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

Xiaoyue Shan*

Job Market Paper

October 27, 2020

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 as the minority gender in study 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 increases female attrition. Women who experience minority status 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 perpetuate itself and lead to a vicious cycle of female underrepresentation in male-dominated fields.

*I thank my supervisors, Ulf Zölitz and Michel Maráchal, 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, 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, as well as seminar attendants at University of Zurich, IZA, Graduate Institute Geneva, Jinan University, GEEZ, AYEW, ETH-Zurich, 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 the minority gender in a wide range of majors and professions, such as STEM and economics. Moreover, when climbing the career ladder, their minority status becomes more and more salient as some women decide to leave these fields—a phenomenon commonly known as the "leaky pipeline".[1] For example, only about a third of students in economics are women in the United States; the share further drops to 28% and 14% for assistant and full professors (Lundberg, 2018). The leaky pipeline is concerning because it implies the low presence of women in highly-skilled and highly-paid professions, especially in high-ranking positions. That further implies a gender pay gap and potential misallocation of talent across sectors (Blau and Kahn, 2017; Hsieh et al., 2019).

This paper examines whether women's minority status causes them to "leak" out of male-dominated fields. I conduct a field experiment in an introductory economics course at a Swiss university by randomly assigning students to study groups of four. The economics course is compulsory for students majoring in business, economics, and informatics, or enrolled in related minor programs. Most students take the course in their first semester at college, a period when they meet new peers and form new friendships. Although the course is male-dominated overall, the random assignment creates exogenous variation in the group gender composition. Some women are assigned to female-minority groups and experience minority status, while other women become members of gender-balanced or female-dominated groups. To identify the impact of minority status on women, I compare those in female-minority groups to those in other groups.

Results show that minority status in the study group increases women's voluntary course dropout. 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, for men, assignment to male-minority groups does not significantly affect

[1] This paper uses *minority* status, instead of *token* status, to describe the underrepresentation of women, because the tokenism concept (Laws, 1975; Kanter, 1977) focuses more on the perfunctory intention of the dominant or majority group to include the *tokens* for the appearance of equality.

[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.

their dropout behavior. I further show that the effect of minority status on women is primarily a product of the skewed gender ratio rather than other peer characteristics, such as ability, personality traits, or preferences. The impact of minority status on women's dropout remains large in magnitude and statistically significant even after controlling for a rich set of peer characteristics including high school achievement, the Big-Five personality traits, competitiveness, as well as gender attitudes.

To explain why women in the minority are more likely to drop out, I focus on two underlying mechanisms capturing the academic and social domain of college life: educational expectations and peer-to-peer interaction.[3] Educational expectations are students' expectations about future educational outcomes and capture their self-confidence about 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. 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 less confident 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; Niederle and Vesterlund, 2011; Buser et al., 2014; 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 and Ryan, 2008; Spencer et al., 2016). Being the minority gender 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 stereotypically male 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

[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.

to their social exclusion, women may feel that they belong less to the field 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 exhibit more homophily than women. Given three randomly assigned peers, men interact more frequently with male peers than female peers. Men also value male peers more than female peers. In contrast, women's preference for peers depends less on peers' gender.

Finally, I ask 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—personality traits, socioeconomic preferences, and gender stereotypes. In contrast, women with low math achievement or exhibiting typically male characteristics are not significantly affected minority status. The findings suggest that economics as a male-dominated field is losing the highly-skilled and typically-female women, further implying a loss of talent and diversity.

This paper contributes to the literature studying women in male-dominated settings from two perspectives. First, regarding why women are underrepresented, this paper shows that women's minority status can exacerbate their low presence in male-dominated fields and create a vicious cycle of female underrepresentation. This is the first study using a field experiment to examine whether and why minority status causes female attrition. Second, regarding how to increase women's representation, the findings of this paper suggest that small groups where women are not in the minority can serve as a light-touch and cost-effective policy tool to keep more women in male-dominated settings.

