that a more salient minority status can increase women's dropping out of male-dominated settings. Using photos of classes at the U.S. Military Academy in the 1970s and 1980s, Huntington-Klein & Rose (2018) show that women with fewer female classmates are less likely to advance to the next year. Based on longitudinal data, Griffith (2010) and Bostwick & Weinberg (2018) find that female students in STEM programs are less likely to persist in the field if they have a lower fraction of female peers in the cohort.[4] Nevertheless, all these studies rely on limited administrative data, and only one of them employs random assignment of peers to alleviate endogeneity concerns.

Taken together, this paper makes three main contributions to the literature. First, I conduct a field experiment to show that minority status can cause women to drop out of male-dominated fields. The finding implies the existence of a "minority trap" (as analogous to "poverty traps"): underrepresented minority groups are trapped in a vicious cycle of being in the minority. Second, by combining administrative data and rich survey data, this paper provides evidence on potential mechanisms of the minority trap. Women in the minority are more likely to leave a male-dominated field potentially because they are less integrated to the social environment and have lower confidence about their future achievement in the field. These two mechanisms are fairly generic, implying that other underrepresented groups may face a similar minority trap.

The third contribution of the paper is that it provides a new possibility for policy makers to increase women's presence in male-dominated settings. To attract more women or men into gender-incongruent majors or professions, existing studies have examined policies related to role models, mentors, information, and gender bias (Blau et al., 2010; Carnes et al., 2015; Avilova & Goldin, 2018; Porter & Serra, 2019; Delfino, 2020). The results of this paper suggest that small peer groups where women are not in the minority may serve as a light-touch and cost-effective policy tool. Rather than looking at the gender composition of large cohorts or classes which

[4]In another setting, Oosterbeek & Van Ewijk (2014) find no significant impact of gender composition in teaching sessions on women's dropout of business or economics. Li et al. (2013) and Anil et al. (2017) also examine how class ability or gender composition affects dropout in schools; Fischer (2017) studies the effect of class ability composition on women's STEM persistence. More broadly, a large body of literature shows that peers' gender and ability can affect educational outcomes. For example, peers' gender composition can impact academic achievement in school (Lavy & Schlosser, 2011; Hill, 2015; Gong et al., 2019) and university (Ficano, 2012; Schneeweis & Zweimüller, 2012; Hill, 2017; Booth et al., 2018). Other than gender, peers' cognitive or non-cognitive abilities (Zimmerman, 2003; Carrell et al., 2009; Booij et al., 2017; Carrell et al., 2018; Zou, 2019; Zárate, 2019), cultures (Bursztyn et al., 2019), and personality traits (Golsteyn et al., 2020) can also affect other students' educational outcomes.

are difficult to manipulate,[5] this paper focuses on small peer groups which are ubiquitous in educational and workplace settings (Devine et al., 1999; Lazear & Shaw, 2007) and offer more scope for policy makers. Given a male-dominated class or cohort, small subgroups in which women are not in the minority can even counteract the effect of an unbalanced environment.

My findings are also consistent with a few recent studies examining gender disparities in group dynamics based on laboratory experiments. Coffman (2014) shows that women are less confident about their ability in male-typed domains and are thus less likely to contribute their ideas to a group. Chen & Houser (2019) find that women assigned to mixed-gender groups are more likely to shy away from leadership in gender-incongruent areas. Born et al. (2020) also show that women in male-majority teams have lower leadership aspirations partly because they are less confident. Studying mixed-gender teams in an accounting program, Stoddard et al. (2020) find evidence that token women are perceived as less influential by peers in a laboratory exercise than women in female-majority teams. However, all these results are found in the laboratory and it is not clear to what extent these results can be extrapolated to the world beyond (Levitt & List, 2007; Falk & Heckman, 2009). By showing that group gender composition can also affect group interactions, women's self-confidence, and high-stakes decisions in a real-world setting, this paper bridges the gap between laboratory experiments and naturally-occurring environments.

The article proceeds as follows. Section 2 describes the experiment design, measurement of key variables, and the data. After introducing the empirical strategy in Section 3, Section 4 presents the main results. Section 5 conducts robustness checks and heterogeneity analyses. I conclude in Section 6.

2 The Experiment and Data

2.1 The Experiment Design

The experiment took place in an introductory economics course at a Swiss university. It is a male-dominated course, with the overall share of female students ranging from 30% to 40% depending on the cohort. About 90% of the students take the course in their first semester at

[5]For example, the average size of cohorts is 17 in Bostwick & Weinberg (2018); the average class size is 32 in Huntington-Klein & Rose (2018) and 39 in Oosterbeek & Van Ewijk (2014).

university and the remaining 10% are repeating the course.

Figure 1 shows the timeline of the experiment. In mid-September, when the Fall semester started, students completed a baseline survey and registered for study groups. The baseline survey elicited information about students' demographics, high school background, educational expectations, personality traits, and socioeconomic preferences. Afterward, I randomly assigned students who had voluntarily registered into groups of four. The group assignment was stratified at the broad level of study programs: economics-business, informatics, and others. I conducted the first round of the experiment in 2018, and replicated it in 2019.

Study groups differed in gender composition as a result of the random assignment. In total, the experiment sample includes 620 students in 155 study groups. Among 259 women, 53 were assigned to female-minority groups and consequently experienced minority status, while the rest were assigned to gender-balanced or female-dominated groups. Similarly, 28 out of 361 men were assigned to male-minority groups.

After the group assignment, students received an email introducing their group members with names, email addresses, phone numbers, and WhatsApp account information. WhatsApp is the most popular communication tool among students; more than 90% of students use it to communicate with their friends. Students also received emails encouraging group members to attend lectures and tutorial sessions, review lecture content, and work on problem sets together. To foster social interaction, each group was provided with a voucher worth 20 Swiss Francs (CHF) for drinks at the local university bar.

At the end of the semester but before the final exam, students filled out the endline survey, which elicited educational expectations again and one-to-one academic and social interaction among group members. In mid-December, the final exam took place, but some students did not take the exam, implying that they had dropped out of the course during the semester.

2.2 Data and Descriptive Statistics

The data used in this paper come from two sources: administrative education data and survey data at baseline and endline. Table 1 presents the descriptive statistics separately for women and men. Panel A shows the summary statistics of baseline characteristics, including high school background and educational expectations. Panel B summarizes endline educational expectations

and self-reported interaction with peers. Panel C shows descriptive statistics for dropout and final course grades.

In the baseline survey, students reported high school final grades for mathematics and the first language: the grades range from 1 to 6.[6] The survey elicited three aspects of educational expectations from the short run to the long run: (1) the expected grade for the course, (2) the expected probability of passing the first year in the university,[7] and (3) the expected highest academic degree—high school, bachelors, masters, or doctoral degree. To reduce concerns of multiple hypotheses testing and increase measurement accuracy, I combine the three questions and create a normalized index for the overall educational expectation. Higher values of expectation represent higher optimism or academic self-efficacy. The baseline survey also elicited the Big-Five personality traits—conscientiousness, extraversion, agreeableness, openness, and neuroticism—using the 15-item inventory.

The endline survey in 2019 also elicited pairwise relations within study groups. As Figure 2 illustrates, each student reported the intensity of interaction with each of his/her three peers. The intensity of academic/social interaction refers to the frequency of studying together or joint social activities: never, less than once per month, once per month, two or three times a month, once per week, and more than once a week. Also, for the first time, the survey asked students to report their willingness to pay (WTP) for peers in a hypothetical scenario: how much they would be willing to pay to have someone as their study mate again. They can choose any amount between 0 and 100 CHF.[8] The WTP measurement helps to uncover how peer-to-peer interaction depends on students' preferences for peers.

For both academic and social interaction, I first standardize the pairwise interaction fre-

[6] Around 90% of students in the course obtained their Matura—high school degree—in Switzerland. International students reported the grade according to the Swiss scale from point 1 to 6, with point 4 as the passing grade. More than 85% of students listed German as their first language in high school.

[7] In Swiss universities, most study programs have an assessment policy. In the first year at university, students must attend and pass the compulsory courses. If they fail a course, they have to retake it in the second year. If they have not successfully passed all assessment courses by the end of the second year, they cannot continue the study program anymore. As a result, the first year is usually called the "assessment year", and first-year students are referred to as "assessment students".

[8] Students report their WTP only for peers with effective academic/social interactions. For peers without any effective interaction, I assume the WTP is equal to zero—the results are similar if I do not impose this assumption. In addition to the self-organized face-to-face interactions as elicited by the survey, students can also meet in lectures and tutorial sessions, or communicate through social media like WhatsApp. In a two-step interaction framework as in Mayer & Puller (2008), group members first meet each other and decide whether to have effective interactions. Given the frequent nudges that the experiment provides to encourage group members to meet each other, I believe that most students have met or communicated with their group members. Therefore, if two students "never" study or attend social events together, that indicates a failed attempt of the second-step interaction—due to a mismatch in individual preferences.

quency. Then, I calculate the average interaction with the three peers and standardize the average value across individuals. Finally, I use the standardized mean of academic and social interaction as an indicator for the overall interaction with peers.[9] For example, in Figure 2, the interaction intensity of student A with his/her peers is the normalized average of interaction$_{AB}$, interaction$_{AC}$, and interaction$_{AD}$.

The key outcome variable of interest is dropout from the course. A student who drops out from the course does not attend the final exam for the course and the exam grade is the final grade. Typically, the course is mandatory for successful completion of a student's registered major or minor program. As a result, dropout from the course translates into dropout from the study program, unless the student repeats the course in the following year. Furthermore, students majoring in business, economics, or informatics can only repeat the course once. Taken together, dropout from the course is a high-stakes decision with significant consequences. I also observe the final grade for the course, which ranges from 1–6. Students who drop out receive a 1—the lowest possible grade.[10]

3 Empirical Strategy

I use the following regression model to estimate the causal impact of minority status on student outcomes:

$$Y_{ig} = \alpha + \beta \cdot \text{Minority}_{ig} + \gamma \cdot \text{Minority}_{ig} \times \text{Male}_{ig} + \delta \cdot X_{ig} + \varepsilon_{ig}. \tag{1}$$

Y_{ig} is the outcome of student i in group g. The primary outcome variable is an indicator for course dropout, and the two intermediate outcomes (the two potential mechanisms) are the

[9] I do not observe all the pairwise interactions, because the response rate for the endline survey is around 70%. For each observed pairwise link, either both students report the frequency, or only one student reports it. When constructing the individual-level index, if both sides report the frequency, I use the average value to increase measurement accuracy. The frequencies reported by the two sides are positively correlated with a correlation coefficient of .77 for academic interactions and .69 for social interactions. If only one side answers the question, I use the answer for both sides to increase the sample size.

[10] To figure out what students typically do after dropping out of the course, I track them through course records and a short follow-up survey in July 2020. Appendix Figure B1 summarizes the pathways that students took after dropping out in 2018. Of a total of about 1,300 students in 2018, 208 students dropped out of the course without taking the final exam. Of those 208 students, 71% dropped out of the course permanently and 29% repeated the course the next year. Among those who did not repeat the course, 32% did not have a valid university email in July 2020, meaning that they no longer studied at the university. Among those retaking the course, 34% either dropped out again or failed the exam. In summary, while a small fraction of students repeated and passed the course successfully after the dropout, most dropouts resulted in permanent separation from the economics course.

endline educational expectation and peer-to-peer interaction. Minority_{ig} is a dummy variable indicating whether student i is the minority gender in group g. More specifically, $\text{Minority}_{ig} = 1$ if a woman is assigned to a female-minority group or a man is assigned to a male-minority group. Male_{ig} is an indicator for student gender. X_{ig} is a battery of individual characteristics such as gender, major, course-retaking status, personality traits, high school grades, and experiment year. ε_{ig} is the residual term.

As this paper uses a randomized experiment, the preferred inference method is randomization inference (RI)—also known as permutation tests or randomization tests. RI constructs nonparametric distributions of test statistics by replicating the randomization, rather than relying on asymptotic assumptions that may not hold in randomized experiments (Fisher, 1935; Imbens & Wooldridge, 2009; Young, 2019). More and more experimental studies use RI to test whether a treatment has any impact on the sample and report the RI p-values (Cohen et al., 2010; Fujiwara & Wantchekon, 2013; Gneezy et al., 2019; Alan et al., 2019; Lindqvist et al., 2020). Throughout the paper, I derive the RI p-values by re-randomizing students into study groups following the same stratification design and re-estimating (β, γ) for 1,000 times. Under the null hypothesis of no treatment effect, the exact p-value equals the fraction of estimated coefficients larger in absolute values than the actual coefficient. For robustness, I also present standard errors clustered at the study group level.

Because study groups are randomly assigned, β identifies the causal effect of minority status on women, and γ estimates the gender gap in the impact of minority status. To confirm that the randomization was successful, I test whether individual characteristics are balanced between students assigned to be in the minority and students not assigned to be in the minority. As Table 2 shows, minority status does not significantly predict any of the baseline characteristics. Notice that most baseline characteristics are measured through the baseline survey (with a take-up rate of 88%), and the two mechanisms are observed in the endline survey (with a take-up rate of 66%). However, I find no evidence suggesting unbalanced survey response: minority status does not significantly predict survey response at baseline or endline. In Section 5, I further show that my results on mechanisms are robust when using Lee Bounds (Lee, 2009) analysis and Inverse Probability Weighted (IPW) regressions.

Additionally, I test whether the group assignment is random. If the study group assignment

is random within assignment strata, group dummies should not significantly predict baseline characteristics such as gender, high school background, and personality traits. I test this in two steps. I first regress each baseline characteristic on study program and year dummies and derive the residuals. Then, I regress the residuals on study group dummies and test the joint significance of group dummies. As Appendix Table B1 shows, I find no evidence suggesting that the group assignment is systematically related to individual characteristics.

4 Main Results

4.1 The Impact of Minority Status on Dropout

Figure 3 shows how dropout varies with group gender composition and own gender. Women assigned to female-minority groups have a much higher dropout rate than women assigned to other groups. The difference in means between women in female-minority groups and women in other groups is about 8.5 percentage points. For men, the differences across gender compositions are very small.

Table 3 presents the estimated impacts of minority status on dropout, using ordinary least squares (OLS) regressions. The dependent variable is an indicator for dropping out of the course, which is compulsory for students' registered major or minor programs. All regressions include the basic controls—student major, course-retaking status, and year fixed effects. Column (1) include the whole sample of women or men, while columns (2)–(4) only include students who have completed the baseline survey. I gradually include high school controls and the Big-Five personality traits in the regressions.

Results show that minority status in the study group significantly increases women's dropout rate. Overall, women assigned to be in the minority become about 10 percentage points more likely to drop out of the course than women in other groups. The estimated impact is robust to controlling for high school background and personality traits, as well as using different inference methods. The effect size is substantial, considering that the average dropout rate of women is about 13%. The impact of minority status on men's dropout is about 10 percentage points smaller, meaning that the effect size is close to zero. In Appendix Tables B3 and B4, I split the sample by gender, and separately estimate the effect of minority status on women and men. The split-sample analyses lose some power, but the results are virtually the same: women

assigned to female-minority groups are more likely to drop out, while minority status does not significantly affect men's dropout behavior.

The focus of this paper is women's dropping out of economics. Given the salient gap in dropout between women assigned to be in the minority and other women, I concentrate my analysis on the binary comparison. In Appendix Table B5, I employ different models to analyze the impact of group gender composition on the dropout behavior of women and men. The results imply the same pattern as Figure 3 shows. When using the linear model, students' dropout rate decreases with the number of women per study group. Overall, non-linear models are fitting the data better, with larger R-squared values. However, for women, the fully saturated model is not much better than the simple binary comparison by minority status.

Appendix Table B6 also examines how minority status affects the final grade for the course. Notice that if a student drops out, the grade is simply point 1, the lowest grade. When including all students in the analysis, I find that minority status has a negative impact on women's grades. Nevertheless, when only focusing on students who take the final exam, minority status does not affect women's grades significantly. That suggests the overall negative impact on women's grades is driven by the dropout effect. The gender gap in the impact of minority status on grades is not statistically significant.

Peer Gender or Other Peer Traits? Having shown that women in male-dominated groups are more likely to drop out, I further ask whether this effect is a product of peer gender or other peer characteristics. As the literature on peer effects documents, various student characteristics such as ability and personality traits can impact peers' educational outcomes. It is therefore valid to conjecture that the dropout effect of minority status can be driven by other characteristics of male peers. For instance, male students are on average more competitive and self-confident, and display stronger gender stereotypes. These characteristics may also affect women's dropout behavior.[11]

To formally answer the question, I check how the dropout effect of minority status changes when controlling for other peer traits. More specifically, I control for the average value of peers' characteristics (the leave-own-out means) when regressing female dropout on the minority in-

[11] In effect, I find that some peer traits like conscientiousness and agreeableness do have a significant impact on women's dropout (see Appendix Figure B9).

dicator. I consider the following peer traits measured at baseline: high school background, personality traits, gender attitudes, and self-reported preferences like risk preference and patience. I only include students with non-missing values of peer characteristics—all of their peers have completed the baseline survey, to make sure that the change in coefficient can not be driven by differences in the estimation sample. Figure 4 plots the estimated effects of minority status on female dropout without peer controls and with different sets of peer controls. Results show that the impact of minority status on female dropout remains strong and statistically significant as a rich set of peer characteristics is controlled for. The impact even becomes stronger and gets more precisely estimated. The finding suggests that women in the minority are discouraged primarily by the unbalanced gender ratio rather than other peer characteristics.

Gender Difference in the Impact of Minority Status Knowing that minority status in the study group increases women's dropout but does not significantly affect men's dropout, a follow-up question to ask is: what is the source of the gender gap in the effect of minority status? One possible reason is that women and men are systematically different in their ability, personality traits, or preferences, so that men are simply less responsive to the gender composition of the study group. This explanation is not very plausible, because as Appendix Tables B3 and B4 show, minority status does not significantly affect men's dropout even after controlling for high school grades and personality traits. Similarly, the impact of minority status on women's dropout is robust to the inclusion of a wide range of women's own and their peers' characteristics.

The more plausible reason is that women are overall underrepresented in the economics course and the corresponding study fields such as business, economics and informatics, while men are the majority gender. In other words, the course and the study programs are stereotypically male. Nevertheless, such a perception may get strengthened or weakened as students are exposed to different social networks with different gender compositions. Intuitively speaking, if a woman in a male-typed field is assigned to male-dominated peer group, she may believe more firmly that the field is male-dominated and she as a women is less belonged to it. I test the conjecture by examining how minority status in the group affects students' beliefs about the overall gender

ratio in the course.[12]

Figure 5 shows female students' believed share of women in the course—their perceived minority status of women in the course—by group gender composition. At baseline, all women have similar beliefs, at a level slightly lower than the actual share. Interestingly, in the end of the semester, those assigned to female-minority groups adjust their beliefs further away from the actual level: their perceived minority status of women in the course is even more salient or stronger. In contrast, women assigned to other gender compositions maintain their beliefs at a similar level as at baseline. The regression results presented in Table 4 tell the same story: being in the minority in a study group significantly lowers women's believed share of women in the course but raises men's believed share of women. The finding suggests that being in the minority in a small social network can raise the salience of women's minority status in the overall environment and make women more likely to drop out.

4.2 Mechanism: Educational Expectations

This section examines the impact of minority status on educational expectations—one potential mechanism through which minority status drives dropout. The intuition is that minority status in peer groups lowers students' academic self-efficacy or self-confidence—especially for women in a male-dominated field—and consequently leads to more dropout. Figure 6 shows how the overall educational expectation at endline varies with group gender composition. Similar to dropout behavior, a large gap in expectation exists between women in female-minority groups and women in other groups. The overall educational expectation of minority women is lower by about 0.4 standard deviations, and the difference is significant at the 5% level. Again, the cross-composition differences for men are small and insignificant.

Table 5 presents the main results: the impact of minority status on the overall educational expectation at endline (a standardized indicator). I find that minority status lowers women's expectations for future academic achievement by about 0.4 standard deviations. The effect is statistically significant at the 5% level. The effect size and significance level are robust to controlling for high school background, personality traits, and the expectation at baseline. For men, the effect is much smaller in magnitude and is closer to zero, but the gender gap is not

[12]The belief is elicited with a survey question asking "out of the approximately 1,000 students in the course, how many are women". Notice that the question is only available in 2019, at the baseline and endline, so the analysis only uses the sample in 2019.

statistically significant. Appendix Table B8 presents the estimation results when splitting the sample by gender. The split sample results are very similar: women assigned to be in the minority have significantly lower expectations than other women, while being in the minority does not significantly affect men's expectations.

Next, in Figure 7, I show the impacts of minority status on different aspects of educational expectations: the expected grade for the course, the expected probability of successfully passing the first year at university, the highest academic degree expected to be obtained, and the likelihood of expecting to obtain a PhD degree. I find that both women and men assigned as the minority gender expect lower grades for the course. Minority status lowers women's expected probability of passing the first year at university, but the effect fails to reach significance at conventional levels. Women in the minority also tend to expect lower academic degrees. They are 10 percentage points less likely to expect a doctoral degree than other women.

Having shown that minority status has a negative impact on women's educational expectations, I proceed by asking what the expectations reflect. Do they reflect students' assessment of own ability, their educational aspirations, or the comparison to a reference group? I analyze the potential underlying drivers of expectations by examining the following measurements of the expected grade: (1) the gap between the expected grade and the realized grade, as a proxy for confidence in one's ability; (2) the aspired minimal grade for the course, as a proxy for educational aspiration;[13] and (3)–(4) the believed grade of the average woman/man in the previous cohort, which reflects potential comparison to same-gender or opposite-gender peers. I examine the impact of minority status on the four outcomes, separately for women and men, and plot the coefficients in Figure 8.

Results show that women in the minority in study groups significantly underestimate their ability and have lower grade aspirations, but their beliefs about same-gender and opposite-gender peers' ability are similar to women not in the minority. The results suggest that minority status lowers women's expectations mainly through the self-assessment and aspiration channels, rather than the social comparison channel. Minority status also (insignificantly) lowers men's expected grades, and the negative impact is potentially driven by the social comparison effect.

[13]The aspiration question exists in the 2019 baseline/endline survey: "what is the minimal grade that you want to achieve for the course".

4.3 Mechanism: Peer-to-Peer Interaction

To test the second mechanism, this section estimates the impact of minority status on peer-to-peer interaction within study groups. Figure 9 shows how students' overall interaction with peers depends on group gender composition. The difference between women in female-minority groups and other women is significant at the 5% level, but the difference between men in male-minority groups and other men is not significant. The figure also highlights that gender-balanced groups induce more peer-to-peer interaction than unbalanced groups.[14]

Table 6 shows the impact of minority status on a student's interaction with his/her three peers in the group. The dependent variable is the normalized index for the overall frequency of academic and social interaction with peers. Results show that women assigned to female-minority groups report significantly less interaction with group peers. The impact of minority status on men is much smaller in magnitude and closer to zero. Again, the results are similar when conducting split sample analysis by gender (see Appendix Table B9).