Existing studies in the literature can also be broadly summarized from the two perspectives. In terms of why women are underrepresented, studies show that it can be a result of both individual factors and contextual factors. Individual or internal factors include preferences, personality traits, beliefs, and gender stereotypes (Kamas and Preston, 2012; Wiswall and Zafar, 2015; Bian et al., 2017; Wiswall and Zafar, 2018), while contextual factors include teachers,

families, and peers (Bettinger and Long, 2005; Lavy, 2008; Cheng et al., 2017; Anelli and Peri, 2019; Brenøe and Zölitz, 2020; Zölitz and Feld, 2020). In terms of how to increase women's presence and reduce their attrition, recent studies find that policies targeting role models, mentors, information, and gender bias can be effective (Blau et al., 2010; Carnes et al., 2015; Avilova and Goldin, 2018; Porter and Serra, 2019).[4]

More closely, this paper relates to four recent studies investigating the impact of class or cohort gender composition on women's educational persistence in male-dominated environments. Huntington-Klein and Rose (2018) show that women with more female classmates at the U.S. Military Academy are more likely to advance to the next year. Similarly, Griffith (2010) and Bostwick and Weinberg (2018) find that female graduate students in STEM fields are more likely to persist if they have a higher percentage of female peers. However, Oosterbeek and Van Ewijk (2014) find no significant impact of gender composition in teaching sessions on women's dropout from business or economics.[5]

This paper differs from these four studies in two important ways. First, because these studies primarily rely on administrative data, they provide little evidence on why peers' gender composition influences dropout or not. In contrast, this paper helps to open up the "black box" of gender peer effects by combining administrative data and rich survey data. My results suggest that academic self-confidence and social integration are two potential channels through which peer gender affects dropout behavior. Second, this paper focuses on the gender composition in small peer groups rather than large classes or cohorts.[6] Small groups capture the effects of closely connected peers instead of a general environment. More importantly, small groups can serve as a light-touch policy tool, while classes or cohorts are in practice difficult to manipulate. Given a male-dominated class or cohort, small subgroups in which women are not in the minority

[4]More broadly, this paper builds on a large body of literature studying the effects of peers' gender and ability composition on educational outcomes. For example, studies show that peers' gender composition impacts academic achievement in school (Lavy and Schlosser, 2011; Hill, 2015; Gong et al., 2019) and university (Ficano, 2012; Schneeweis and 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.

[5]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.

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

can even counteract the effect of an unbalanced environment.

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 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.[7] 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,[8] 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.

[7] 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.

[8] 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".

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.[9] 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 frequency. 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.[10] 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

[9] 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 and 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.

[10] 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.

observe the final grade for the course, which ranges from 1–6. Students who drop out receive a 1—the lowest possible grade.[11]

3 Empirical Strategy

I use the following regression model to estimate the causal impact of minority status separately on women and men:

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

Y_{ig} is the outcome of student i in group g. The primary outcome variable is an indicator for course dropout. I also consider two intermediate outcomes (the two potential mechanisms): 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. X_{ig} is a battery of individual controls such as major, course-retaking status, personality traits, high school grades, and experiment year. ε_{ig} is the residual term. I show both randomization inference p-values and clustered standard errors at the study group level for most results.

Because study groups are randomly assigned, β identifies the causal effect of minority status on women or men. To confirm that the randomization was successful, I test whether individual characteristics at baseline and survey responses are balanced between students assigned as the minority gender and students not assigned as the minority gender. Table 2 presents the coefficients of minority status and the clustered standard errors separately for women and men. Except for extraversion, minority status does not significantly predict any of the baseline characteristics or survey completion. It is not a big threat for identification because extraversion does not significantly predict any of the outcome variables. As Section 4 shows, the estimation

[11] 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.

results are also robust to controlling extraversion and other personality traits.

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 baseline characteristics 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. Appendix Table B1 presents the F-statistics and p-values of the tests. 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 and significant at the 10% level. For men, in contrast, group gender composition does not significantly affect their dropout.

Table 3 shows the impact of minority status on course dropout, separately for women and men. The dependent variable is an indicator for dropout from the course, which is compulsory for students' registered major or minor programs. Panel A uses ordinary least squares (OLS) regressions, while Panel B uses probit regressions and reports the estimated marginal effects of minority status. All regressions include the basic controls—student major, course-retaking status, and year fixed effects. Regressions in columns (2) and (5) also control for students' high school background, and finally, regressions in columns (3) and (6) control for the Big-Five personality traits.

Results show that minority status in the study group significantly increases the female dropout rate. Overall, women assigned as the minority gender 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. Results using OLS and probit regressions are also very similar. The effect size is huge, considering that the average dropout rate of women is about 13%. By contrast, I do not find a significant effect of minority status in the study group on men's dropout behavior.

Appendix Table B2 shows that on average, one additional female peer in the study group lowers women's dropout rate by 4 percentage points. However, given the strong non-linearity as shown in Figure 3, the linear-model is not a good choice to analyze the effect of group gender composition on dropout.[12] Appendix Table B3 also examines how minority status affects students' final grades for the course. If a student drops out, the grade is simply point 1, the lowest grade. Both women and men have lower grades when assigned as the minority gender in a study group, but the differences are not significant at conventional significance levels. Overall, women respond to their minority status to a greater extent than men.

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 4 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 4 presents the main results: the impact of minority status on the overall educational expectation of women and men at endline. I find that minority status lowers women's expectation for future academic achievement by about 0.42 standard deviations. The effect is

[12] As Appendix Table B2 presents, compared to minority women, those assigned to gender-balanced, male-minority, and female-only groups all have a relatively lower dropout rate. Women assigned to male-minority groups have the lowest dropout rate. They are 14% less likely to drop out than women in female-minority groups—the gap is significant at the 5% level.

statistically significant at the 5% level. The effect size and significance level are robust to controlling for high school background, personality traits, and the overall educational expectation at baseline. For men, minority status also lowers the overall educational expectation, but the effect is much smaller in magnitude and is insignificant.[13]

Next, in Table 5, I examine the impact 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, and the highest academic degree expected to be obtained. Columns (7)–(8) focus on 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 9 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 this impact reflects. Lower expectations can be a result of rational assessment or underestimation of one's own ability, or a result of social comparison to a certain reference group. Expectations may also reflect students' aspirations. I analyze the potential underlying drivers of lower 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 gap between the expected grade and the believed grade for the average woman/man in the previous cohort, as a proxy for comparison with same-gender or opposite-gender peers; and (3) the aspired minimal grade for the course, as a proxy for aspiration.[14] I examine the impact of minority status on the four outcomes, separately for women and men, and plot the coefficients in Figure 5.

Results show that women who experience minority status significantly underestimate their

[13]Similarly, I conduct a detailed comparison across different gender compositions in Appendix Table B4 for women. Results show that compared to women in female-minority groups, those in other groups all have higher educational expectation, and the gap between female-minority groups and female-only groups is the largest. On average, one additional female peer raises women's overall expectation by about 0.15 standard deviations.

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

ability and have lower grade aspirations. Meanwhile, they expect (insignificantly) lower grades for themselves when compared to the average man or woman. For men, those in male-minority groups also predict grades lower than what they actually achieve and have lower aspirations for the grade, but to a lesser extent than women—the effects are also insignificant. Interestingly, minority status has an especially strong social comparison effect on men: those in male-minority groups expect significantly lower grades for themselves relative to the average man or woman.

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 6 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.[15]

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. Again, I separately look at women and men. Results show that women assigned to female-minority or male-dominated groups report significantly less interaction with group peers. In contrast, minority status does not significantly lower men's overall interaction with peers. Table 7 separately looks at academic and social interaction with study group peers. 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.

In Appendix Table B5, I examine the effect of minority status on the number of peers with frequent academic or social interaction rather than the average interaction with all three peers. Two students have *frequent* academic interaction if they study together more than once per month; they have *frequent* social interaction if they meet for social events at least once per

[15] 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.

month. Similarly, I find that women who experience minority status are less likely to frequently meet peers for academic or social activities.[16]

4.4 Why Are Women Isolated in Male-Dominated Groups?

Having shown that women in the minority interact less frequently with their peers, I continue by asking 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 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.[17] The variable of interest, $\text{Female}^p = \{0, 1\}$, indicates whether peer p is female or male. δ_i controls for individual fixed effects and I cluster standard errors at the individual level. Note that I use students in gender-mixed groups for analysis, because peer gender is homogeneous in female-only or male-only groups.

Table 8 presents the results. Columns (1)–(2) examine the intensity of interaction with peers, and columns (3)–(4) look at the willingness to pay for peers—all outcome variables are standardized. Results suggest that men have a salient preference for same-gender peers: given three randomly assigned group members, men interact much more frequently with male peers than with female peers. They are also willing to pay significantly more for male peers. In contrast, women's preferences are more gender neutral. The results suggest that women's

[16] In Appendix Figures B2 and B3, 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 very low. Overall, I find that women in female-minority groups report a larger fraction study mates from out of the group, and these self-selected study mates are more likely to be female. The findings suggest that women assigned to female-minority groups have a demand for female peers.

[17] 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.

exclusion seems to be driven by men preferring to interact with other men rather than women's preferences.[18]

4.5 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 educational expectations and peer-to-peer interaction as two potential mediators of the impact of minority status on dropout. Figure 7 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 statistically 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 and 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.

[18] Note that students report their willingness to pay only for peers with whom they have had effective interaction. If a student reports to have never interacted with a peer, I replace the WTP with zero. In Appendix Table B6, I show that the results are very similar quantitatively and qualitatively if I only include pairwise observations with non-missing values of WTP.