Next, I examine the disaggregated results by looking at different measurements of peer-to-peer interaction: the average academic interaction with peers, the average social interaction with peers, the number of peers studying together more than once per month, and the number of peers having social events together for at least once per month. Figure 10 shows the results separately for women and men. I find that minority status significantly lowers women's interaction with peers, in terms of both academic and social activities. Men in male-minority groups are also less likely to meet for social events with assigned peers, but their academic interaction with peers is not significantly lower.[15]

Why Are Women Isolated in Male-Dominated Groups? After showing that women in the minority interact less with their peers, a natural question to ask is whether the exclusion of women is driven by women's preference for female peers or men's preference for male peers. Put differently, I ask, is it because women want to interact more with women or because men tend

[14] Note that I only use data from the 2019 cohort for analysis in this section, because the endline survey in 2018 did not include the peer-to-peer questions.

[15] In Appendix Figures B4 and B5, I show suggestive evidence on whether minority status affects interaction with students from out of the assigned group. The analysis is based on voluntary nominations of out-of-group study mates in the endline survey, but the response rate is low. Overall, I find that women in female-minority groups report a larger fraction of study mates from outside group, and these self-selected study mates are more likely to be female. The findings suggest that women assigned to female-minority groups still have a demand for peers, especially female peers. As the randomly assigned group can not meet the demand, they have to make some effort to find their own peers from out of the group.

to interact more with men? I test this by looking at how students' interaction with peers and evaluation of peers depend on peer gender. I use the following regression model, which utilizes within-individual cross-peer variations in pairwise relations:

$$Y_i^p = \alpha + \beta \cdot \text{Female}^p + \delta_i + \varepsilon_i^p \qquad p \in \{1, 2, 3\}. \tag{2}$$

The outcome variable, Y_i^p, is the reported relationship between student i and each of his/her three peers, $p \in \{1, 2, 3\}$. I consider academic and social interaction between i and p, as well as i's willingness to pay for p.[16] The variable of interest, $\text{Female}^p = \{0, 1\}$, indicates whether peer p is female or male. δ_i controls for individual fixed effects. I cluster standard errors at the individual level (results are very similar when standard errors are clustered at the group level). Again, I separate the analysis by gender. To use the variation in peer gender, I only include women in gender-balanced and male-minority groups when analyzing women's preferences. Similarly, I only include men in gender-balanced and female-minority groups.

Table 7 presents the results separately for women and men. Panel A examines the intensity of interaction with peers and Panel B looks at the willingness to pay for peers (the dependent variables are standardized). Results show that women's interaction with peers and evaluation of peers do not significantly depend on peers' gender. In contrast, men have a salient preference for same-gender peers. Given three randomly assigned study mates, men interact much more frequently with male peers than female peers, and they are willing to pay significantly more for male peers. The pattern is similar for gender-balanced and female-minority groups. The results suggest that women's exclusion is driven by men preferring to interact with other men rather than women's preferences for female peers.[17]

4.4 Mediation Analysis

The previous results show that minority status raises women's dropout and lowers their educational expectations and interaction with peers. In this section, I present suggestive evidence on

[16] As mentioned in Section 2, in the endline survey, students report their interaction with each peer and the hypothetical willingness to pay (WTP) for the peer, that is, how much they are willing to pay to have that peer as their study mate again. They report the WTP by choosing any amount between 0 and 100 CHF.

[17] Note that students report their willingness to pay (WTP) only for peers with whom they have interacted. If a student reported that he/she never interacted with a peer, I replace the missing value with zero. In Appendix Table B10, I show that the results are very similar quantitatively and qualitatively if I only include pairwise observations with non-missing values of the WTP.

educational expectations and peer-to-peer interaction as two potential mediators of the impact of minority status on dropout. Figure 11 shows the overall correlations between dropout and the two potential mediators. I find that students with higher educational expectation at endline or who have interacted more frequently with peers are less likely to drop out. A one standard deviation increase in educational expectations is related to a 3.7 percentage point decreases in the dropout rate. Similarly, a one standard deviation increase in peer-to-peer interaction translates into a 2.8 percentage point decrease in dropout. Both correlations are significant.

In Appendix Section C.2, I further show: (1) how the inclusion of mediators changes the estimated impact of minority status on dropout, and (2) what fractions of the impact are explained by the mediators based on the approach of Heckman & Pinto (2015). I find that the effect of minority status on dropout decreases when the endline educational expectation and peer-to-peer interaction are controlled for (see Appendix Table C2). I also find that each mediator explains more than 10% of the effect of minority status on women's dropout (see Appendix Figure C2). Note that this mediation analysis should be taken with caution. In the Appendix, I discuss the assumptions and limits fully.

5 Robustness, Heterogeneity, and Discussions

5.1 Robustness

Endogenous Attrition. Note that the results on the two mechanisms use data from the endline survey, for which the response rate is about 70%. Although I find that minority status does not significantly predict survey response, the attrition rate can be correlated with observed or unobserved characteristics. To erase concerns about non-random attrition, I report the estimated effects of minority status on educational expectation and interaction with peers using Lee bounds (Lee, 2009), with or without using Inverse Probability Weighted (IPW) regressions. Appendix Table B11 shows that the lower and upper bounds are within the 95% confidence intervals of the OLS estimations, suggesting that endogenous attrition is not driving the results.

Multiple Test. Appendix Section C.1 shows that most results survive the multiple test correction. After the Romano-Wolf correction (Romano & Wolf, 2005a,b), which controls the familywise error rate (FWER), all significant coefficients of minority status (the impacts on

women) remain statistically significant. For robustness, I also use the procedure of Benjamini & Yekutieli (2001), which controls the false discovery rate (FDR). Again, most significant coefficients of minority status remain significant.

5.2 Heterogeneous Effects: Which Women Drop Out?

From a policy perspective, it is important to know what kind of women are more susceptible to the impact of minority status. Losing different types of women have different implications. For example, it implies a loss of talent for the field that women leave if high-skilled women drop out, while losing low-skilled women may seem less concerning. A study field becomes less diversified if women who stay display similar characteristics as men, while those who leave exhibit typical female characteristics. I analyze the heterogeneous effects of minority status along two dimensions: (1) academic ability as measured by high school math grade, and (2) gender typicality—how typical a student's personality traits, socioeconomic preferences, and gender attitudes are female or male.

Heterogeneity by Math Ability. Students' high school math grade serves as a proxy for academic ability because the math grade is the strongest predictor of performance in this economics course. Appendix Figure B6 plots how different individual characteristics predict the final grade. Among all baseline characteristics, high school math achievement has the largest predictive effect and explanatory power on the final grade. I divide students into two sub-groups: those with below-median and above-median math grades. Figure 12 plots the estimated coefficient of minority status for each subgroup of students.

I find that the impact of minority status on women is driven by those with high math grades, suggesting that economics as a male-dominated field is loosing talented women. When assigned as the minority gender, they become more likely to drop out, have lower expectations, and interact much less with their peers in the group. In contrast, minority status does not significantly affect the outcomes of women with below-median math grades. Men with low math achievement also become more likely to drop out when assigned as the minority gender. There are two potential explanations for this striking effect. First, women with higher math achievement may have better outside options. They can switch to another study field and perform well. Second, high math achievement is correlated with higher expectations and aspirations at baseline. These

higher expectations and aspirations potentially translate to stronger disappointment and cause dropout.

Heterogeneity by Gender Typicality. Next, I examine the heterogeneous impact by gender typicality: whether a student's social characteristics are typically female or male. As stressed by Lorber et al. (1991) and Butler et al. (2004), gender is a social construct and the binaries of women versus men omit important variations within a gender. A women can display social characteristics that are typical of men; a man can also display characteristics that are typical of women. I define gender typicality in two steps. First, I regress an indicator for female on students' baseline Big-Five personality traits, competitiveness, risk preference, patience, and gender attitudes, and derive the predicted probability of being female. Second, I standardize the predicted probability and call it the female typicality index. Women or men with an above-zero female typicality index are categorized as *typically female*, and those with a below-zero female typicality index are categorized as *typically male*.[18]

Figure 13 presents the estimated effects of minority status on dropout, educational expectation, and interaction with peers by student gender and gender typicality. Results show that the impact of minority status on women are mainly driven by women with typically female characteristics. In contrast, minority status does not significantly affect the outcomes of women with typically male characteristics. The finding suggests that due to the unbalanced gender ratio, women who have traditionally female personalities and preferences are more likely to drop out of economics, and those who stay are more like their male peers.

5.3 Discussion

Other Potential Mechanisms? In additional to educational expectations and interaction with peers, minority status may also affect educational persistence through other channels. For example, students who experience minority status may invest less in the course and consequently become less prepared for the final exam. They may also develop less interest in the course or the selected major, and drop out as a result. More generally, minority status in the study group

[18] More specifically, to avoid mechanical correlations between own gender and the predicted gender, I exclude the student himself/herself when predicting the female typicality index. Appendix Figure B8 plots the distribution of the female typicality index by gender. Risk preference and patience are measured with single-item questions asking how risk-taking and patient a student perceives himself/herself.

for a specific course may have broad impact on students' self-esteem or self-efficacy.

I test alternative mechanisms by regressing other outcome variables measured in the survey on minority status. Figure 14 plots the estimated coefficients of minority status. I find no evidence suggesting that other mechanisms are at play. Women in the minority do not report fewer study hours for the course, a lower share of lectures or tutorials attended, or less satisfaction with the course. Also, minority status in the study group does not seem to affect their general self-efficacy, self-esteem, or beliefs about their relative math ability and diligence level.

Are Study Groups Beneficial? To this point, all results apply to students in randomly assigned study groups. One question that remains unanswered is whether study groups themselves are beneficial. This paper cannot provide causal evidence on that because the study groups are offered on a voluntary basis and students' select into study groups correlate with observable or observably characteristics.[19] I test whether the group participation decision is correlated with observed student characteristics in Appendix Figure B10. Although women are 7 percentage points more likely to sign up for study groups compared to men, the women who sign up are not significantly different from women who do not sign up in terms of high school background and personality traits.

With the potential selection bias in mind, I analyze how group participation correlates with relevant outcomes. Table 8 examines whether students registered for study groups differ from those without groups in terms of dropout behavior. Results show that women in study groups are about 4 percentage points less likely to drop out of the course, even after controlling for high school background, the Big-Five personality traits, competitiveness, gender bias, and the baseline expectation. Men with study groups are not significantly different from those without study groups. Students with a study group also have higher educational expectations at endline than students without study groups, although the difference is not statistically significant.

Based on the suggestive evidence that study groups are overall beneficial for students, especially for women, I continue by asking what the optimal group assignment should be in order to lower women's dropout rate. I conduct a simple counterfactual analysis in my experimental setting. By fixing the sample size and the observed dropout rate for each type of group gen-

[19]The overall sign-up rate for study groups is about 30%. Study groups are offered voluntarily not only due to ethical concerns but also because students who want to interact with others are the population of higher policy interest.

der composition, I abolish female-minority groups and assign students into gender-balanced, male-minority, or male-only groups. For each counterfactual group assignment, I calculate the average female and male dropout rate. As Appendix Figure B11 shows, in the optimal counterfactual scenario, 620 students are allocated into 86 gender-balanced groups, 29 male-minority groups, and 40 male-only groups. Such a combination of study groups can decrease women's dropout rate by 2 percentage points and keep men's dropout rate the same.

External Validity. A final concern about this paper regards the external validity of the results. This paper uses a field experiment in an introductory economics course at university. The design has the following features that adds to its external validity. First, around 35% of the students in the course are female, so it represents a typical male-dominated setting. Second, the course is compulsory for students majoring in economics, business, and computer science, and those enrolled in related minors. That means the results do not only apply to economics. Third, small peer groups are very common in education and the workplace, and first semester in college is a natural period when students form new friendships. In my setting, before conducting the experiment, about 50% of students reported that they studied with others for the course and more than 65% the study mates were new peers that they met after coming to college. Finally, a size of 4 is not uncommon for small groups. Based on self-reported social network data, most endogenously formed networks among students in the course were of sizes 3–5.

The study has two limitations regarding its generalizability. First, due to the group size, this paper can only examine one type of minority status which is when the share of women is 1/4. If the group size is 5, I would be able to compare the effects of two scenarios of minority status on women: when the share of women is 1/5 or 2/5. Suppose that women's educational persistence in male-dominated settings increases with the share of female peers—as suggested by findings of Griffith (2010), Huntington-Klein & Rose (2018), and Bostwick & Weinberg (2018), then the impact of minority status on women's dropout in the 1/4 scenario will be between the 1/5 and 2/5 scenarios.

The second limitation relates to students' selection into study groups, that is, their voluntary participation in the experiment. It raises the concern that the average treatment effects (ATE) of minority status estimated by the experimental sample may not be externally valid for the population of students (especially women) as a whole. However, it is less of a concern

because the observable characteristics are balanced between women registered for study groups and women without study groups (see Appendix Figure B10), and students who want to interact with others arguably have higher policy interest. That being said, it is still valuable to check how the ATE estimations are different after correcting for the participation bias. In Appendix Table B12, I present the estimated impacts of minority status on women's dropout and educational expectation after correcting for participation on observables and unobservables, using the approach of Andrews & Oster (2019). Overall, I find that the impacts of minority status on women are robust. Even after considering a wide range of observables such as student major, course-retaking status, experiment year, high school background, and personality traits, the unobservables would have to be more important than the observables ($\Phi > 2$) in order to deliver an ATE of zero in the population.

6 Conclusion

This paper uses a field experiment in an economics course to show that women's minority status can drive them out of male-dominated fields. Results show that women assigned to be in the minority become 10 percentage points more likely to drop out of the course and their major or minor study programs. The finding suggests the existence of a "minority trap", meaning that underrepresented groups are trapped in a vicious cycle of being in the minority. I present suggestive evidence on two mechanisms of the minority trap: lower confidence about future achievement and lower integration into the social environment.

Heterogeneous analyses show that the effects of minority status on women are driven by those with above-median math achievement rather than those with low math achievement, suggesting that economics as a male-dominated field is losing talented women. Meanwhile, women with typically-female characteristics, like personality traits and gender attitudes, are more susceptible to the impact of minority status than those with typically-male characteristics. Taken together, the findings suggest that the unbalanced gender ratio is causing a loss of both talent and diversity for economics or other male-dominated fields.

Given the current efforts to promote gender equality and close gender gaps in male-dominated fields, this paper has important policy implications. To break the minority trap and retain more women in these fields, educational institutions should avoid female-minority groups whenever

students are assigned to natural peer groups like study groups, joint projects, laboratories, or offices. A back-of-the-envelope calculation shows that the female dropout rate in my experiment could decrease by 2 percentage points if female-minority groups were abolished. Working in teams is common practice in a wide range of educational and workplace settings. Over time it has become much more prevalent in firms and organizations, as well as in academic research (Devine et al., 1999; Lazear & Shaw, 2007; Rath & Wohlrabe, 2016). Compared to previous studies examining cohort or class gender composition which is difficult to manipulate in practice, this paper's focus on small peer groups offers more scope for policy interventions. By introducing and designing small groups in an appropriate way, male-dominated majors or professions can hopefully raise women's integration to the social environment, boost their self-confidence, and retain more women in the field.

References

Alan, S., Boneva, T., & Ertac, S. (2019). Ever failed, try again, succeed better: Results from a randomized educational intervention on grit. *The Quarterly Journal of Economics*, 134(3), 1121–1162.

Andrews, I. & Oster, E. (2019). A simple approximation for evaluating external validity bias. *Economics Letters*, 178, 58–62.

Anelli, M. & Peri, G. (2019). The effects of high school peers' gender on college major, college performance and income. *The Economic Journal*, 129(618), 553–602.

Anil, B., Guner, D., Delibasi, T., & Uysal, G. (2017). Does classroom gender composition affect school dropout? *Economic Research Forum Working Paper 1084*.

Arulampalam, W., Booth, A. L., & Bryan, M. L. (2007). Is there a glass ceiling over Europe? Exploring the gender pay gap across the wage distribution. *Industrial and Labor Relations Review*, (pp. 163–186).

Autor, D. H. (2014). Skills, education, and the rise of earnings inequality among the "other 99 percent". *Science*, 344(6186), 843–851.

Avilova, T. & Goldin, C. (2018). What can UWE do for economics? In *AEA Papers and Proceedings*, volume 108 (pp. 186–90).

Bengtsson, C., Persson, M., & Willenhag, P. (2005). Gender and overconfidence. *Economics Letters*, 86(2), 199–203.

Benjamini, Y., Krieger, A. M., & Yekutieli, D. (2006). Adaptive linear step-up procedures that control the false discovery rate. *Biometrika*, 93(3), 491–507.

Benjamini, Y. & Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics*, (pp. 1165–1188).

Bettinger, E. P. & Long, B. T. (2005). Do faculty serve as role models? The impact of instructor gender on female students. *American Economic Review*, 95(2), 152–157.

Bian, L., Leslie, S.-J., & Cimpian, A. (2017). Gender stereotypes about intellectual ability emerge early and influence children's interests. *Science*, 355(6323), 389–391.

Blau, F. D., Currie, J. M., Croson, R. T., & Ginther, D. K. (2010). Can mentoring help female assistant professors? Interim results from a randomized trial. *American Economic Review*, 100(2), 348–52.

Blau, F. D. & Kahn, L. M. (2017). The gender wage gap: Extent, trends, and explanations. *Journal of Economic Literature*, 55(3), 789–865.

Booij, A. S., Leuven, E., & Oosterbeek, H. (2017). Ability peer effects in university: Evidence from a randomized experiment. *The Review of Economic Studies*, 84(2), 547–578.

Booth, A. L., Cardona-Sosa, L., & Nolen, P. (2018). Do single-sex classes affect academic achievement? An experiment in a coeducational university. *Journal of Public Economics*, 168, 109–126.

Born, A., Ranehill, E., & Sandberg, A. (2020). Gender and willingness to lead: Does the gender composition of teams matter? *Review of Economics and Statistics*, forthcoming.

Bostwick, V. K. & Weinberg, B. A. (2018). Nevertheless she persisted? Gender peer effects in doctoral STEM programs. *National Bureau of Economic Research Working Paper 25028*.

Brenøe, A. A. & Zölitz, U. (2020). Exposure to more female peers widens the gender gap in stem participation. *Journal of Labor Economics*, 38(4), 000–000.

Bursztyn, L., Egorov, G., & Jensen, R. (2019). Cool to be smart or smart to be cool? Understanding peer pressure in education. *The Review of Economic Studies*, 86(4), 1487–1526.

Buser, T., Niederle, M., & Oosterbeek, H. (2014). Gender, competitiveness, and career choices. *The Quarterly Journal of Economics*, 129(3), 1409–1447.

Buser, T., Niederle, M., & Oosterbeek, H. (2020). Can competitiveness predict education and labor market outcomes? Evidence from incentivized choice and survey measures. *Unpublished manuscript*.

Butler, J. et al. (2004). *Undoing gender*. Psychology Press.

Cadinu, M., Maass, A., Frigerio, S., Impagliazzo, L., & Latinotti, S. (2003). Stereotype threat: The effect of expectancy on performance. *European Journal of Social Psychology*, 33(2), 267–285.

Carnes, M., Devine, P. G., Manwell, L. B., Byars-Winston, A., Fine, E., Ford, C. E., Forscher, P., Isaac, C., Kaatz, A., Magua, W., et al. (2015). Effect of an intervention to break the gender bias habit for faculty at one institution: A cluster randomized, controlled trial. *Academic Medicine: Journal of the Association of American Medical Colleges*, 90(2), 221.

Carrell, S. E., Fullerton, R. L., & West, J. E. (2009). Does your cohort matter? Measuring peer effects in college achievement. *Journal of Labor Economics*, 27(3), 439–464.

Carrell, S. E., Hoekstra, M., & Kuka, E. (2018). The long-run effects of disruptive peers. *American Economic Review*, 108(11), 3377–3415.

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

Chen, J. & Houser, D. (2019). When are women willing to lead? The effect of team gender composition and gendered tasks. *The Leadership Quarterly*, 30(6), 101340.

Cheng, A., Kopotic, K., & Zamarro, G. (2017). Can parents' growth mindset and role modelling address STEM gender gaps? *EDRE Working Paper 2017-07*.

Coffman, K. B. (2014). Evidence on self-stereotyping and the contribution of ideas. *The Quarterly Journal of Economics*, 129(4), 1625–1660.

Cohen, J., Dupas, P., et al. (2010). Free distribution or cost-sharing? Evidence from a randomized malaria prevention experiment. *Quarterly Journal of Economics*, 125(1), 1.

Croson, R. & Gneezy, U. (2009). Gender differences in preferences. *Journal of Economic Literature*, 47(2), 448–74.

Delfino, A. (2020). Breaking gender barriers: Bringing men into the pink-collar jobs of the future. *Unpublished manuscript*.

Devine, D. J., Clayton, L. D., Philips, J. L., Dunford, B. B., & Melner, S. B. (1999). Teams in organizations: Prevalence, characteristics, and effectiveness. *Small Group Research*, 30(6), 678–711.

Dryler, H. (1998). Parental role models, gender and educational choice. *British Journal of Sociology*, (pp. 375–398).

Falk, A. & Heckman, J. J. (2009). Lab experiments are a major source of knowledge in the social sciences. *Science*, 326(5952), 535–538.

Ficano, C. C. (2012). Peer effects in college academic outcomes–Gender matters! *Economics of Education Review*, 31(6), 1102–1115.

Fischer, S. (2017). The downside of good peers: How classroom composition differentially affects men's and women's STEM persistence. *Labour Economics*, 46, 211–226.

Fisher, R. A. (1935). *The Design of Experiments*. Edinburgh: Oliver & Boyd.

Fujiwara, T. & Wantchekon, L. (2013). Can informed public deliberation overcome clientelism? Experimental evidence from Benin. *American Economic Journal: Applied Economics*, 5(4), 241–55.

Gneezy, U., List, J. A., Livingston, J. A., Qin, X., Sadoff, S., & Xu, Y. (2019). Measuring success in education: the role of effort on the test itself. *American Economic Review: Insights*, 1(3), 291–308.

Golsteyn, B., Non, A., & Zölitz, U. (2020). The impact of peer personality on academic achievement. *Journal of Political Economy*, forthcoming.