5 Heterogeneity and Discussions

5.1 Robustness

Peer Gender or Other Peer Traits? I first ask whether the observed effects of minority status are only about gender or whether they are instead driven by other peer characteristics. For instance, women with three male peers can be more pessimistic due to the skewed gender composition or because the male peers are, on average, more competitive, self-confident, or have stronger gender stereotypes. This is an important potential concern because some peer traits such as conscientiousness and agreeableness do have significant effects on women's dropout (see Appendix Figure B7).

To formally answer the question, I check how the estimated coefficient of minority status changes when controlling for other peer traits in the regression.[19] Figure 8 shows the effects of minority status on women's dropout with and without peer traits as controls. I only include observations with non-missing peer traits so that the change in coefficient can not be driven by changes in the estimation sample. I first control for each set of the following peer traits separately: high school background, preferences and gender attitudes, and personalities. Finally, I control for all the peer traits. Results show that the impact of minority status on women's dropout does not shrink, but actually increases, as other peer characteristics are included. The impact also generally remains statistically significant. These results suggest that the impact of minority gender is not driven other observable characteristics correlated with peer gender.

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). Appendix Table B7 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.

[19] More specifically, I control for the average value of peers' characteristics, the leave-own-out means.

Multiple Test. Appendix Section C.1 shows that most results survive the multiple test correction. After the Romano-Wolf correction (Romano and Wolf, 2005a,b), which controls the familywise error rate (FWER), all significant results remain statistically significant. For robustness, I also use the procedure of Benjamini et al. (2006), which controls the false discovery rate (FDR). Again, most significant results 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 the male-dominated field that women leave. For example, it implies a loss of talent if high-skilled women drop out, while losing low-skilled women is relatively less concerning. A study field becomes less diversified if women who stay display similar characteristics as men, while those who leave exhibit traditionally female characteristics. Therefore, 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 as reflected by personality traits, socioeconomic preferences, and gender attitudes.

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 B4 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 subgroups: those with below-median and above-median math grades. Figure 9 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 achieve-

ment 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*.[20]

Figure 10 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.

[20] 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 B6 plots the distribution of the female typicality index by gender.

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 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 11 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. The paper cannot provide causal evidence about that because study groups are offered on a voluntary basis. In effect, students with certain characteristics are more likely to sign up for a study group. As Appendix Figure B8 shows, women, first-semester students, and those with higher openness or lower conscientiousness are more likely to select into study groups.[21]

With the selection bias in mind, I analyze the correlations between group registration and relevant outcomes, controlling for a variety of individual characteristics. Table 9 examines whether women registered for study groups differ from those without groups in terms of dropout behavior and educational expectations. Results show that women in study groups are about 4 percentage points less likely to drop out of the course and their overall educational expectation at endline is higher by 0.13 standard deviations, even after controlling for high school background,

[21] The overall sign-up rate for study groups is about 30%. We do not randomize the offer of study groups due to the lack of power. Forcing all students into study groups also seems less relevant from a policy perspective. However, future studies randomizing the offer as well as the composition of study groups would be extremely valuable.

personality traits, preferences, gender attitudes, and the baseline expectation.[22]

Based on the suggestive evidence that study groups are overall beneficial for students, especially 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 group gender composition and gender, I replace female-minority groups with gender-balanced, female-majority, and male-only groups. Appendix Figure B9 shows how the counterfactual dropout rate changes with the number of gender-balanced or female-majority groups. I find that the optimal combination of study groups can decrease women's dropout rate by 2 percentage points while keeping men's dropout rate the same.

6 Conclusion

This paper shows that women can leave male-dominated fields as a result of their minority status. I conduct a field experiment in an introductory economics course by randomly assigning students to small study groups. Results show that women who experience minority status in study groups become 10 percentage points more likely to drop out of the course and their majors or minors. I find two potential reasons why minority status raises female dropout. Women assigned as the minority gender interact less frequently with their study group peers; women in the minority also form more pessimistic expectations about their future academic achievement.

The negative effects of minority status on women are driven by high-skilled women 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 personalities and preferences are more susceptible to the impact of minority status than those who exhibit typically male characteristics. The result suggests that due to women's low presence, those who stay in economics are more like their male peers in terms of social characteristics. I further show that the impact of minority status is mainly a product

[22] For women's dropout, the results are similar using OLS regressions. I find no significant difference between men with study groups and men without study groups. See Appendix Table B8 for details.

of the unbalanced gender ratio rather than other peer traits. The effect of minority status on female dropout remains significant and gets even more precisely estimated when controlling for various peer traits such as high school achievement, personality traits, preferences, and gender attitudes.