Gong, J., Lu, Y., & Song, H. (2019). Gender peer effects on students' academic and noncognitive outcomes: Evidence and mechanisms. *Journal of Human Resources*, (pp. 0918–9736R2).

Griffith, A. L. (2010). Persistence of women and minorities in STEM field majors: Is it the school that matters? *Economics of Education Review*, 29(6), 911–922.

Heckman, J. J. & Pinto, R. (2015). Econometric mediation analyses: Identifying the sources of treatment effects from experimentally estimated production technologies with unmeasured and mismeasured inputs. *Econometric Reviews*, 34(1-2), 6–31.

Herring, C. (2009). Does diversity pay?: Race, gender, and the business case for diversity. *American Sociological Review*, 74(2), 208–224.

Hill, A. J. (2015). The girl next door: The effect of opposite gender friends on high school achievement. *American Economic Journal: Applied Economics*, 7(3), 147–77.

Hill, A. J. (2017). The positive influence of female college students on their male peers. *Labour Economics*, 44, 151–160.

Hoogendoorn, S., Oosterbeek, H., & Van Praag, M. (2013). The impact of gender diversity on the performance of business teams: Evidence from a field experiment. *Management Science*, 59(7), 1514–1528.

Hsieh, C.-T., Hurst, E., Jones, C. I., & Klenow, P. J. (2019). The allocation of talent and us economic growth. *Econometrica*, 87(5), 1439–1474.

Huntington-Klein, N. & Rose, E. (2018). Gender peer effects in a predominantly male environment: Evidence from west point. In *AEA Papers and Proceedings*, volume 108 (pp. 392–95).

Ibarra, H. (1992). Homophily and differential returns: Sex differences in network structure and access in an advertising firm. *Administrative Science Quarterly*, (pp. 422–447).

Imbens, G. W. & Wooldridge, J. M. (2009). Recent developments in the econometrics of program evaluation. *Journal of economic literature*, 47(1), 5–86.

Kamas, L. & Preston, A. (2012). The importance of being confident; gender, career choice, and willingness to compete. *Journal of Economic Behavior & Organization*, 83(1), 82–97.

Kanter, R. M. (1977). *Men and women of the corporation*. Basic Books.

Lavy, V. (2008). Do gender stereotypes reduce girls' or boys' human capital outcomes? Evidence from a natural experiment. *Journal of Public Economics*, 92(10-11), 2083–2105.

Lavy, V. & Schlosser, A. (2011). Mechanisms and impacts of gender peer effects at school. *American Economic Journal: Applied Economics*, 3(2), 1–33.

Laws, J. L. (1975). The psychology of tokenism: An analysis. *Sex Roles*, 1(1), 51–67.

Lazear, E. P. & Shaw, K. L. (2007). Personnel economics: The economist's view of human resources. *Journal of Economic Perspectives*, 21(4), 91–114.

Lee, D. S. (2009). Training, wages, and sample selection: Estimating sharp bounds on treatment effects. *The Review of Economic Studies*, 76(3), 1071–1102.

Levitt, S. D. & List, J. A. (2007). What do laboratory experiments measuring social preferences reveal about the real world? *Journal of Economic Perspectives*, 21(2), 153–174.

Li, Q., Zang, W., & An, L. (2013). Peer effects and school dropout in rural China. *China Economic Review*, 27, 238–248.

Lindqvist, E., Östling, R., & Cesarini, D. (2020). Long-run effects of lottery wealth on psychological well-being. *The Review of Economic Studies*, 87(6), 2703–2726.

Lorber, J., Farrell, S. A., et al. (1991). *The social construction of gender*. Sage Newbury Park, CA.

Lundberg, S. (2018). The 2018 report of the committee on the status of women in the economics profession. *American Economic Association*.

Mayer, A. & Puller, S. L. (2008). The old boy (and girl) network: Social network formation on university campuses. *Journal of Public Economics*, 92(1-2), 329–347.

Mengel, F. (2020). Gender differences in networking. *The Economic Journal*, 130(630), 1842–1873.

Newson, R. & Team, A. S. (2003). Multiple-test procedures and smile plots. *The Stata Journal*, 3(2), 109–132.

Nguyen, H.-H. D. & Ryan, A. M. (2008). Does stereotype threat affect test performance of minorities and women? A meta-analysis of experimental evidence. *Journal of Applied Psychology*, 93(6), 1314.

Niederle, M. & Vesterlund, L. (2007). Do women shy away from competition? Do men compete too much? *The Quarterly Journal of Economics*, 122(3), 1067–1101.

Niederle, M. & Vesterlund, L. (2011). Gender and competition. *Annual Review of Economics*, 3(1), 601–630.

Oosterbeek, H. & Van Ewijk, R. (2014). Gender peer effects in university: Evidence from a randomized experiment. *Economics of Education Review*, 38, 51–63.

Porter, C. & Serra, D. (2019). Gender differences in the choice of major: The importance of female role models. *American Economic Journal: Applied Economics*.

Rath, K. & Wohlrabe, K. (2016). Recent trends in co-authorship in economics: evidence from RePEc. *Applied Economics Letters*, 23(12), 897–902.

Reuben, E., Wiswall, M., & Zafar, B. (2017). Preferences and biases in educational choices and labour market expectations: Shrinking the black box of gender. *The Economic Journal*, 127(604), 2153–2186.

Romano, J. P. & Wolf, M. (2005a). Exact and approximate stepdown methods for multiple hypothesis testing. *Journal of the American Statistical Association*, 100(469), 94–108.

Romano, J. P. & Wolf, M. (2005b). Stepwise multiple testing as formalized data snooping. *Econometrica*, 73(4), 1237–1282.

Rosenberg, M., Schooler, C., Schoenbach, C., & Rosenberg, F. (1995). Global self-esteem and specific self-esteem: Different concepts, different outcomes. *American Sociological Review*, (pp. 141–156).

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

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

Schwarzer, R., Jerusalem, M., Weinman, J., Wright, S., & Johnston, M. (1995). Measures in health psychology: A user's portfolio. Causal and control beliefs. *Windsor, UK: Nfer-Nelson*, (pp. 35–7).

Shrum, W., Cheek Jr, N. H., & MacD, S. (1988). Friendship in school: Gender and racial homophily. *Sociology of Education*, (pp. 227–239).

Spencer, S. J., Logel, C., & Davies, P. G. (2016). Stereotype threat. *Annual Review of Psychology*, 67, 415–437.

Stinebrickner, T. & Stinebrickner, R. (2012). Learning about academic ability and the college dropout decision. *Journal of Labor Economics*, 30(4), 707–748.

Stoddard, O., Jessica, P., & Karpowitz, C. (2020). Strength in numbers: A field experiment in gender, influence, and group dynamics. *Unpublished manuscript*.

Tinto, V. (1975). Dropout from higher education: A theoretical synthesis of recent research. *Review of Educational Research*, 45(1), 89–125.

Wiswall, M. & Zafar, B. (2015). Determinants of college major choice: Identification using an information experiment. *The Review of Economic Studies*, 82(2), 791–824.

Wiswall, M. & Zafar, B. (2018). Preference for the workplace, investment in human capital, and gender. *The Quarterly Journal of Economics*, 133(1), 457–507.

Woolley, A. W., Chabris, C. F., Pentland, A., Hashmi, N., & Malone, T. W. (2010). Evidence for a collective intelligence factor in the performance of human groups. *Science*, 330(6004), 686–688.

Young, A. (2019). Channeling Fisher: Randomization tests and the statistical insignificance of seemingly significant experimental results. *The Quarterly Journal of Economics*, 134(2), 557–598.

Zárate, R. A. (2019). Social and cognitive peer effects: Experimental evidence from selective high schools in Peru. *Unpublished manuscript*.

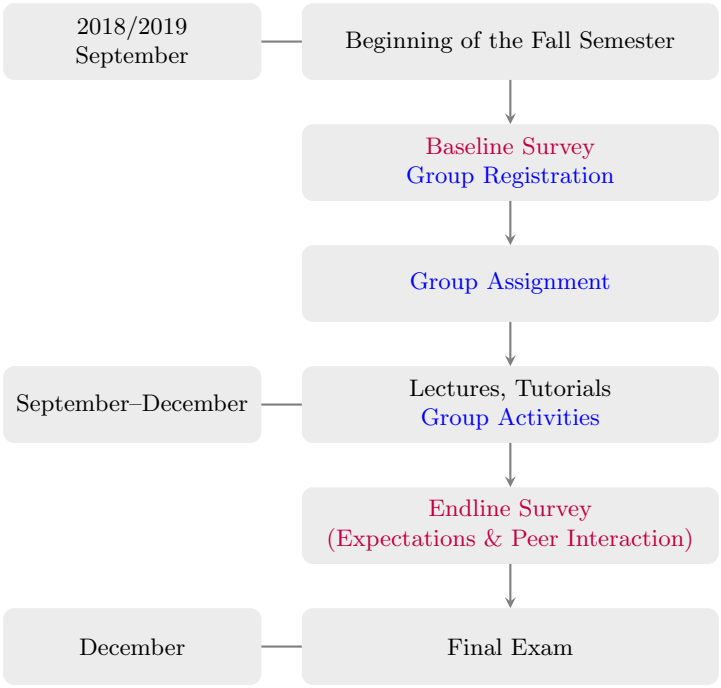
Zimmer, L. (1988). Tokenism and women in the workplace: The limits of gender-neutral theory. *Social Problems*, 35(1), 64–77.

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. & Feld, J. (2020). The effect of peer gender on major choice in business school. *Management Science*, forthcoming.

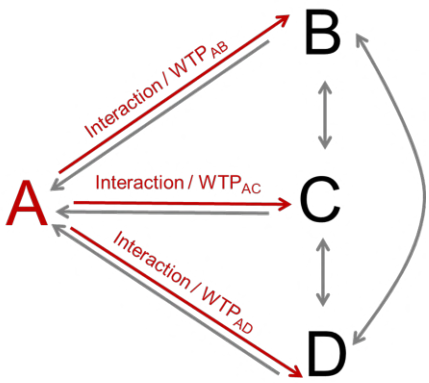
Zou, J. (2019). Understanding the peer effects of non-cognitive ability on academic outcomes. *Unpublished manuscript*.

Figure 1: The Experiment Timeline



Notes: The figure plots timeline of the experiment which took place in 2018 and 2019 following the same design. The endline survey measures two important things: expectations about future academic achievement and interaction with peers.

Figure 2: Pairwise Interaction and Willingness to Pay (WTP) Within a Study Group



Notes: The figure shows the one-to-one or pairwise relations in a study group of four. Each arrow represents a directed one-to-one relationship between two students. In the endline survey, a student reports the frequency of interaction with each of the peers in the group and the willingness to pay for each peer.

Table 1: Descriptive Statistics

	Women			Men		
	N	Mean	(SD)	N	Mean	(SD)
Panel A: Baseline Characteristics						
High School Math Grade	235	4.54	(0.78)	319	4.52	(0.86)
High School Language Grade	235	4.91	(0.54)	319	4.65	(0.57)
High School in German	235	0.85	(0.36)	319	0.86	(0.35)
Expected Grade	235	4.60	(0.49)	319	4.74	(0.52)
Expected Probability of Passing First Year	235	75.3	(16.5)	319	76.5	(19.2)
Expected Highest Degree is MA	235	0.83	(0.38)	319	0.76	(0.43)
Expected Highest Degree is PhD	235	0.09	(0.28)	319	0.12	(0.33)
Overall Educational Expectation	235	-0.10	(0.93)	319	0.06	(1.05)
Panel B: Endline Characteristics						
Expected Grade	187	4.35	(0.66)	247	4.53	(0.61)
Expected Probability of Passing First Year	187	67.1	(19.5)	247	72.3	(20.9)
Expected Highest Degree is MA	187	0.78	(0.42)	247	0.76	(0.43)
Expected Highest Degree is PhD	187	0.08	(0.27)	247	0.10	(0.30)
Overall Educational Expectation	187	-0.13	(0.96)	247	0.10	(1.02)
Academic Interaction (2019)	146	0.04	(1.00)	202	-0.03	(1.00)
Social Interaction (2019)	146	-0.08	(0.99)	202	0.06	(1.00)
Overall Peer Interaction (2019)	146	-0.03	(0.98)	202	0.02	(1.01)
Panel C: Educational Outcomes						
Dropout	259	0.13	(0.33)	361	0.09	(0.29)
Final Grade	259	3.70	(1.30)	361	3.97	(1.22)

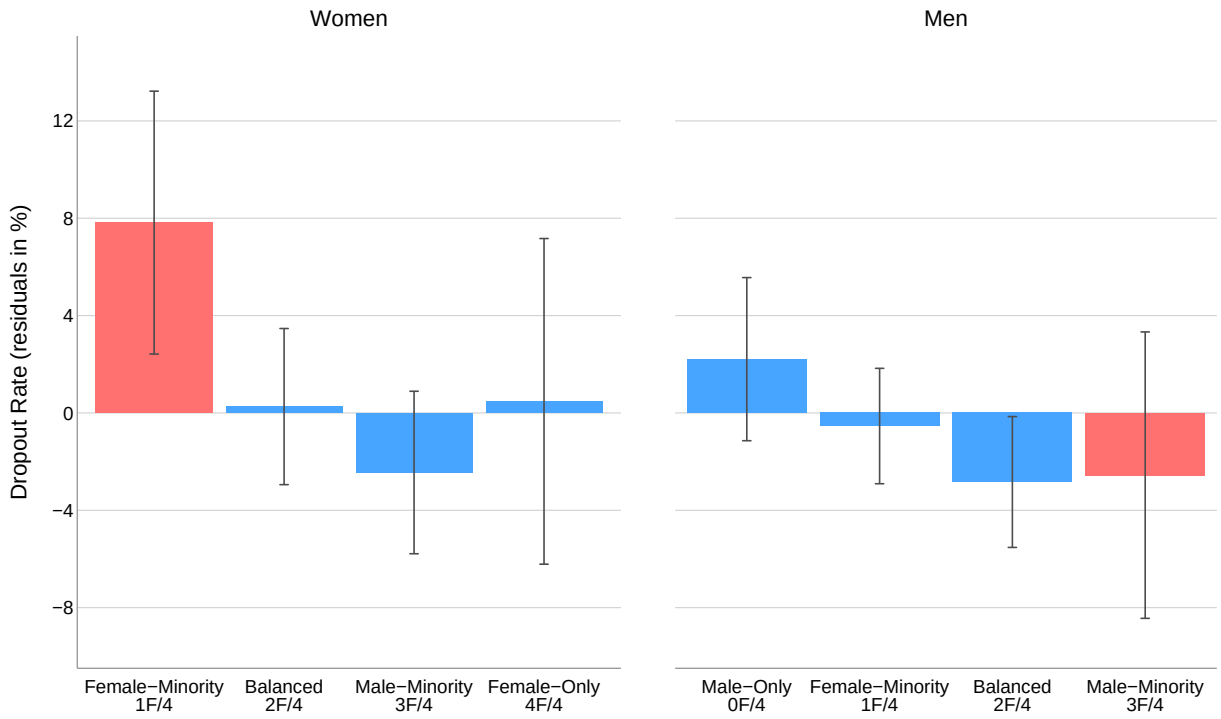
Notes: *Overall educational expectation* is a combined index for three dimensions of expectations: (1) the expected grade for the course, (2) the expected probability of passing the first year at the university, and (3) the highest academic degree expected to obtain. The index is created as follows: standardize the expectation for each dimension, take the mean of standardized values, and finally standardize the mean. *Academic/Social interaction* is the standardized average academic or social interaction with assigned peers. *Overall interaction* is the combined index for both academic and social interaction. Variables using administrative data have the full sample size, that is, 259 for women and 361 for men. For other variables, the number of observations depends on survey response.

Table 2: Does Minority Status Predict Baseline Characteristics and Survey Response?

	_b[Minority Status]	[RI p-val]	(CL S.E.)
Panel A: Baseline Characteristics			
High School Math Grade	-0.034	[0.731]	(0.094)
High School Language Grade	-0.037	[0.627]	(0.065)
High School in German	-0.028	[0.536]	(0.045)
Conscientiousness	0.123	[0.439]	(0.120)
Extraversion	-0.130	[0.317]	(0.126)
Agreeableness	0.019	[0.890]	(0.114)
Openness	0.066	[0.602]	(0.128)
Neuroticism	0.041	[0.798]	(0.140)
Course Retaking	-0.028	[0.405]	(0.030)
Baseline Educational Expectation	0.009	[0.929]	(0.101)
Panel B: Survey Response			
Baseline Survey Completion	-0.007	[0.867]	(0.040)
Endline Survey Completion	-0.020	[0.735]	(0.057)

Notes: The table tests whether minority status predicts baseline characteristics and survey response. The response rate is 88% for the baseline survey and 66% for the endline survey. I regress each characteristic on the minority indicator, controlling for gender, and the fixed effects of study program and experiment year. See Appendix Table B2 for the balance test separately for women and men. Randomization Inference (RI) p-values are in brackets, and clustered standard errors are in parentheses. $*p < .1, **p < .05, ***p < .01$.

Figure 3: Group Gender Composition and Dropout



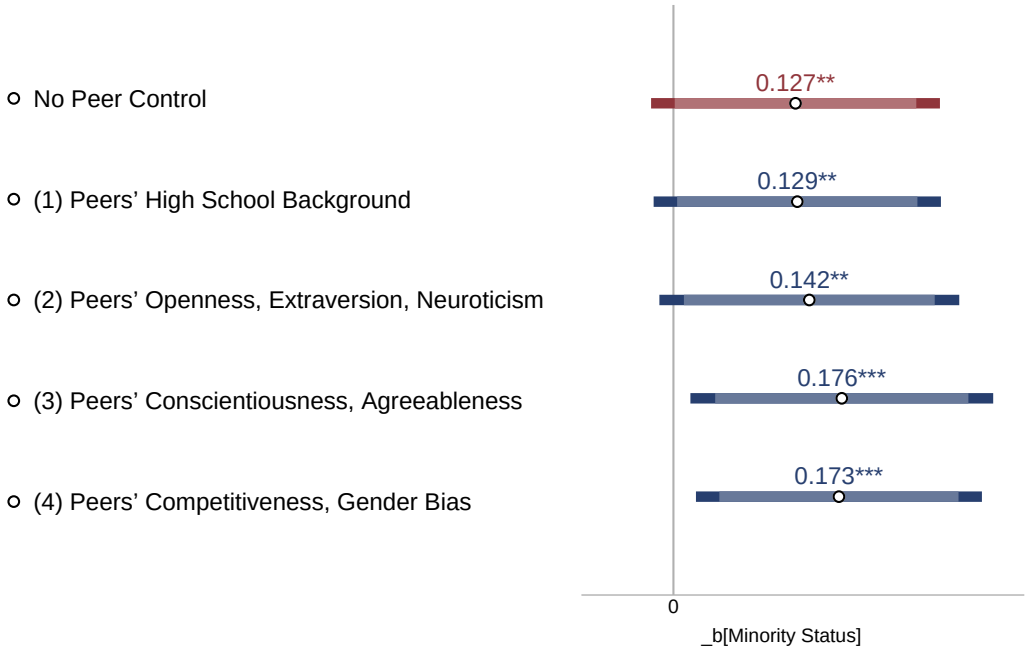
Notes: The figure shows the residuals of dropout rate, after controlling for assignment strata (year and study program) fixed effects, by group gender composition. The gap in residuals of dropout between women in female-minority groups and women in other groups is about 8.5 percentage points; the gap between men in male-minority groups and other men is about 2 percentage points. The raw dropout rate is 13% for women and 9% for men (see Appendix Figure B2 for the distribution of the dropout rate by group gender composition). Error bars indicate standard error of the mean.

Table 3: The Impact of Minority Status on Dropout

DV: Dropout	(1)	(2)	(3)	(4)
Minority Status	0.094	0.107	0.101	0.100
[RI p-val]	[0.016]**	[0.005]***	[0.007]***	[0.008]***
(CL S.E.)	(0.056)*	(0.059)*	(0.059)*	(0.059)*
Minority Status × Male	-0.119	-0.139	-0.127	-0.130
[RI p-val]	[0.049]**	[0.026]**	[0.041]**	[0.039]**
(CL S.E.)	(0.085)	(0.086)	(0.088)	(0.090)
Basic controls & gender	Y	Y	Y	Y
High school controls	N	N	Y	Y
Personality traits	N	N	N	Y
Observations	620	554	554	554
R-squared	0.027	0.033	0.053	0.068

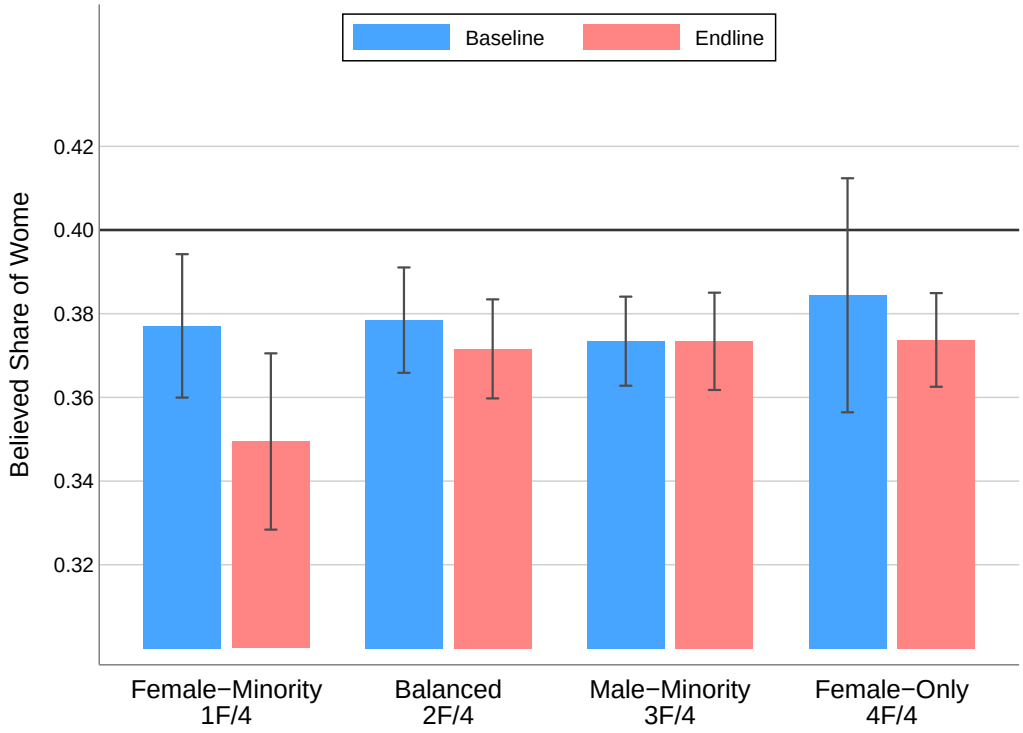
Notes: The table presents the impact of minority status on dropout, using OLS regressions. The dependent variable is an indicator for dropout. Basic controls include student major, course-retaking status, and year fixed effects. High school controls include grades for math and first language, and whether the first language is German. Personality traits refer to the Big-Five personality traits. Results are virtually the same when I split the sample by gender and estimate the impact of minority status separately for women and men (see Appendix Table B3). Randomization Inference (RI) p-values are in brackets, and clustered standard errors are in parentheses. $*p < .1, **p < .05, ***p < .01$.

Figure 4: The Impact of Minority Status on Female Dropout With/Without Peer Controls



Notes: The figure plots the estimated impact of minority status on women's dropout when controlling for different peer traits. The top row includes no peer traits; then I gradually include each of the following sets of peer traits: (1) high school math and language grades and whether the first language is German, (2) openness, extraversion, and neuroticism, (3) agreeableness and conscientiousness, and (4) competitiveness and gender bias. Competitiveness is measured with the sing-item question from Buser et al. (2020). Gender bias is measured with questions on gender role attitudes from the World Values Survey (wave 6). All regressions control for women's own major, retaking status, year fixed effects, high school background, the Big-Five personality traits, competitiveness, and gender bias. All regressions only include female observations with non-missing values, so that the change in estimated effects is not driven by changes in the estimation sample. 95% and 90% confidence intervals are based on standard errors clustered at the group level; significance stars are based on randomization inference (RI) p-values. *RI p <.1, **RI p <.05, ***RI p <.01.

Figure 5: Believed Share of Women in the Course



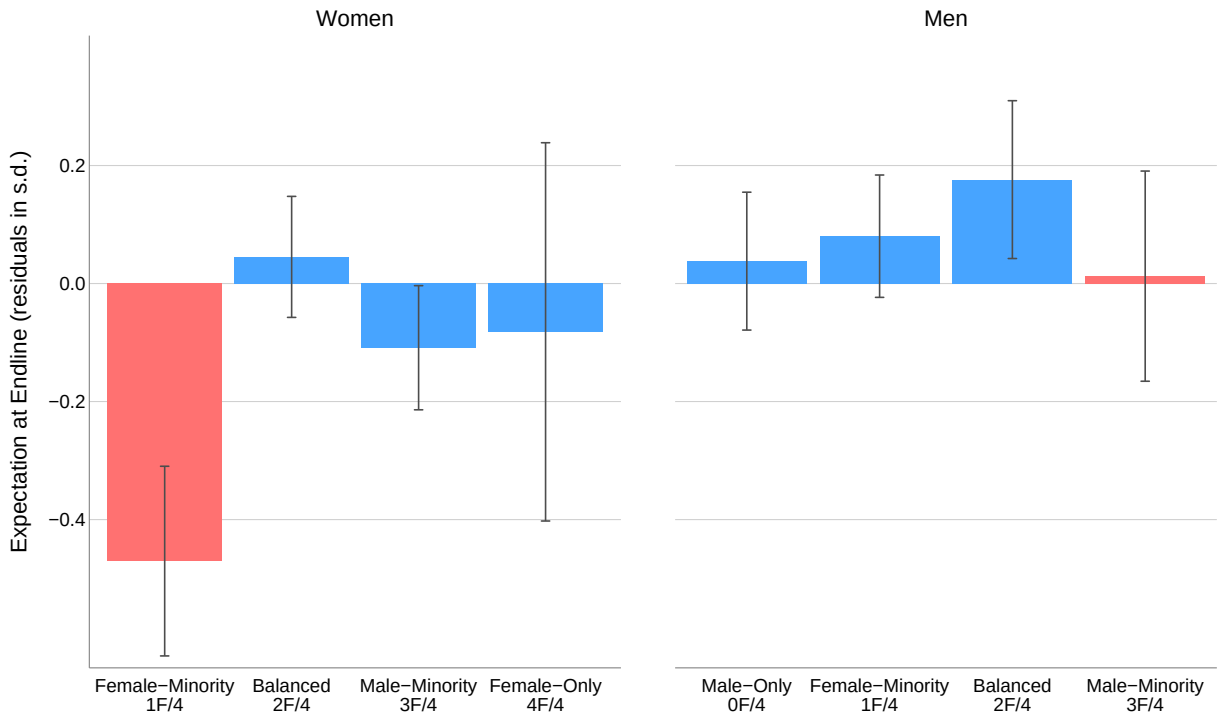
Notes: The figure shows female students' believed share of women in the economics course by group gender composition at the beginning and end of the semester in 2019. The actual share of women is 40%, as indicated by the black line. Error bars indicate the standard error of the mean. Appendix Figure B3 shows the distributional pattern of men's beliefs.

Table 4: The Impact of Minority Status on the Believed Share of Women in the Course

	(1)	(2)	(3)	(4)
DV: Believed Share of Women in the Course at Endline				
Minority Status	-0.036	-0.034	-0.036	-0.040
[RI p-val]	[0.017]**	[0.031]**	[0.019]**	[0.006]***
(CL S.E.)	(0.021)*	(0.021)	(0.022)	(0.024)
Minority Status × Male	0.093	0.091	0.086	0.078
[RI p-val]	[0.002]***	[0.002]***	[0.001]***	[0.001]***
(CL S.E.)	(0.034)***	(0.034)***	(0.030)***	(0.030)**
Basic controls & gender	Y	Y	Y	Y
High school controls & baseline belief	N	N	Y	Y
Personality traits	N	N	N	Y
Observations	267	252	252	252
R-squared	0.090	0.098	0.184	0.235
Mean of DV	0.389	0.389	0.389	0.389
SD of DV	0.081	0.082	0.082	0.082

Notes: The table presents the impact of minority status on students' believed gender ratio in the course, using OLS regressions. The dependent variable is the believed share of women measured in the 2019 endline survey. Basic controls include student major, course-retaking status, and year fixed effects. High school controls include grades for math and first language, and whether the first language is German. Baseline belief is the believed share of women at baseline. Personality traits refer to the Big-Five personality traits. Results are similar when I separately estimated the effect of minority status for women and men (see Appendix Table B7). Randomization Inference (RI) p-values are in brackets, and clustered standard errors are in parentheses. $*p < .1, **p < .05, ***p < .01$.

Figure 6: Group Gender Composition and Educational Expectation at Endline



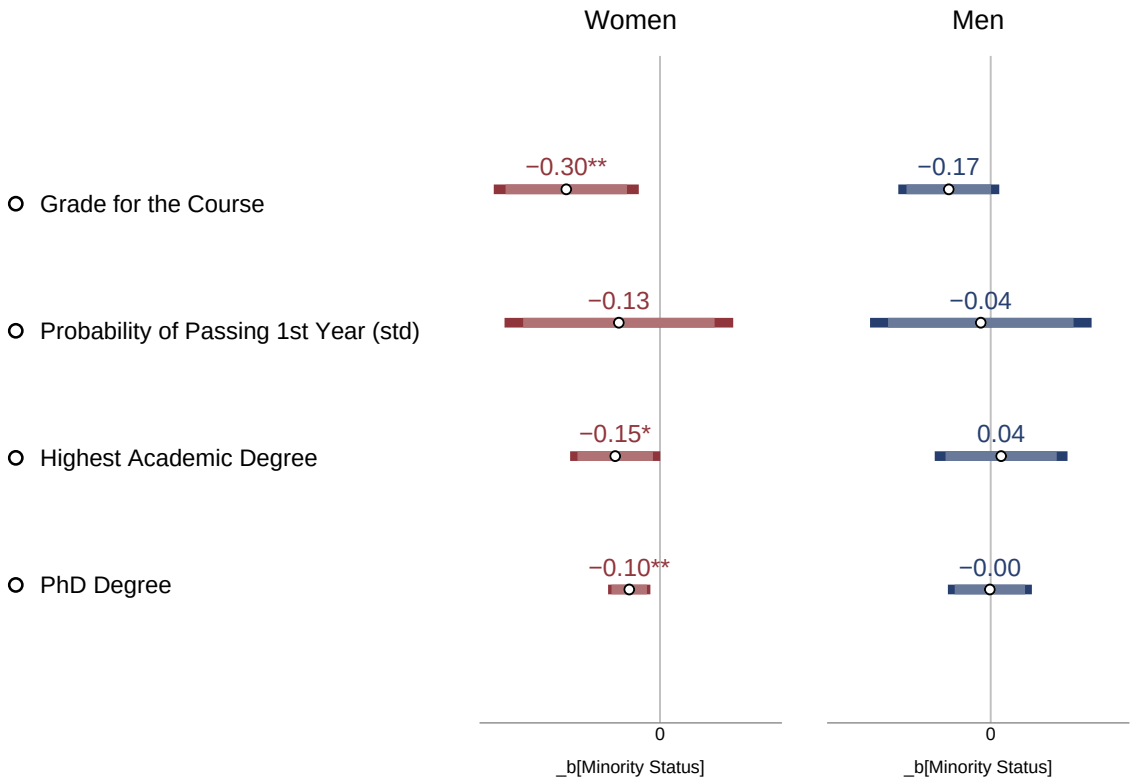
Notes: The figure plots residuals of the overall educational expectation at endline, after controlling for assignment strata (year and study program) fixed effects, by group gender composition. The gap in expectation residuals between women in female-minority groups and women in other groups is more than 0.4 standard deviations; the gap between men in male-minority groups and other men is less than 0.1 standard deviations. Error bars indicate standard error of the mean.

Table 5: The Impact of Minority Status on the Overall Educational Expectation

	(1)	(2)	(3)	(4)
DV: Overall Educational Expectation at Endline				
Minority Status	-0.429	-0.423	-0.401	-0.404
[RI p-val]	[0.005]***	[0.005]***	[0.009]***	[0.009]***
(CL S.E.)	(0.178)**	(0.178)**	(0.178)**	(0.170)**
Minority Status × Male	0.321	0.263	0.241	0.209
[RI p-val]	[0.177]	[0.233]	[0.275]	[0.349]
(CL S.E.)	(0.270)	(0.268)	(0.267)	(0.264)
Basic controls	Y	Y	Y	Y
Baseline expectation	N	Y	Y	Y
High school controls	N	N	Y	Y
Personality traits	N	N	N	Y
Observations	434	408	408	408
R-squared	0.080	0.269	0.282	0.295

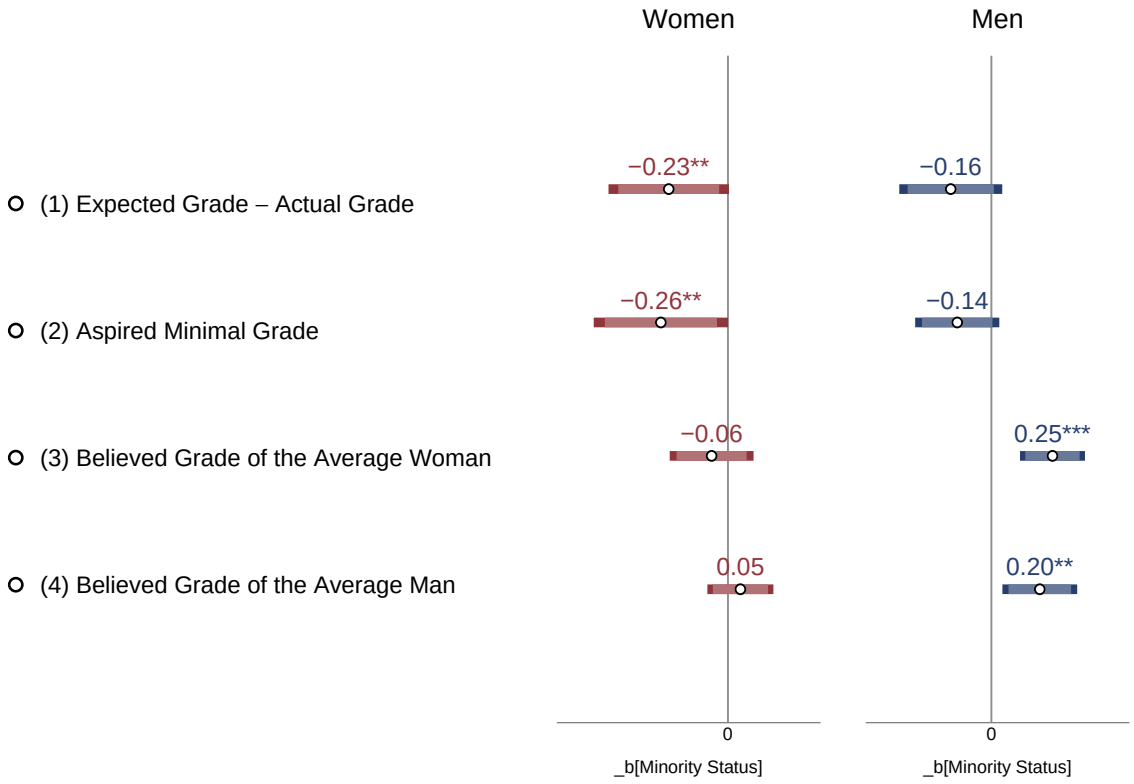
Notes: The table examines the impact of minority status on the overall educational expectation at endline. All models are estimated using OLS regressions. Basic controls include student major, course-retaking status, and year fixed effects. Baseline expectation is the overall expectation measured at the beginning of the course. High school controls include high school math and language grades, and whether the first language is German. Personality traits refer to the Big-Five personality traits. Split-sample results by gender are similar (see Appendix Table B8). Standard errors in parentheses are clustered at the group level. Randomization Inference (RI) p-values are in brackets, and clustered standard errors are in parentheses. $*p < .1, **p < .05, ***p < .01$.

Figure 7: Impact on Different Aspects of Educational Expectations



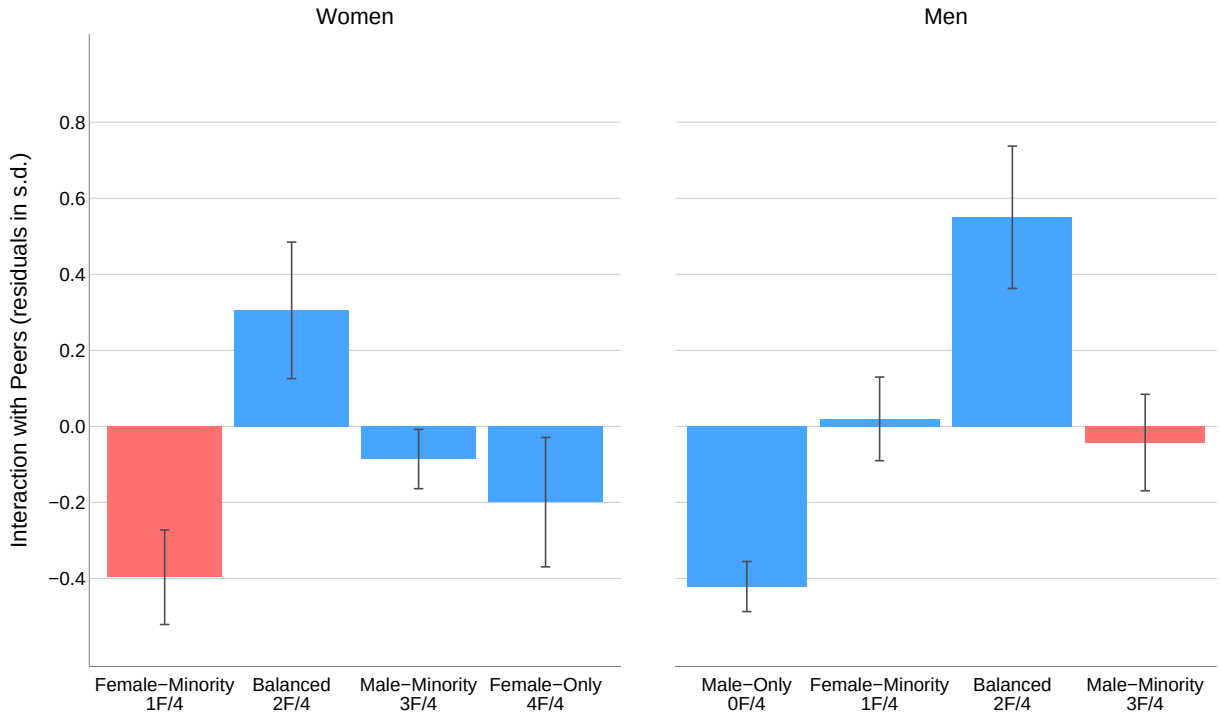
Notes: The graph shows the impact of minority status on different aspects of endline educational expectations, separately for women and men. The dependent variables are respectively: the expected grade for the course, the expected probability of passing the first year, the expected highest academic degree, and an indicator for whether the expected highest degree is a PhD. All regressions control for major, course-retaking status, year fixed effects, the baseline level of the dependent variable, high school controls, and Big-Five personality traits. 95% and 90% confidence intervals are based on standard errors clustered at the group level; significance stars are based on randomization inference (RI) p-values. *RI p <.1, **RI p <.05, ***RI p <.01.

Figure 8: Driving Forces of Lower Expected Grades



Notes: The graph plots the estimated effects of minority status on the following outcomes: (1) the gap between the expected and actual grade—a positive gap implies overestimation of own ability while a negative gap implies underestimation; (2) the aspired minimal grade for the course, which captures educational aspiration; (3)–(4) the believed grade for the average woman/man in the previous cohort, which reflects the social comparison effect. Measures (2)–(4) are available only in 2019. For measure (1), results including and excluding dropouts (lowest grades) are very similar. All regressions control for the baseline level of the outcome variable, major, retaking status, year fixed effects if applicable, high school grades, whether the first language is German, and the Big-Five personality traits. 95% and 90% confidence intervals are based on standard errors clustered at the group level; significance stars are based on randomization inference (RI) p-values. *RI p <.1, **RI p <.05, ***RI p <.01.

Figure 9: Group Gender Composition and Interaction with Peers



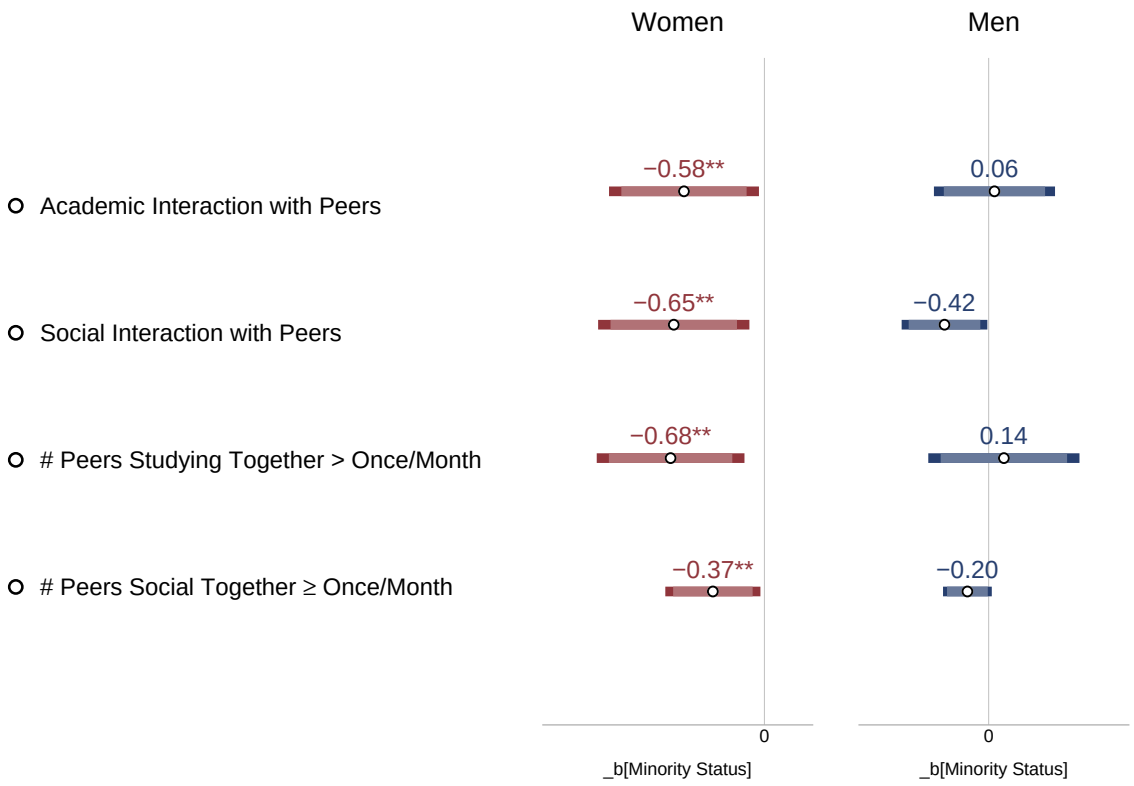
Notes: The figure plots the overall intensity of interaction with peers, after controlling for study program fixed effects, by group gender composition. The raw gap in expectation between women in female-minority groups and women in other groups is 0.45 standard deviations; the gap between men in male-minority groups and other men is about 0.05 standard deviations. Error bars indicate standard error of the mean.

Table 6: Minority Status and Peer-to-Peer Interaction

	(1)	(2)	(3)	(4)
DV: Overall Frequency of Interaction with Peers				
Minority Status	-0.495	-0.536	-0.519	-0.563
[RI p-val]	[0.005]***	[0.005]***	[0.007]***	[0.003]***
(CL S.E.)	(0.219)**	(0.229)**	(0.232)**	(0.246)**
Minority Status × Male	0.478	0.480	0.432	0.416
[RI p-val]	[0.050]**	[0.080]*	[0.115]	[0.133]
(CL S.E.)	(0.325)	(0.339)	(0.337)	(0.343)
Basic controls	Y	Y	Y	Y
High school controls	N	N	Y	Y
Personality traits	N	N	N	Y
Observations	348	318	318	318
R-squared	0.072	0.081	0.104	0.114

Notes: The DV is the standardized index for the overall interaction with peers. All regressions use OLS regressions and include the basic controls: course-retaking status and major fixed effects. High school controls include math and language grades, and whether the first language is German. Personality traits refer to the Big-Five personality traits. Results are very similar when I split the sample by gender and estimate the impact of minority status separately for women and men (see Appendix Table B9). Randomization Inference (RI) p-values are in brackets, and clustered standard errors are in parentheses. $*p < .1, **p < .05, ***p < .01$.

Figure 10: Impact on Different Measurements of Peer-to-Peer Interaction



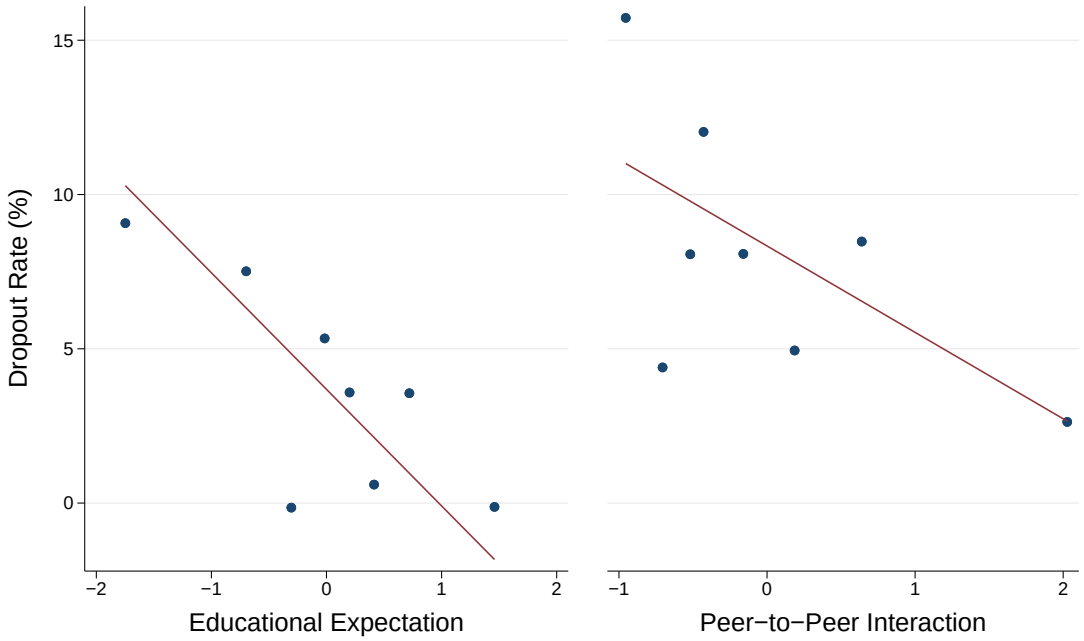
Notes: The graph shows the impact of minority status on peer-to-peer interaction measured in different ways: the average academic interaction with peers, the average social interaction with peers, the number of peers studying together more than once per month, and the number of peers having social events together for at least once per month. The first two measurements are standardized indicators. All regressions control for major, course-retaking status, high school controls, and personality traits. 95% and 90% confidence intervals are based on standard errors clustered at the group level; significance stars are based on randomization inference (RI) p-values. *RI p <.1, **RI p <.05, ***RI p <.01.

Table 7: Gender Homophily in Pairwise Interaction and Evaluation

	(1)	(2)	(3)	(4)	(5)	(6)
		Men			Women	
	2F/4	1F/4	Both	2F/4	3F/4	Both
Panel A: Interaction with Peers						
Female Peer	-0.458	-0.316	-0.381	0.123	-0.016	0.059
	(0.245)*	(0.119)**	(0.128)***	(0.087)	(0.058)	(0.055)
Observations	123	147	270	135	114	249
R-squared	0.683	0.643	0.674	0.857	0.800	0.859
Mean of DV	0.394	0.057	0.210	0.355	-0.209	0.097
SD of DV	1.365	0.967	1.175	1.316	0.548	1.073
Panel B: Willingness to Pay for Peers						
Female Peer	-0.341	-0.375	-0.359	0.023	0.037	0.029
	(0.216)	(0.145)**	(0.125)***	(0.114)	(0.049)	(0.065)
Observations	123	147	270	135	114	249
R-squared	0.397	0.608	0.522	0.902	0.553	0.880
Mean of DV	0.032	0.128	0.084	0.310	-0.204	0.074
SD of DV	0.967	1.050	1.012	1.471	0.469	1.156

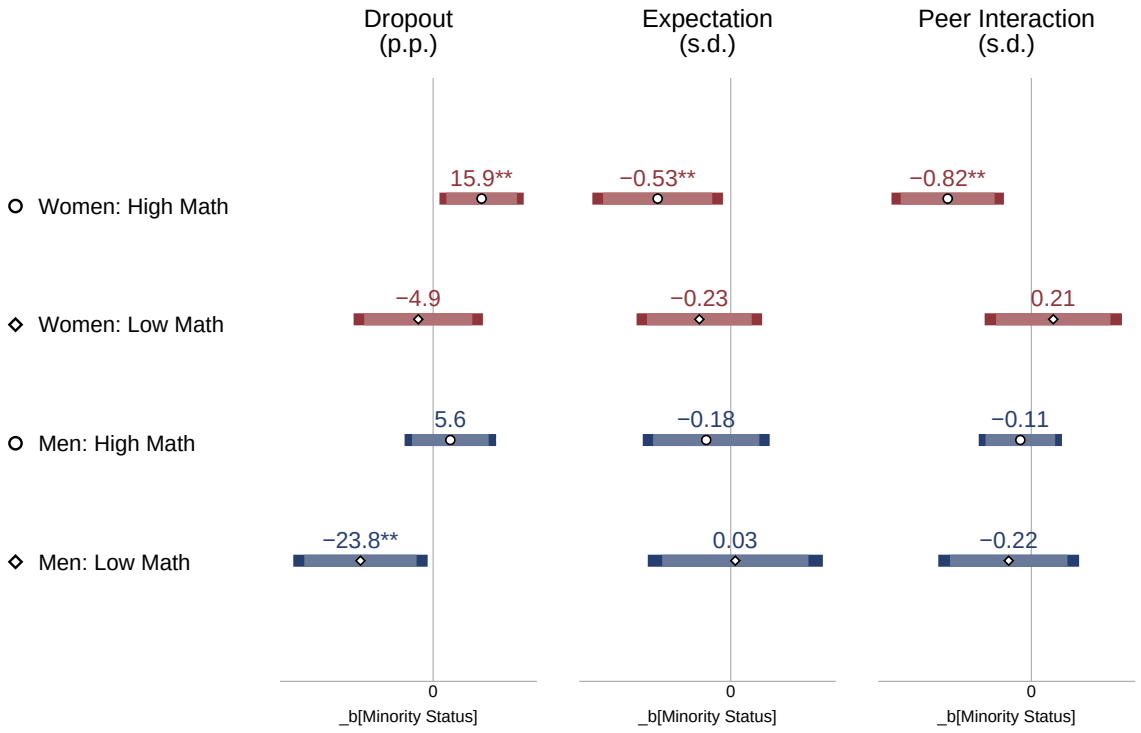
Notes: The table examines how students' interaction with peers and willingness for study group peers depend on the gender of the peer. The dependent variable in Panel A is the standardized frequency of interaction with peers; in Panel B, the dependent variable is the standardized willingness to pay. Columns (1)–(3) focus on women, separately for those in gender-balanced, male-minority groups, and those in both types of groups. Columns (2)–(4) focus on men, separately for those in gender-balanced, female-minority groups, and those in both types of groups. All columns use pairwise observations and OLS regressions, and control for individual fixed effects. Standard errors are clustered at the individual level and are shown in parentheses. The significance levels are very similar when standard errors are clustered at the group level. $*p < .1, **p < .05, ***p < .01$.

Figure 11: Correlations between Dropout and Intermediate Outcomes



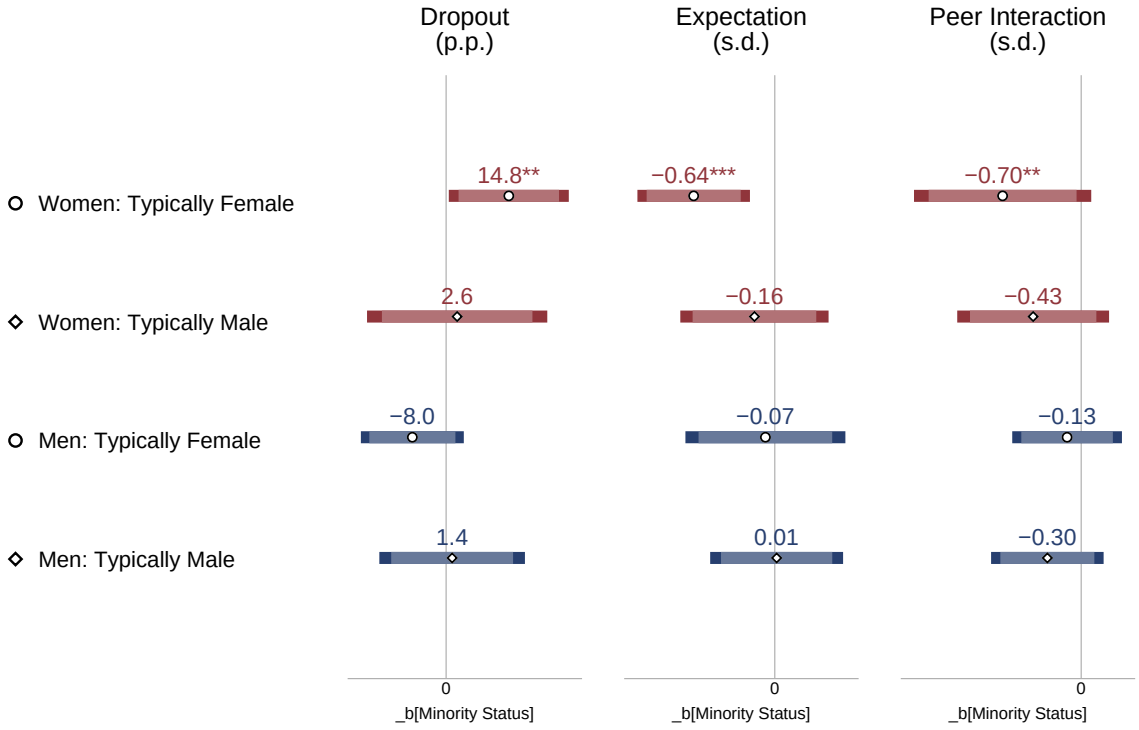
Notes: The binned scatter plots show how dropout correlates with the two intermediate outcomes: the overall educational expectation at endline and the intensity of interaction with peers. Both plots control for gender, major, course-taking status, and year fixed effects. The slopes in the two sub-figures are respectively *-.037*** and *-.028***.

Figure 12: Heterogeneous Impact of Minority Status by High School Math Grade



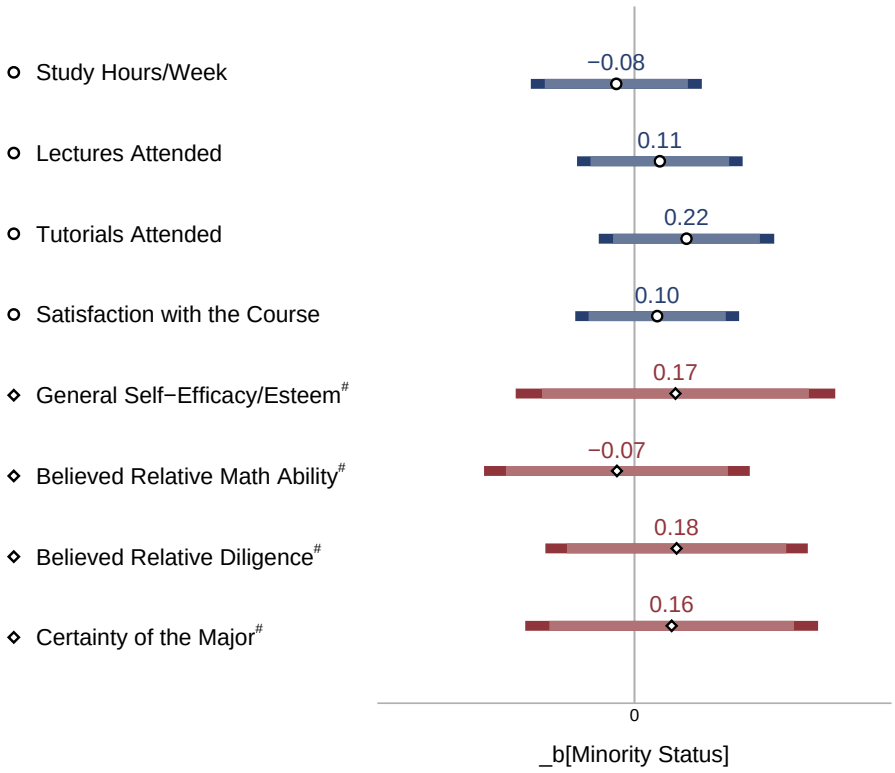
Notes: The graph shows the estimated effects of minority status on three outcomes: dropout, endline educational expectation, and peer-to-peer interaction. *High Math* refers to math grades equal to and above 4.5 (the median), while *Low Math* refers to grades below 4.5. See Appendix Figure B7 for the distribution of math grade by gender. Results are similar if students with math grades equal to 4.5 are categorized as *Low Math*. Control variables include high school background, baseline educational expectation (only in the middle panel), major, course-retaking status, and year fixed effects. 95% and 90% confidence intervals are based on standard errors clustered at the group level; significance stars are based on randomization inference (RI) p-values. *RI p <.1, **RI p <.05, ***RI p <.01.

Figure 13: Heterogeneous Impact of Minority Status by Gender Typicality



Notes: The graph shows the estimated effects of minority status on three outcomes: dropout, endline educational expectation, and peer-to-peer interaction. *Typically Female* students have an above-zero female typicality index, and *typically male* students have a below-zero index. Control variables include math grade, baseline educational expectation (only in the middle panel), major, course-retaking status, and year fixed effects. 95% and 90% confidence intervals are based on standard errors clustered at the group level; significance stars are based on randomization inference (RI) p-values. *RI p <.1, **RI p <.05, ***RI p <.01.

Figure 14: Does Minority Status Affect Female Dropout Through Other Channels?



Notes: The figure shows the estimated effects of minority status on different outcomes of women (results are all insignificant for men). Outcomes marked with # are only available in 2019. All regressions use standardized outcome variables, and control for major, retaking status, and year fixed effects. General self-efficacy/self-esteem is measured with a sub-scale of the General Self-Efficacy Scale (Schwarzer et al., 1995) and the Global Self-Esteem Scale (Rosenberg et al., 1995); other outcomes are measured with own survey questions. 95% and 90% confidence intervals are based on clustered standard errors at the group level. All results are also insignificant when using randomization inference.

Table 8: Group Registration and Dropout Behavior

	(1)	(2)	(3)	(4)
Panel A: Women's Dropout				
Having a Study Group	-0.060	-0.039	-0.045	-0.046
	(0.027)**	(0.027)	(0.027)*	(0.026)*
Observations	950	755	748	744
R-squared	0.049	0.082	0.099	0.104
Panel B: Men's Dropout				
Having a Study Group	-0.017	0.017	0.012	0.011
	(0.019)	(0.019)	(0.019)	(0.019)
Observations	1,361	947	937	927
R-squared	0.013	0.036	0.049	0.063
High school controls	N	Y	Y	Y
Personality traits	N	N	Y	Y
Additional controls	N	N	N	Y

Notes: The table examines whether women/men in study groups differ from those without study groups. All regressions control for major, course-retaking status, and year fixed effects. High school controls include high school math and language grades, and whether the first language is German. Additional individual controls include competitiveness, gender bias, and baseline expectation. Robust standard errors are in parentheses. $*p < .1, **p < .05, ***p < .01$.

A Conceptual Framework

I consider a simple dynamic model about voluntary dropout in college. Suppose that a student of gender $g = \{f = \text{female}, m = \text{male}\}$ is currently enrolled in college, and his/her existing academic achievement is a^g. The student has two options: *persistence* or *dropout*. Persistence means that he/she stays in college for another period and then enters the labor market, while dropout means that he/she withdraws from college and enters the labor market now. Suppose the expected utility of working per period in the labor market is U_l, which depends on academic achievement. If the student drops out now, his/her academic achievement is fixed at a^g, and the expected working utility is equal to $U_l(a^g)$ for each period. Using β to denote the discount factor, the value of dropout is simply

$$V^g_{dropout} = U_l(a^g) + \beta U_l(a^g) + \beta^2 U_l(a^g) + ... + \beta^T U_l(a^g), \tag{3}$$

where $T + 1$ denotes the number of periods in the labor market.

Instead, if the student persists in college, he/she derives utility from education in the next period, and after graduation, he/she derives utility form working in each period. Suppose the expected utility of education increases with the extent of social integration (S^g) in college: $U_c(S^g)$ and $U_c'(S^g) > 0$. The expected working utility after graduation also depends on the academic achievement to be obtained by graduation, which is expected to be $A^g \geq a^g$.[20] Suppose the expected academic achievement depends on existing achievement (a^g), the intended effort investment (e^g), as well as an error term (ε^g) which captures expectation bias due to incomplete information: $A^g = f(a^g, e^g) + \varepsilon^g$, and $f_a, f_e > 0$.

Taken together, the value of persisting in college is

$$V^g_{persist} = U_c(S^g) + \beta U_l(A^g) + \beta^2 U_l(A^g) + ... + \beta^T U_l(A^g). \tag{4}$$

By comparing $V^g_{dropout}$ to $V^g_{persist}$, the student decides to drop out or stay in college. He/She drops out of college when $V^g_{dropout} > V^g_{persist}$, i.e.,

$$\mathrm{d}^g = 1 \quad \text{if} \quad (1 + \widetilde{\beta})U_l(a^g) > U_c(S^g) + \widetilde{\beta} U_l(A^g), \tag{5}$$

where $\widetilde{\beta} = \beta + \beta^2 + ... + \beta^T$. Denote the probability of dropout as $\Pr(\mathrm{d}^g)$.

Proposition 1. *The probability of dropout from college decreases with the extent of social integration and the expected academic achievement:*

$$\frac{\partial \Pr(\mathrm{d}^g)}{\partial S^g} < 0 \quad \& \quad \frac{\partial \Pr(\mathrm{d}^g)}{\partial A^g} < 0.$$

From the decision rule as specified in Equation (5), it is straightforward to see that the probability of dropout *decreases* with social integration in college and the expectation about

[20] The student may also fail to graduate after another period of education. That case is captured by $A^g = a^g$.

future academic achievement. As S^g increases, the short-run utility of education in college goes up; as A^g increases, the total discounted utility of working in the labor market goes up.

Suppose contextual factors, including peers, can affect the level of social integration and expectations about academic achievement. This paper especially focuses on one contextual factor – the gender minority status. A female student experiences minority status if the peer group that she is exposed to is male-dominated. Similarly, a male student experiences minority status if he is exposed to a female-dominated peer environment.

Hypothesis 1. *Compared to a student of the non-minority gender, a student experiencing minority status interacts less with peers and feels less socially integrated in college:*

$$\Delta S^g_M \equiv S^g|_M - S^g|_{NonM} \leq 0.$$

As the principle of gender homophily (Shrum et al., 1988; Ibarra, 1992; Mengel, 2020) implies, people tend to interact more with same-gender peers than opposite-gender peers. Therefore, in a peer environment with a skewed gender ratio, the dominant gender may interact closely among themselves, while the minority gender may get socially marginalized. In addition, men typically exhibit more gender homophily than women, suggesting that minority status can decrease female students' social integration to a greater extent than male students: $\Delta S^f_M < \Delta S^m_M \leq 0$.

Hypothesis 2. *Minority status can lower a student's academic self-efficacy and the expectation about future academic achievement:*

$$\Delta A^g_M \equiv A^g|_M - A^g|_{NonM} \leq 0.$$

Note that the expected academic achievement is composed of two parts: a production function of existing achievement and effort investment, $f(a^g, e^g)$, and an error term representing expectation bias (ε^g). I conjecture that minority status mainly affects A^g through the error term: a student exposed to minority status becomes less confident about his/her ability, even if his/her academic potential stays the same. Due to the gender gap in self-confidence, I further hypothesize that the negative impact of minority status on the expected achievement is stronger for women than for men: $\Delta A^f_M < \Delta A^m_M \leq 0$.

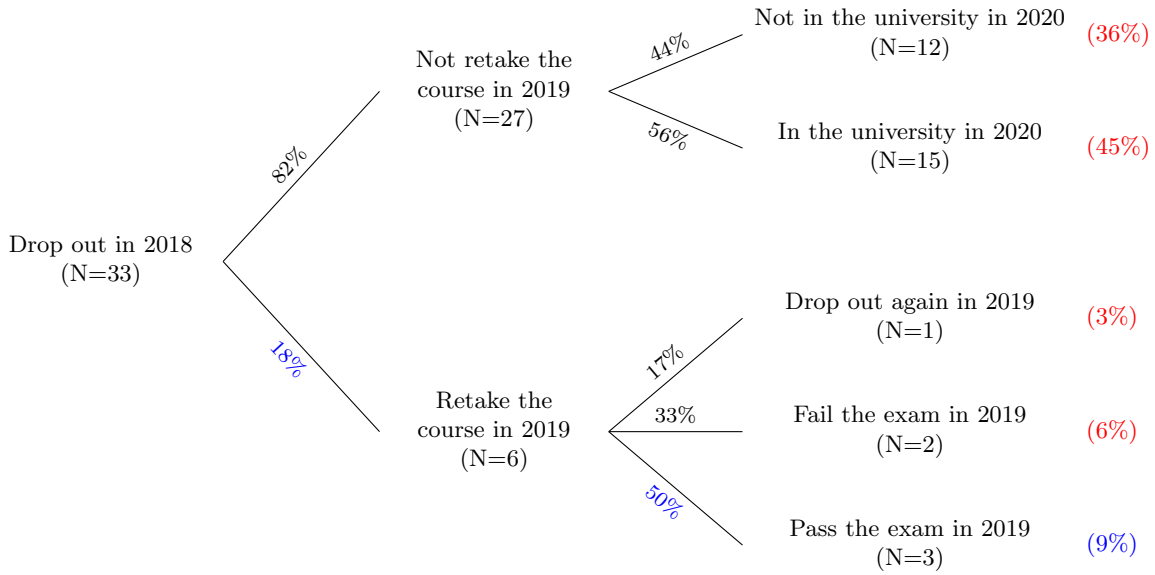
Proposition 2. *Given Proposition 1 and Hypotheses 1–2, minority status can increase the probability of dropout from college:*

$$\Delta \Pr(\mathrm{d}^g)_M \equiv \Pr(\mathrm{d}^g)|_M - \Pr(\mathrm{d}^g)|_{NonM} \geq 0.$$

As discussed above, minority status can lower a student's integration with college peers and his/her expected academic achievement, and consequently leads to a higher dropout rate. Furthermore, due to gender differences in homophily preference and self-confidence, minority status can induce more dropouts among women: $\Delta \Pr(\mathrm{d}^f)_M > \Pr(\mathrm{d}^m)_M \geq 0$.

B Additional Figures and Tables

Figure B1: What Happens after Dropping out



Note: The figure shows the different pathways that students took after dropping out of the economics course in 2018. Among the 33 students, 82% did not retake the course in 2019, suggesting that they were permanently separated from the course and the previously chosen majors or minors (as the course is compulsory). By September 2020, 36% of these students no longer had a valid university email address, meaning that they had dropped out from the university. For those who repeated the course in 2019, only half of them managed to pass the exam, but at the cost of one additional year in college.

Table B1: Test of Random Group Assignment

	Joint Significance of Study Group Dummies	
	F-stat	p-value
Gender	0.886	0.814
Baseline Education Expectation	0.869	0.845
High School Math Grade	0.873	0.837
High School Language Grade	0.922	0.719
High School in German	0.988	0.527
Conscientiousness	0.830	0.911
Extraversion	1.097	0.238
Agreeableness	0.925	0.713
Openness	1.103	0.224
Neuroticism	1.172	0.112

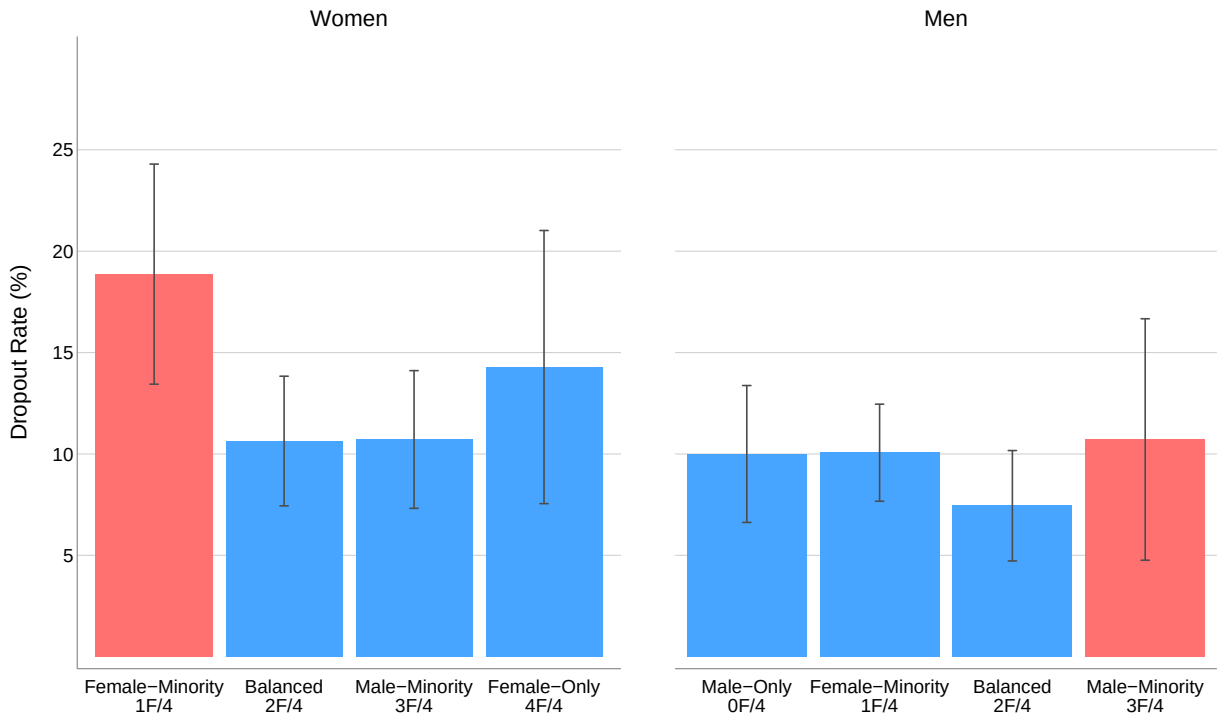
Notes: I first regress each characteristic on study program and year fixed effects and derive the residuals. Then, I regress the residuals on the group dummies and test the joint significance of group dummies.

Table B2: Balance Test by Gender

	Women			Men		
	_b[Minority]	[RI p-val]	(CL S.E.)	_b[Minority]	[RI p-val]	(CL S.E.)
Panel A: Baseline Characteristics						
High School Math Grade	-0.08	[0.559]	(0.120)	0.114	[0.543]	(0.181)
High School Language Grade	-0.023	[0.828]	(0.089)	-0.144	[0.256]	(0.103)
High School in German	-0.068	[0.307]	(0.065)	0.023	[0.787]	(0.070)
Conscientiousness	0.082	[0.633]	(0.143)	-0.017	[0.934]	(0.250)
Extraversion	-0.283	[0.110]	(0.159)*	-0.008	[0.973]	(0.227)
Agreeableness	0.043	[0.772]	(0.141)	0.055	[0.804]	(0.201)
Openness	0.040	[0.816]	(0.169)	0.057	[0.779]	(0.263)
Neuroticism	0.135	[0.454]	(0.186)	0.126	[0.582]	(0.235)
Baseline Educational Expectation	0.038	[0.819]	(0.130)	-0.074	[0.737]	(0.197)
Course-Retaking Status	-0.094	[0.065]*	(0.056)*	0.078	[0.077]*	(0.069)
Panel B: Survey Response						
Baseline Survey Completion	-0.022	[0.668]	(0.050)	-0.030	[0.713]	(0.071)
Endline Survey Completion	-0.025	[0.776]	(0.076)	0.020	[0.844]	(0.103)

Note: The table tests whether minority status predicts the baseline characteristics and survey response of women and men. For the baseline survey, the response rate is around 88% for both women and men. The the endline survey, the response rate is 69% for women and 64% for men. I regress each characteristic on the minority indicator, controlling for the fixed effects of study program and experiment year. Randomization Inference (RI) p-values are in brackets, and clustered standard errors are in parentheses. $*p < .1, **p < .05, ***p < .01$.

Figure B2: Group Gender Composition and Dropout



Notes: The figure shows the raw dropout rate by student gender and group gender composition. The raw dropout rate is 13% for women and 9% for men. Error bars indicate standard error of the mean.

Table B3: The Impact of Minority Status on Dropout by Gender (OLS)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Women				Men			
DV: Dropout								
Minority Status	0.101	0.109	0.101	0.105	-0.015	-0.018	-0.013	-0.010
[RI p-val]	[0.062]*	[0.032]**	[0.047]**	[0.037]**	[0.813]	[0.794]	[0.849]	[0.888]
(CL S.E.)	(0.057)*	(0.058)*	(0.057)*	(0.058)*	(0.061)	(0.063)	(0.065)	(0.067)
Basic controls	Y	Y	Y	Y	Y	Y	Y	Y
High school controls	N	N	Y	Y	N	N	Y	Y
Personality traits	N	N	N	Y	N	N	N	Y
Observations	259	235	235	235	361	319	319	319
R-squared	0.031	0.046	0.088	0.114	0.029	0.033	0.046	0.062

Notes: The table presents the impact of minority status on dropout, separately for women and men, using OLS regressions. Basic controls include student major, course-retaking status, and year fixed effects. High school controls include grades for math and first language, and whether the first language is German. Personality traits refer to the Big-Five personality traits. Results using probit regressions are very similar (see Appendix Table B4.) Randomization Inference (RI) p-values are in brackets, and clustered standard errors are in parentheses. $*p < .1, **p < .05, ***p < .01$.

Table B4: The Impact of Minority Status on Dropout by Gender (Probit)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Women				Men			
DV: Dropout								
Minority Status	0.097	0.099	0.098	0.095	-0.013	-0.012	-0.003	0.005
[RI p-val]	[0.060]*	[0.027]**	[0.025]**	[0.030]**	[0.827]	[0.838]	[0.967]	[0.927]
(CL S.E.)	(0.048)**	(0.045)**	(0.041)**	(0.041)**	(0.054)	(0.057)	(0.059)	(0.057)
Basic controls	Y	Y	Y	Y	Y	Y	Y	Y
High school controls	N	N	Y	Y	N	N	Y	Y
Personality traits	N	N	N	Y	N	N	N	Y
Observations	259	235	235	235	361	319	319	319
Pseudo R-squared	0.041	0.064	0.133	0.172	0.046	0.058	0.082	0.115

Notes: The table presents the impact of minority status on dropout, separately for women and men. Use probit regressions and report the estimated marginal effects. Basic controls include student major, course-retaking status, and year fixed effects. High school controls include grades for math and first language, and whether the first language is German. Personality traits refer to the Big-Five personality traits. Randomization Inference (RI) p-values are in brackets, and clustered standard errors are in parentheses. $*p < .1, **p < .05, ***p < .01$.

Table B5: Group Gender Composition and Dropout (Different Models)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Women				Men			
DV: Dropout								
Minority Status	0.101				-0.015			
[RI p-val]	[0.062]*				[0.830]			
(CL S.E.)	(0.057)*				(0.061)			
# Women per group		-0.040	-0.170			-0.028	-0.062	
[RI p-val]		[0.113]	[0.000]***			[0.155]	[0.001]***	
(CL S.E.)		(0.024)*	(0.106)			(0.016)*	(0.044)	
(# Women per group)^2			0.027				0.013	
[RI p-val]			[0.000]***				[0.072]*	
(CL S.E.)			(0.020)				(0.017)	
Gender-Balanced				-0.085				-0.006
[RI p-val]				[0.057]*				[0.876]
(CL S.E.)				(0.061)				(0.065)
Female-/Male- Majority				-0.124				0.021
[RI p-val]				[0.012]**				[0.514]
(CL S.E.)				(0.062)**				(0.063)
Female-/Male- Only				-0.099				0.068
[RI p-val]				[0.177]				[0.061]*
(CL S.E.)				(0.073)				(0.065)
Observations	259	259	259	259	361	361	361	361
R-squared	0.031	0.028	0.033	0.033	0.029	0.034	0.035	0.035

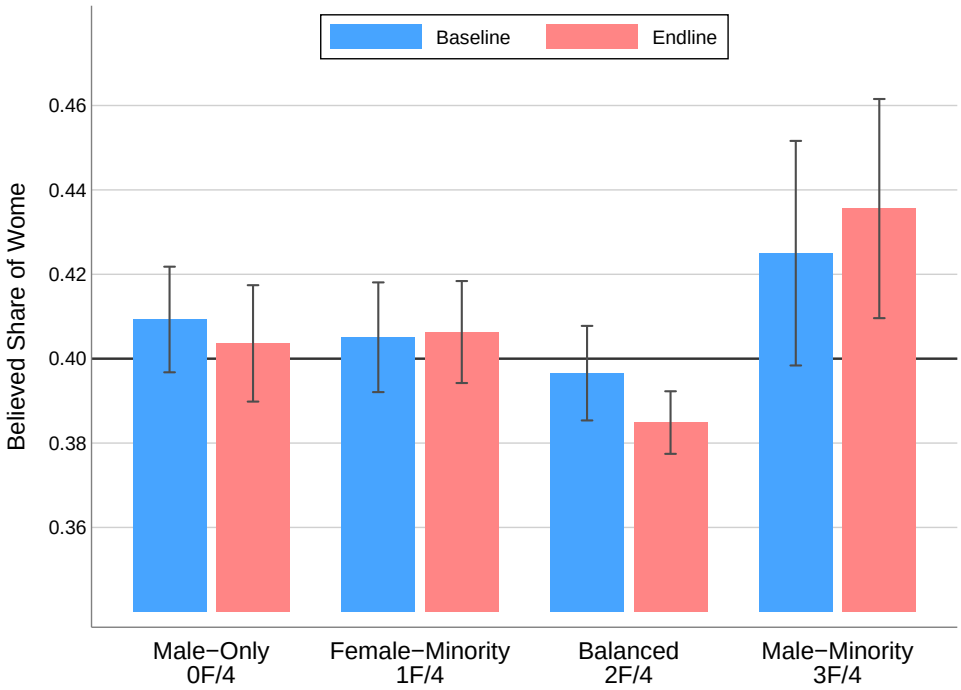
Notes: The table analyzes the impact of group gender composition on women's and men's dropout, using different models. Column (1) & (5) conduct the simple binary comparison between students assigned to be in the minority and other students; columns (2) & (6) use a linear model by estimating the marginal effect of an additional women in a study group; columns (3) & (7) further relax the model by including a squared term of the number of women per group; columns (4) & (8) are the fully saturated model including dummies for all possible gender compositions. All columns use OLS regressions and control for major, course-retaking status, and year fixed effects. Randomization Inference (RI) p-values are in brackets, and clustered standard errors are in parentheses. $*p < .1, **p < .05, ***p < .01$.

Table B6: The Impact of Minority Status on the Final Grade

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	All Students				Students Taking the Exam			
DV: Final Grade								
Minority Status	-0.299	-0.349	-0.280	-0.287	-0.003	-0.018	0.048	0.025
[RI p-val]	[0.049]**	[0.025]**	[0.063]*	[0.051]*	[0.985]	[0.863]	[0.632]	[0.784]
(CL S.E.)	(0.218)	(0.228)	(0.221)	(0.218)	(0.149)	(0.155)	(0.145)	(0.144)
Minority Status × Male	0.189	0.297	0.188	0.197	-0.194	-0.136	-0.187	-0.165
[RI p-val]	[0.456]	[0.243]	[0.471]	[0.446]	[0.241]	[0.449]	[0.242]	[0.326]
(CL S.E.)	(0.338)	(0.346)	(0.344)	(0.350)	(0.227)	(0.230)	(0.198)	(0.203)
Basic controls & gender	Y	Y	Y	Y	Y	Y	Y	Y
High school controls	N	N	Y	Y	N	N	Y	Y
Personality traits	N	N	N	Y	N	N	N	Y
Observations	620	554	554	554	553	502	502	502
R-squared	0.029	0.039	0.135	0.155	0.046	0.058	0.188	0.205

Notes: The table examines the impact of minority status on students' final grades ranging from point 1 to 6. All columns use OLS regressions, and include the basic controls: major, course-retaking status, and year fixed effects. High school controls are math and language grades and whether the first language is German. Personality traits are the Big-Five personality traits. Randomization Inference (RI) p-values are in brackets, and clustered standard errors are in parentheses. $*p < .1, **p < .05, ***p < .01$.

Figure B3: Male Students' Believed Share of Women in the Course



Notes: The figure shows male students' believed share of women in the economics course by group gender composition at the beginning and end of the semester in 2019. The actual share of women is 40%, as indicated by the black line. Error bars indicate the standard error of the mean.

Table B7: The Impact of Minority Status on the Believed Share of Women by Gender

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Women				Men			
DV: Believed Share of Women in the Course at Endline								
Minority Status	-0.039	-0.038	-0.048	-0.053	0.061	0.058	0.057	0.042
[RI p-val]	[0.041]**	[0.062]*	[0.029]**	[0.025]**	[0.039]**	[0.056]*	[0.027]**	[0.094]*
(CL S.E.)	(0.020)*	(0.021)*	(0.022)**	(0.025)**	(0.025)**	(0.024)**	(0.020)***	(0.018)**
Basic controls	Y	Y	Y	Y	Y	Y	Y	Y
HS controls & BSL belief	N	N	Y	Y	N	N	Y	Y
Personality traits	N	N	N	Y	N	N	N	Y
Observations	110	102	102	102	157	150	150	150
R-squared	0.130	0.141	0.187	0.280	0.044	0.041	0.228	0.282
Mean of DV	0.370	0.368	0.368	0.368	0.402	0.404	0.404	0.404
SD of DV	0.071	0.073	0.073	0.073	0.084	0.085	0.085	0.085

Notes: The table presents the impact of minority status on students' believed gender ratio in the course, using OLS regressions. The dependent variable is the believed share of women measured in the 2019 endline survey. Basic controls include student major, course-retaking status, and year fixed effects. High school (HS) controls include grades for math and first language, and whether the first language is German. Baseline (BSL) belief is the believed share of women at baseline. Personality traits refer to the Big-Five personality traits. Randomization Inference (RI) p-values are in brackets, and clustered standard errors are in parentheses. $*p < .1, **p < .05, ***p < .01$.

Table B8: The Impact of Minority Status on the Overall Educational Expectation

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Women				Men			
DV: Overall Educational Expectation at Endline								
Minority Status	-0.409	-0.441	-0.426	-0.400	-0.032	-0.065	-0.093	-0.101
[RI p-val]	[0.028]**	[0.020]**	[0.022]**	[0.030]**	[0.896]	[0.775]	[0.684]	[0.658]
(CL S.E.)	(0.172)**	(0.171)**	(0.169)**	(0.169)**	(0.199)	(0.195)	(0.198)	(0.207)
Basic controls	Y	Y	Y	Y	Y	Y	Y	Y
Baseline Expectation	N	Y	Y	Y	N	Y	Y	Y
High school controls	N	N	Y	Y	N	N	Y	Y
Personality traits	N	N	N	Y	N	N	N	Y
Observations	187	177	177	177	247	231	231	231
R-squared	0.082	0.226	0.231	0.261	0.065	0.296	0.314	0.335
Mean of DV	-0.133	-0.108	-0.108	-0.108	0.101	0.124	0.124	0.124
SD of DV	0.961	0.966	0.966	0.966	1.019	1.020	1.020	1.020

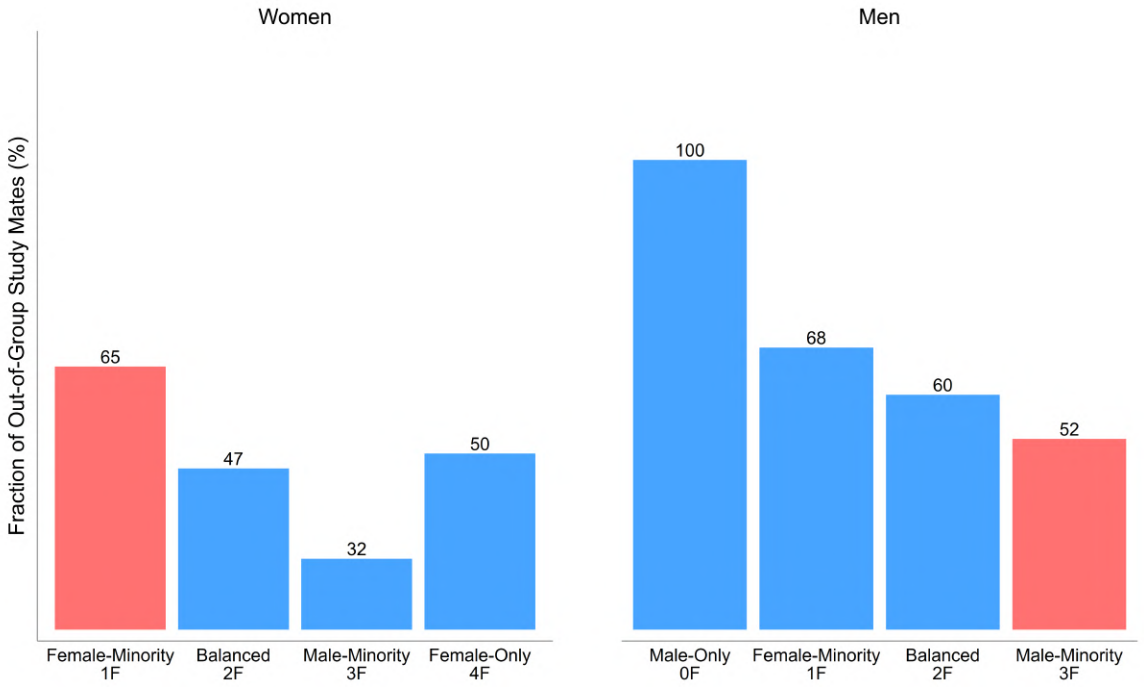
Notes: The table examines the impact of minority status on the overall educational expectation at endline. All models are estimated using OLS regressions. Basic controls include student major, course-retaking status, and year fixed effects. Baseline expectation is the overall expectation measured at the beginning of the course. High school controls include high school math and language grades, and whether the first language is German. Personality traits refer to the Big-Five personality traits. Randomization Inference (RI) p-values are in brackets, and clustered standard errors are in parentheses. $*p < .1, **p < .05, ***p < .01$.

Table B9: The Impact of Minority Status on Peer-to-Peer Interaction by Gender

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Women				Men			
DV: Overall Frequency of Interaction with Peers								
Minority Status	-0.540	-0.604	-0.612	-0.693	-0.107	-0.145	-0.150	-0.204
[RI p-val]	[0.037]**	[0.026]**	[0.028]**	[0.013]**	[0.705]	[0.644]	[0.645]	[0.525]
(CL S.E.)	(0.225)**	(0.241)**	(0.249)**	(0.278)**	(0.175)	(0.180)	(0.175)	(0.190)
Basic controls	Y	Y	Y	Y	Y	Y	Y	Y
High school controls	N	N	Y	Y	N	N	Y	Y
Personality traits	N	N	N	Y	N	N	N	Y
Observations	146	131	131	131	202	187	187	187
R-squared	0.108	0.138	0.177	0.197	0.076	0.084	0.108	0.119
Mean of DV	-0.028	0.025	0.025	0.025	0.020	0.044	0.044	0.044
SD of DV	0.982	1.008	1.008	1.008	1.015	1.033	1.033	1.033

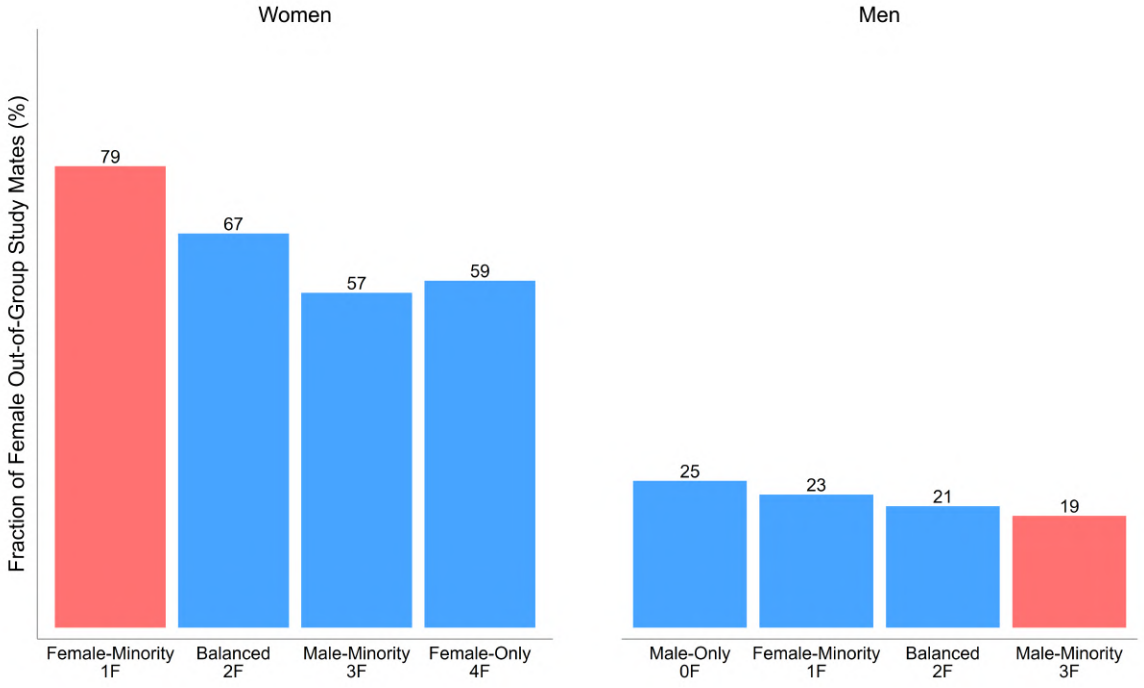
Notes: The DV is the standardized index for the overall interaction with peers. All regressions use OLS regressions and include the basic controls: course-retaking status and major fixed effects. High school controls include math and language grades, and whether the first language is German. Personality traits refer to the Big-Five personality traits. Randomization Inference (RI) p-values are in brackets, and clustered standard errors are in parentheses. $*p < .1, **p < .05, ***p < .01$.

Figure B4: Gender Composition and Fraction of Out-of-Group Study Mates



Notes: In the 2018 endline survey, students reported up to five study mates that they studied most frequently with. The named students could be assigned peers from the study group or endogenous peers from out of the study group. The figure shows how the share of out-of-group study mates varies with group gender composition. The differences across gender compositions are statistically insignificant.

Figure B5: Gender Composition and Fraction of Female Out-of-Group Study Mates



Notes: In 2018, students reported up-to-five frequently-interacted study mates from within or out of the assigned group. In 2019, students report up-to-two study mates from out of the assigned group. I calculate the share of female peers among the reported out-of-group study mates, and plot the share by group gender composition. The differences across gender compositions are statistically insignificant.

Table B10: Gender Homophily in Willingness to Pay for Peers

	(1)	(2)	(3)	(4)	(5)	(6)
	Women			Men		
DV: Willingness to Pay for Peers						
Female Peer	-0.022	-0.026	-0.058	-0.242	-0.365	-0.396
	(0.072)	(0.081)	(0.091)	(0.127)*	(0.127)***	(0.132)***
Observations	148	146	120	168	155	135
R-squared	0.736	0.920	0.913	0.516	0.621	0.613
Group FEs	Y	N	N	Y	N	N
Individual FEs	N	Y	Y	N	Y	Y

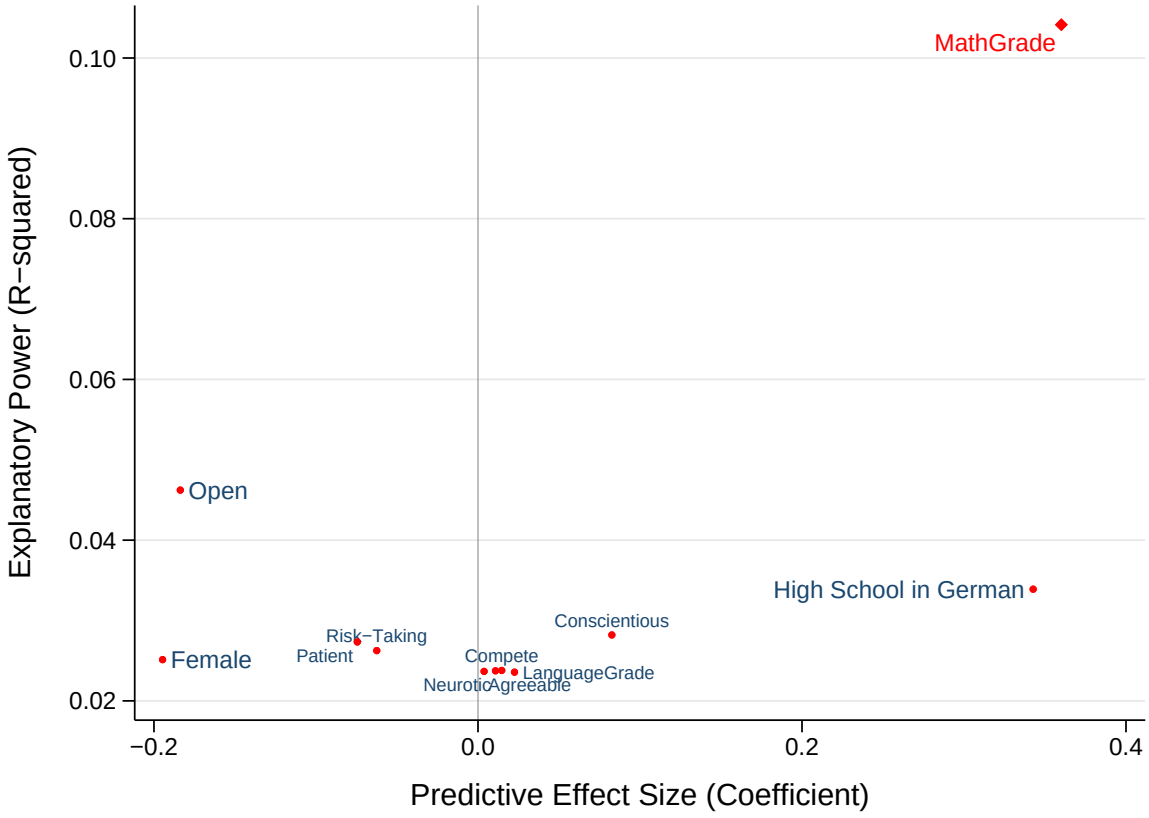
Notes: Use OLS regressions. The dependent variable is the standardized willingness to pay (WTP) for peers. Regressions in columns (1) & (4) control for study group fixed effects, student major, and course-retaking status, and cluster standard errors at the group level. Regressions in columns (2)–(3) and (5)–(6) control for individual fixed effects and cluster standard errors at the individual level. Regressions (1) & (4) include all non-missing pairwise observations reported by all students. Regressions (2) and (5) drop students reporting the WTP for only one peer in the group, by controlling for individual fixed effects. Regressions (3) and (6) only include students who interact with all three peers in the group and report the WTP for the three peers. $*p < .1, **p < .05, ***p < .01$.

Table B11: Effects of Minority Status Using Lee Bounds

	(1)	(2)	(3)
DV: Overall Educational Expectation at Endline			
OLS Estimation	-0.482	-0.475	-0.536
95% Confidence Interval	[-0.837, -0.128]	[-0.817, -0.132]	[-0.883, -0.189]
Lower Bound	-0.571	-0.554	-0.588
Upper Bound	-0.383	-0.392	-0.476
DV: Overall Interaction with Peers			
OLS Estimation	-0.351	-0.362	-0.402
95% Confidence Interval	[-0.701, -0.001]	[-0.718, -0.005]	[-0.756, -0.026]
Lower Bound	-0.430	-0.439	-0.501
Upper Bound	-0.105	-0.147	-0.117
Inverse Probability Weighting	N	Y	Y

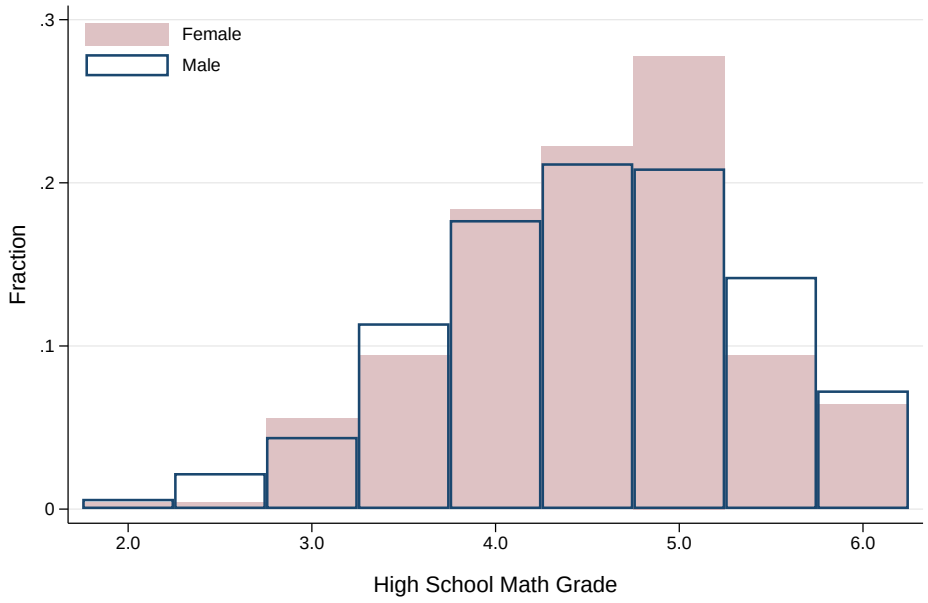
Notes: The table shows the original OLS estimations with 95% confidence intervals and the Lee upper/lower bounds. OLS regressions cluster standard errors at the group level. Columns (2)–(3) conduct inverse probability weighted analyses, where the probability is the predicted likelihood of having non-missing values for the dependent variable. The prediction of the likelihood in column (2) is only based on administrative records—student major, course-retaking status, and year fixed effects. The prediction in column (3) also includes high school background and personality traits.

Figure B6: How Different Characteristics Predict Course Grade



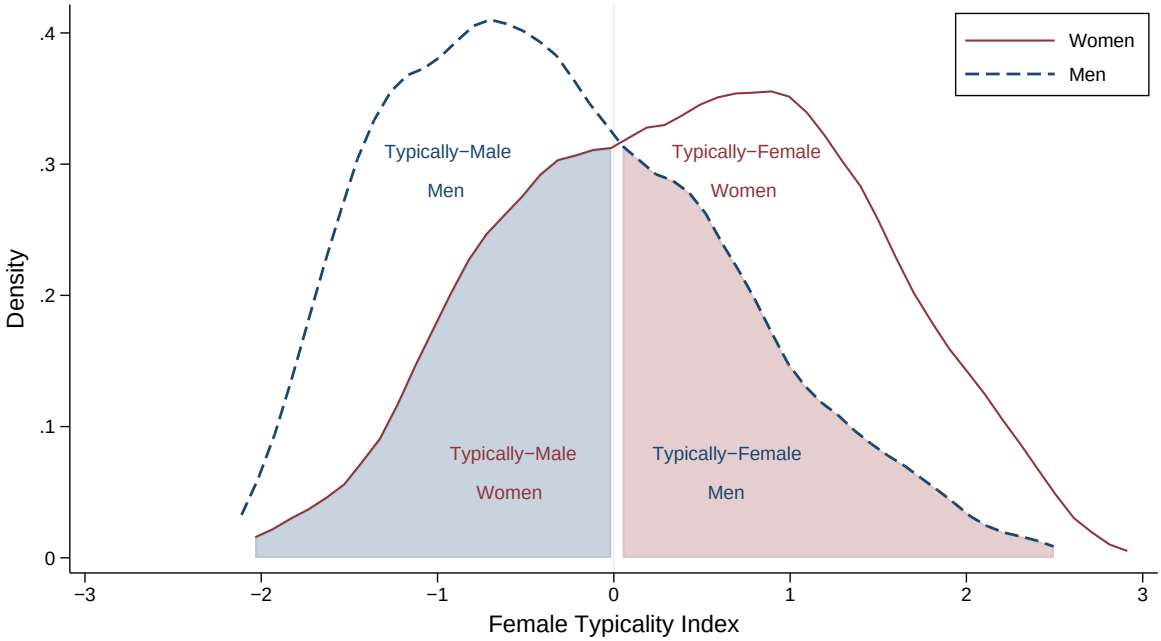
Notes: The figure shows how the final grade is predicted by baseline characteristics, including gender, high school background, personality traits, gender bias, general competitiveness, risk preference, and patience. I regress grade on each characteristic (standardized) separately, controlling for major, course-retaking status, and year fixed effects, and plot the corresponding coefficient and R-squared. Risk preference and patience are measured with single-item questions asking how risk-taking and patient a student perceives himself/herself.

Figure B7: The Distribution of High School Math Grade by Gender



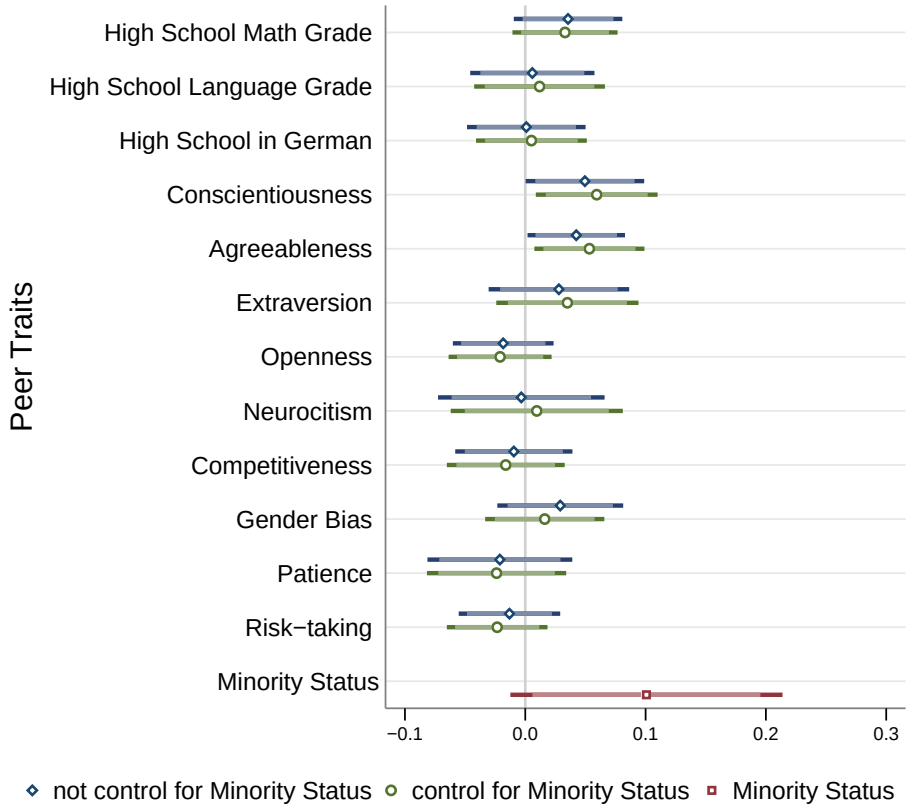
Notes: The grade plots the distribution of self-reported high school math grade by gender. The median grade for both female and male students is *4.5*.

Figure B8: The Distribution of Female Typicality by Gender



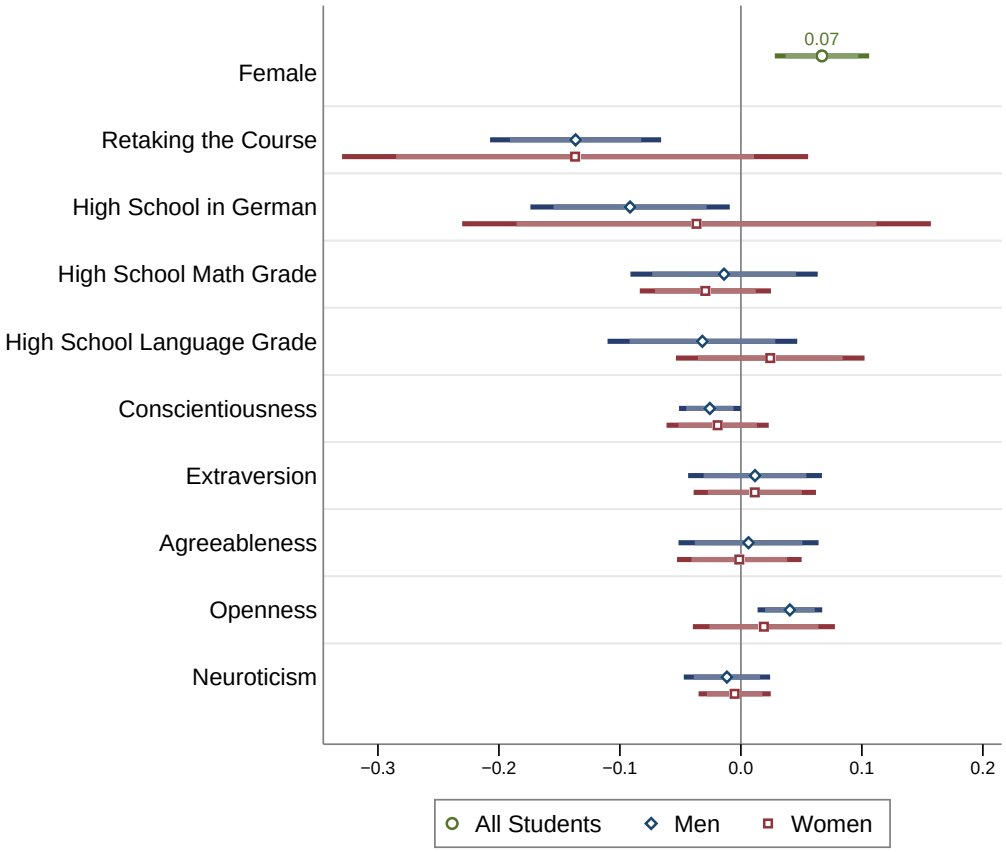
Notes: The figure plots the distribution of the female typicality index by gender. Female typicality is the predicted value of female as a linear combination of the Big-Five personality traits, risk preference, patience, competitiveness, and gender bias. Students with an above-zero female typicality index are defined as *typically female*, while students with a below-zero female typicality index are defined as *typically male*.

Figure B9: The Impact of Peer Traits on Women's Dropout



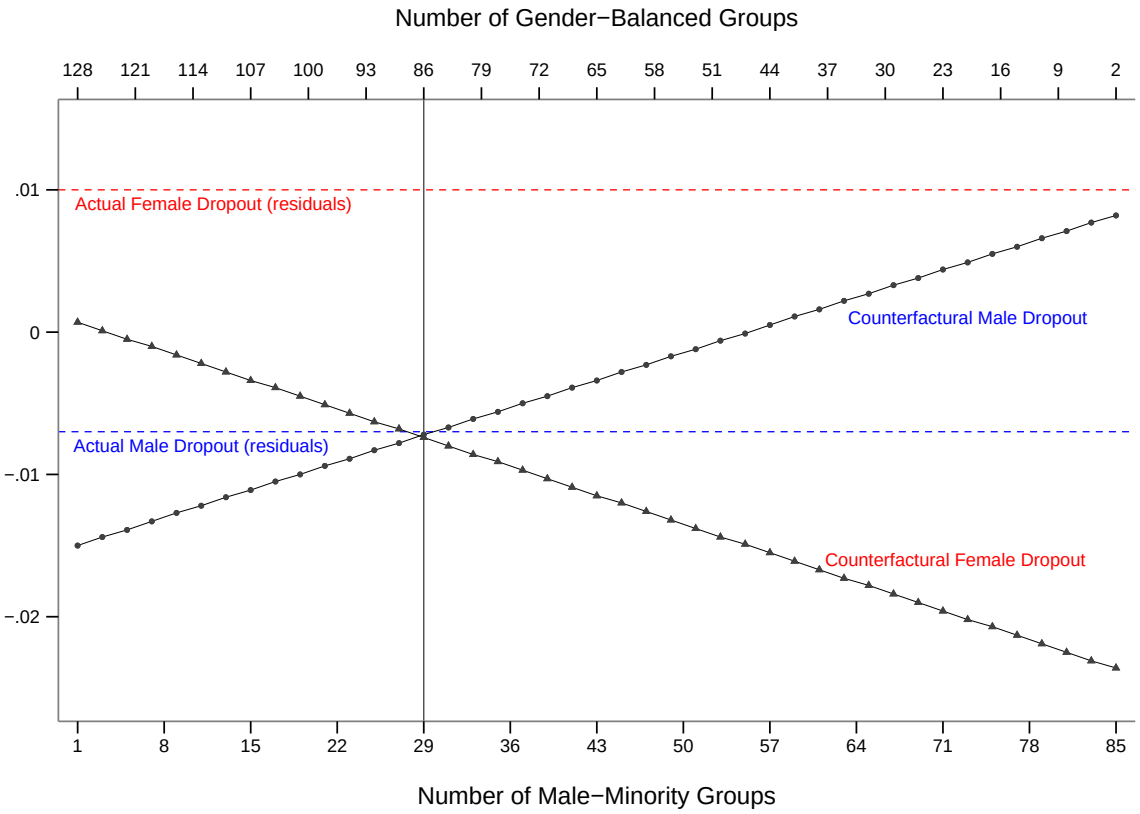
Notes: The figure plots the estimated coefficients of different peer traits with or without controlling for minority status and the coefficient of minority status. Each point corresponds to one OLS regression, where the dependent variable is the dropout indicator for women, and controls include major, course-retaking status, and year fixed effects. Peer traits are measured as the average of three peers. Peer traits not shown here have no significant effects on women's dropout. 95% and 90% confidence intervals are based on clustered standard errors at the group level.

Figure B10: What Kind of Students Register for Study Groups?



Notes: The figure plots the estimated coefficients of different characteristics with 95% and 90% confidence intervals. Use the whole sample of students taking the Economics course in 2018 and 2019, including both those with study groups and those without study groups. Each point corresponds to one OLS regression where the dependent variable is an indicator for group registration, the independent variable of interest is a student characteristic, and the controls include major, retaking status, and year fixed effects. Cluster S.E. at the major level.

Figure B11: Counterfactual Dropout Rate of Women and Men in the Experiment



Notes: The graph shows how the counterfactual dropout rate varies with the number of gender-balanced and female-majority (or male-minority) groups. I first partial out major, course-retaking status, and year fixed effects and derive the residuals of dropout. From the dropout residuals, we know the actual average dropout rate among women (*0.01*) and men (*-0.007*)—the horizontal dash lines, as well as the dropout rate by gender and group gender composition. As shown previously, women in female-majority groups have the lowest dropout rate, and those in gender-balanced groups have the second-lowest dropout rate. Men in gender-balanced and female-majority groups also have a low dropout rate, but those in male-only groups have a higher dropout rate. Fixing the sample size of students—259 women and 361 men—and forcing the number of female-minority groups at zero, I vary the number of female-majority groups, from 1 to 85—with 2 groups a step. And for the rest of women, I assign all of them into gender-balanced groups, which pins down the number of gender-balanced groups. Accordingly, we know the number of men in female-majority and gender-balanced groups. The rest of men are assigned to male-only groups. For each assignment, I then calculate the average dropout rate for women and men—as plotted. We see that the optimal scenario is when women's dropout rate drops to a similar level as men's actual dropout rate (*-0.007*), which is also equal to men's counterfactual dropout rate. In this scenario, we have 29 female-majority groups, 86 gender-balanced groups, and 40 male-only groups.

Table B12: External Validity of the Treatment Effects of Minority Status on Women

			Observable Characteristics		
Baseline Effect	Φ(TE=0)	$\Phi \in$[1, 2]	high school	personality	baseline expectation
Outcome: Dropout					
0.077	2.59	[0.018, 0.047]	N	N	N
0.090	3.78	[0.036, 0.057]	Y	N	N
0.096	2.82	[0.022, 0.050]	Y	Y	N
Outcome: Overall Educational Expectation at Endline					
-0.465	14.87	[-0.434, -0.402]	N	N	N
-0.519	25.96	[-0.447, -0.429]	Y	N	N
-0.532	5.56	[-0.381, -0.297]	Y	Y	Y

Notes: The table presents two measures of external validity as proposed by Andrews & Oster (2019). Φ captures the degree of private information (unobservable characteristics) about the treatment effect (TE) used in participation decisions. $\Phi = 1$ means that private information plays no role at all, and $\Phi = 2$ means that private information and observable characteristics are equally important in the participation decision. Φ(TE=0) is the value of Φ corresponding to a zero average TE for the population. $\Phi \in [1, 2]$ corresponds to lower and upper bounds of TEs when Φ takes values between 1 and 2. The table examines the treatment effects of minority status on two outcomes: dropout and the overall educational expectation at endline. And I consider different sets of observable characteristics in the analysis. The basic effect is the impact of minority status without any controls estimated with the sample without missing values of the observable characteristics.

C Additional Tests and Analyses

C.1 Multiple Hypotheses Test

To account for multiplicity in hypothesis testing, I first control Family-Wise Error Rate (FWER) using the Romano-Wolf (RW) approach (Romano & Wolf, 2005a,b), which corrects the p-values based on a stepdown resampling method. I consider three groups of outcomes: main outcomes, educational expectations, and peer-to-peer interaction. For the coefficients of minority status and the interaction term between minority status and gender, Table C1 presents the clustered p-values, the resampling p-values, and the corrected p-values using the RW approach. The resample p-values are derived from 1,000 bootstrap resampling of individuals stratified at the level of year and study program. I find that the resample p-values are systematically smaller than clustered p-values. After the RW correction, most significant coefficients of minority status remain statistically significant at a similar significance level. The results are similar when I only focus on women and test the p-values of the coefficient of minority status.

Table C1: Romano-Wolf Multiple Hypothesis Correction (FWER)

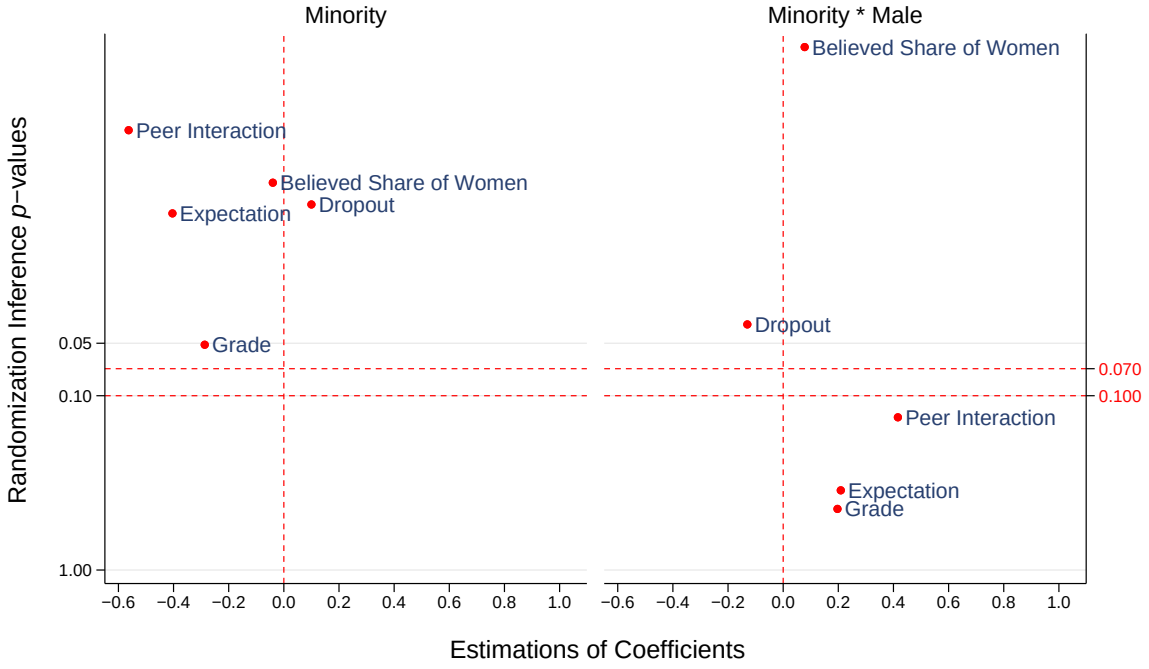
	Minority Status			Minority Status × Male		
	CL p	Resample p	RW p	CL p	Resample p	RW p
Panel A: Main Outcomes						
Dropout	0.097	0.033	0.093	0.163	0.063	0.168
Overall Expectation at Endline	0.017	0.007	0.022	0.235	0.110	0.201
Overall Peer Interaction	0.026	0.003	0.027	0.144	0.029	0.157
Exam Grade	0.987	0.982	0.982	0.394	0.256	0.256
Believed Share of Women at Endline	0.091	0.042	0.093	0.007	0.001	0.002
Panel B: Educational Expectations at Endline						
Overall Expectation	0.017	0.007	0.016	0.235	0.110	0.263
Expected Grade	0.006	0.006	0.012	0.182	0.083	0.216
Expected Probability Pass	0.297	0.166	0.166	0.679	0.586	0.586
Expected Degree	0.194	0.102	0.160	0.426	0.328	0.506
Panel C: Peer Interaction						
Overall Peer Interaction	0.026	0.003	0.009	0.144	0.029	0.061
Academic Interaction	0.041	0.006	0.011	0.075	0.009	0.025
Social Interaction	0.059	0.010	0.011	0.627	0.487	0.519
# Peers (Study > Once/Month)	0.011	0.001	0.002	0.063	0.010	0.025
# Peers (Social ≥ Once/Month)	0.033	0.007	0.011	0.545	0.390	0.519

Note: The table presents the clustered p-values (at the group level), the resampling p-values, and the corrected p-values using the Romano-Wolf (RW) multiple hypothesis correction.

As an alternative test, I control the False Discovery Rate (FDR) for all the main results—dropout, the overall educational expectation, the overall interaction with peers, and the believed share of women in the course at endline. Figure C1 shows the coefficients of minority status and the interaction term between minority status and gender (i.e., *Minority*Male*), the randomization inference p-values, and the corrected significance threshold. Without multiple-test

correction, all the coefficients of minority status are significant at the 10% level. I derive the corrected critical p-value for statistical significance (0.07 in this case), based on the step-up FDR procedure of Benjamini & Yekutieli (2001). According to the corrected threshold, the OLS-estimated impact of minority status on female dropout is no longer significant, while others remain significant. All results for men unsurprisingly remain insignificant.

Figure C1: Multiple Hypothesis Testing (FDR)



Note: The figure shows the distribution of estimation results (randomization inference p-values against coefficients), and the corrected critical p-value indicating the significance level—the *smile plot* based on the step-up FDR procedure of Benjamini & Yekutieli (2001) and `multproc` command in `STATA` (Newson & Team, 2003). Each estimation dot is labeled with the outcome variable. The left panel plots the coefficients of minority status, and the right panel plots the coefficients of the interaction term between minority status and gender. The estimations being tested are based on the pooled regressions with full controls, as shown in column (4) of Tables 3–6. Results are similar when testing the results with only basic controls. The upper horizontal line represents the corrected p-value threshold, and the lower line represents the original p-value threshold: 0.1. Estimations lying above the lower line are significant at 10% level before multiple test correction, but only estimations lying above the upper line remain significant after the correction.

C.2 Mediation Analysis

Table C2 first presents the overall correlations between mediators – educational expectation and peer-to-peer interaction – with dropout, in columns (1)–(2). Columns (3)–(6) show how the inclusion of mediators affects the treatment effect of minority status on dropout. From columns (1)–(2), we see that higher educational expectation at endline or more interaction with peers is correlated with a lower dropout rate. Columns (3)–(4) show that for the same sample, controlling for educational expectation decreases the effect of minority status from 5.4 to 4.2 percentage points, suggesting that lower expectation explains a fraction of the treatment effect.

Similarly, columns (5)–(6) suggest that lower interaction with peers explains part of the impact of minority status on dropout.

Next, I conduct a mediation analysis following Heckman & Pinto (2015), only for women. I analyze the mediating effects of educational expectation and peer interaction separately, because peer interaction was only measured in 2019 – including both endline expectation and peer interaction drastically reduces the sample size and makes the analysis noisy. Assume that women's dropout (y_i) decision is a linear function of minority status, the mediators, and other individual controls (X_i), as shown in Equation (6).

$$y_i = \alpha + \tau \text{Minority}_i + \sum_j \theta^j \text{Mediator}_i^j + \delta X_i + \varepsilon_i \tag{6}$$

The estimated coefficients $\hat{\tau}$, $\hat{\theta}^j$ are presented in columns (1)–(2) and (4)–(5) of Table C3. Columns (3) and (6) show the effects of minority status without including mediators, $\hat{\beta}$ as specified in Equation (1), for observations without missing values of the mediator. Additionally, I store the estimated effects of minority status on each mediator j: $\hat{\beta}^j$. Finally, the fraction of treatment effect explained by mediator j is share^j:

$$\text{share}^j = \frac{\hat{\theta}^j \hat{\beta}^j}{\hat{\beta}} \tag{7}$$

Figure C2 plots the estimated shares of treatment effects explained by different mediators. Each bar represents one set of mediators analyzed: (i) the overall educational expectation at endline, (ii) three dimensions of educational expectations at endline, (iii) the overall peer-to-peer interaction, and (iv) academic and social interaction with peers. I find that the overall educational expectation explains 16% of the impact of minority status on female dropout, while peer-to-peer interaction explains about 14% of the impact. When looking at lower-level expectations, results suggest that the expected grade has the largest contributing effect – 7.6%. That implies the pessimism about short-run educational outcomes, rather than longer-term outcomes, is more malleable. Regarding peer-to-peer interaction, I find that social interaction takes a much larger fraction of the treatment effect (9.6%) than academic interaction (3.8%).

The mediation analysis presented above rely on a few assumptions besides the basic linearity assumption: (i) unmeasured inputs (intermediate outcomes) affected by minority status are independent of the measured inputs, (ii) the contributing effects of mediators (θ^j) are the same for the treatment (minority) and control (non-minority) group, and (iii) the effects of individual controls on dropout do not differ between the treatment and control group. These assumptions that I can not completely validate, accompanied with the low statistical accuracy of results presented in Table C3, imply the need to take the analysis with caution. Another limit of the mediation analysis is that I do not have complete measures of educational expectations or peer-to-peer interactions, as some students do not fill out the endline survey. Students who drop out

are less likely to take the survey. Their under-representation in the mediation analysis suggests the estimated fractions are likely to be biased.

Table C2: Mediators and All Students' Dropout

	(1)	(2)	(3)	(4)	(5)	(6)
	Non-missing Expectation			Non-missing Interaction		
DV: Dropout						
Minority		0.054	0.042		0.074	0.067
		(0.041)	(0.040)		(0.062)	(0.063)
Educational Expectation	-0.041		-0.039			
	(0.019)**		(0.019)**			
Peer-to-Peer Interaction				-0.028		-0.026
				(0.010)***		(0.010)***
Observations	413	413	413	348	348	348
R-squared	0.048	0.024	0.053	0.049	0.046	0.054

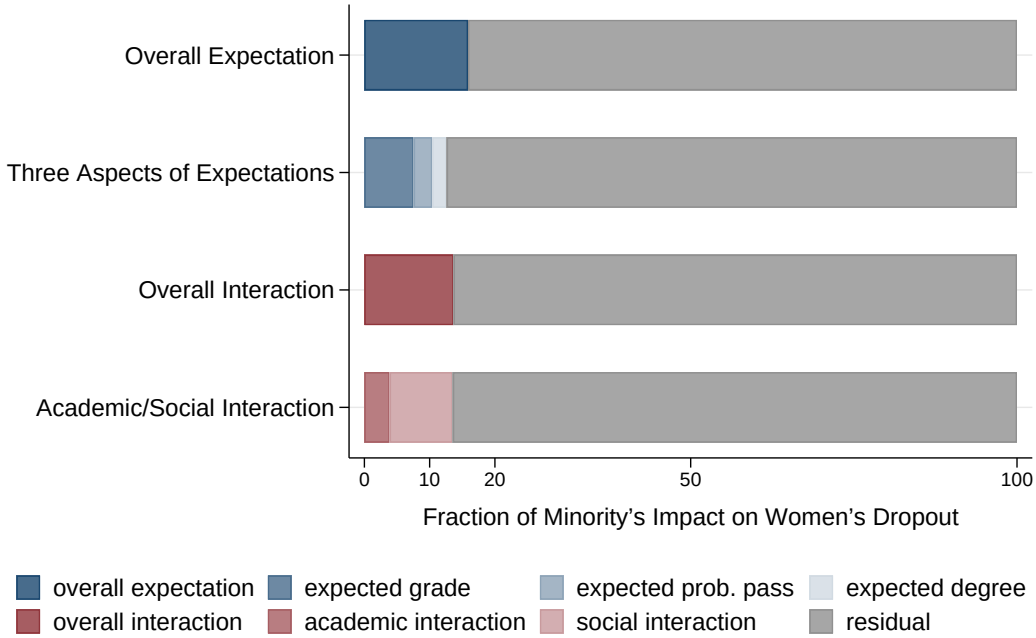
Note: All columns use OLS regressions, include both male and female observations, and control for gender, major, course-retaking, as well as baseline expectation and year fixed effects if applicable. Regressions for peer-to-peer interaction use only observations in 2019. Columns (1)–(2) examine the correlation between educational expectation/peer interaction and dropout. Column (3)–(4) estimate the impact of minority status for observations with measurements of expectation, with or without controlling for expectation. Columns (5)–(6) estimate the impact of minority status for observations with peer interaction measurements, with or without controlling for peer interaction. Standard errors are clustered at the group level. $*p < .1, **p < .05, ***p < .01$.

Table C3: Mediation Analysis for Women's Dropout

	(1)	(2)	(3)	(4)	(5)	(6)
	Non-missing Expectation			Non-missing Interaction		
DV: Women's Dropout						
Minority		0.079	0.067		0.114	0.098
		(0.049)	(0.048)		(0.089)	(0.092)
Educational Expectation	-0.033		-0.028			
	(0.024)		(0.024)			
Peer-to-Peer Interaction				-0.035		-0.029
				(0.019)*		(0.020)
Observations	179	179	179	146	146	146
R-squared	0.083	0.083	0.096	0.038	0.041	0.048

Note: All columns use OLS regressions, and control for major, course-retaking, as well as year fixed effects if applicable. Regressions with peer interaction use only 2019 observations. All mediators use sample-standardized measures. Standard errors are clustered at the group level. $*p < .1, **p < .05, ***p < .01$.

Figure C2: Fractions of Minority Status's Effect on Female Dropout Explained by Mediators



Note: The graph plots the proportions of the effect of minority status on female dropout explained by educational expectation and peer-to-peer interaction. Each bar represents one set of mediators considered.

D Ethical Considerations

In this paper, I show that women assigned to female-minority groups are more likely to drop out of an introductory economics course than women in other groups. In order to assess the ethics of this experiment, it is necessary to think about students' counterfactual decisions and behaviors in the absence of my experiment. The experiment took place in a first-semester course at a Swiss university. The setting has three key features, even without my experiment: 1) students have a demand for new peers and they do study with each other for the course; 2) the course is male-dominated and it is common that students form male-dominated social networks by themselves; 3) a high fraction of students drop out of the course during the semester.

1) Demand for new peers. Based on survey data collected in 2017—before I conducted the experiment, about 50% of students reported that they studied with others for the course, and more than 65% of the study mates were new peers that they met in the university.

2) Male-dominated environment. The fraction of women taking the course is 30 to 40 percent, depending on the cohort. That means women are generally exposed to a male-dominated environment, and it is likely for students to form networks that are male-dominated. Based on students' self-reported social network data, I find that, without the experiment, about 45% of the networks that students formed by themselves are male-dominated (see Figure C3).

3) High dropout rate. A unique feature of the higher education system in Switzerland is that students with a high school degree can almost freely register for a study program at college. Partly because of that, it is also common that students drop out or switch their majors, especially in the first semester at university. Historical administrative records show that the overall dropout rate in the Economics course is always around 9%.

Given the above institutional background, my experiment did two things. First, it offered students with an opportunity of having a study group and meeting new peers. Second, for students who wanted a study group, I randomly assigned them to groups of four. Next, I discuss the ethics of the experiment from two perspectives: 1) how the provision of study groups affected students; 2) conditional on having a group, how the group composition affected students.

1) The provision of study groups. It is very unlikely that providing students with the opportunity of having a study group will harm them. When deciding whether to sign up for a study group, students were explicitly informed of the following: (i) the study groups are

provided for free; (ii) the sign-up decision and participation in group activities have no direct influence on the grade; (iii) if they sign up, they will be randomly assigned to a group with three other students. The voluntary sign-up procedure means that only students with a demand for random study mates registered for study groups and they would very likely find a study mate anyway, with or without the experiment. Furthermore, if they did not like the assigned study mates, they still had the option of finding their own study mates. For students who preferred to study by themselves or had their own study mates, the experiment hardly affected them.

Instead, by meeting students' demand for new peers in a new environment, the experiment is potentially beneficial. Students can use the group to exchange information, discuss course materials, and work on problem sets together. More generally, social interaction with peers can enforce the sense of belonging at university and help students develop social-emotional skills—especially skills related to teamwork, which is ubiquitous and has become much more prevalent in the workplace and academic research (Devine et al., 1999; Lazear & Shaw, 2007; Rath & Wohlrabe, 2016).

Ex-post evidence also suggests that providing study groups does not harm students. If anything, study groups help lower the dropout rate of students, especially women. Figure C4 plots the overall dropout rate across cohorts from 2012 to 2019, with 95% confidence intervals. As the graph shows, the dropout rate fluctuates slightly over time. The years of 2018 and 2019, with the experiment in place, have a similar level of dropout rate as in previous years. The trend is similar after controlling for basic characteristics of students in the course: study program, gender, course retaking status, and nationality. If focusing on the experimental cohorts and comparing students with study groups to students without study groups, I find that study groups are associated with a lower dropout rate and higher educational expectations among women—as discussed in Section 5.3.

2) The random composition of groups. The key ethical concern regarding the random assignment of groups is that it creates female-minority groups, which may increase women's dropout rate—the hypothesis tested in this study. First, even though the tokenism concept famously hypothesizes the challenges faced by women in the minority (Kanter, 1977), rigorous testing of the concept is very limited (Zimmer, 1988). In effect, the literature shows conflicting evidence on whether it's beneficial for women (and men) to have more female peers. For example, Huntington-Klein & Rose (2018) and Bostwick & Weinberg (2018) show that women

are more likely to persist in male-dominated settings if they have more female peers. However, Zölitz & Feld (2020) and Brenøe & Zölitz (2020) find that women become less likely to choose highly-skilled and high-paying majors if they have more female peers. For men, Oosterbeek & Van Ewijk (2014) and Zölitz & Feld (2020) also find opposite effects of a higher share of female peers on men's educational outcomes.

Second, as mentioned above, even without the experiment, students also form networks or groups where women are in the minority. Actually, compared to the self-formed groups, randomly assigned groups are equally or less likely to be female-minority. Simulations before conducting the experiment show that if 30% of the students who sign up for study groups are women, the fraction of female-minority groups is about 45%. If 40% of the registered students are women, the share of female-minority groups is around 30%. In the realized assignment of my experiment, 34% of the study groups are female-minority groups—in contrast to 45% for endogenous groups. The results suggest that random assignment of study groups does not increase the risk of women being assigned to female-minority groups.

Taken together, by providing and randomizing study groups, the experiment is unlikely to and also did not harm students in the setting. If anything, it created additional benefits. Furthermore, randomizing peer groups is very common in educational instructions and economic research. For example, the previously cited studies by Oosterbeek & Van Ewijk (2014), Huntington-Klein & Rose (2018), and Zölitz & Feld (2020) are all based on randomly assigned workgroups, classes, or teaching sessions. Other influential studies by Sacerdote (2001), Carrell et al. (2013), and Booij et al. (2017) also exploit or conduct random assignment of dormitories, squadrons, or tutorial groups. The outcomes analyzed in these studies are also high-stakes decisions and behaviors, such as major choice, academic performance, as well as dropout.

Figure C3: The Distribution of Self-Formed Social Networks by the Share of Women

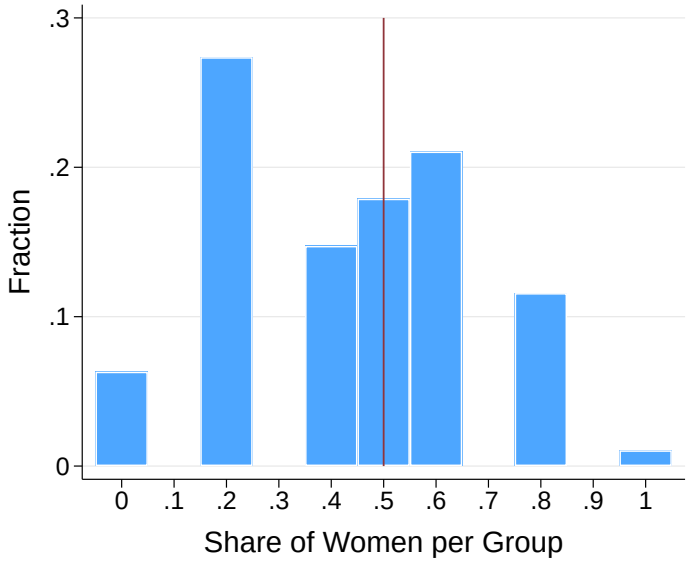


Figure C4: The Overall Dropout Rate by Cohort