Given the current efforts to promote gender equality and close gender gaps in male-dominated fields like STEM and economics, this paper has important policy implications. To help retain more women in these fields, educational institutions should try to lower the salience of their minority status. For example, female-minority groups should be avoided whenever students are assigned to natural peer groups like study groups, joint projects, laboratories, or offices. The female dropout rate in my setting 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 and Shaw, 2007; Rath and Wohlrabe, 2016). This paper shows that group gender composition can affect women's expectations, interaction with peers, and their persistence in a male-dominated setting. Compared to previous studies examining cohort or class gender composition, the focus on small peer groups offers more scope for policy interventions. While it is in practice difficult or impossible to manipulate the composition of cohorts or classes, to change the composition of small groups is much more feasible.

References

Anelli, M. and 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., and Uysal, G. (2017). Does classroom gender composition affect school dropout? *Economic Research Forum Working Paper 1084*.

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

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

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

Bettinger, E. P. and 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., and 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., and 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. and 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., and 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., and Nolen, P. (2018). Do single-sex classes affect academic achievement? an experiment in a coeducational university. *Journal of Public Economics*, 168:109–126.

Bostwick, V. K. and 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. and 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., and 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., and Oosterbeek, H. (2014). Gender, competitiveness, and career choices. *The Quarterly Journal of Economics*, 129(3):1409–1447.

Buser, T., Niederle, M., and 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., and 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., and West, J. E. (2009). Does your cohort matter? measuring peer effects in college achievement. *Journal of Labor Economics*, 27(3):439–464.

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

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

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

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

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.

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

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

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.

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

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

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

Kamas, L. and 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. and 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. and 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.

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

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. and 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. and Team, A. S. (2003). Multiple—test procedures and smile plots. *The Stata Journal*, 3(2):109–132.

Nguyen, H.-H. D. and 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. and Vesterlund, L. (2011). Gender and competition. *Annual Review of Economics*, 3(1):601–630.

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

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

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

Reuben, E., Wiswall, M., and 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. and 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. and Wolf, M. (2005b). Stepwise multiple testing as formalized data snooping. *Econometrica*, 73(4):1237–1282.

Rosenberg, M., Schooler, C., Schoenbach, C., and Rosenberg, F. (1995). Global self-esteem and specific self-esteem: Different concepts, different outcomes. *American Sociological Review*, pages 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. and 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., and Johnston, M. (1995). Measures in health psychology: A user's portfolio. causal and control beliefs. *Windsor, UK: Nfer-Nelson*, pages 35–7.

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

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

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

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

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

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. and 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*.

Figures and Tables

Figure 1: The Experiment Timeline

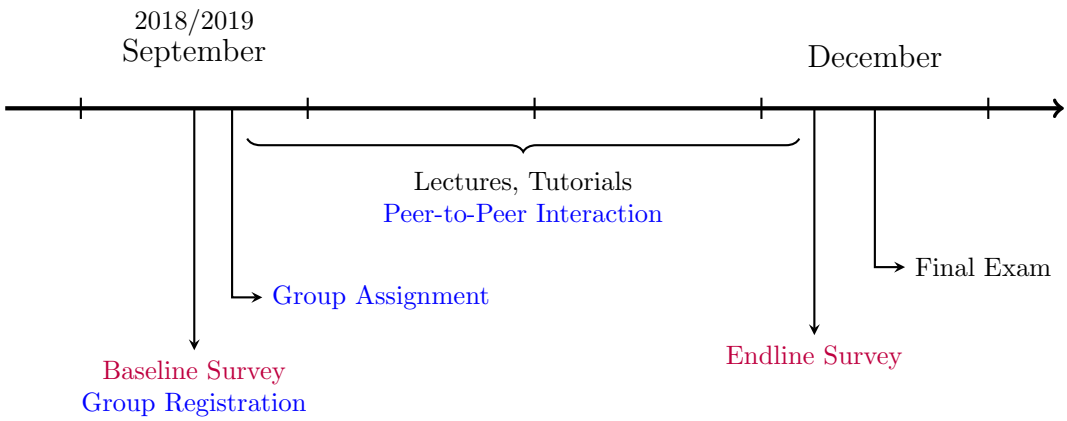


Figure 2: Pairwise Interaction and Willingness to Pay Within a Study Group

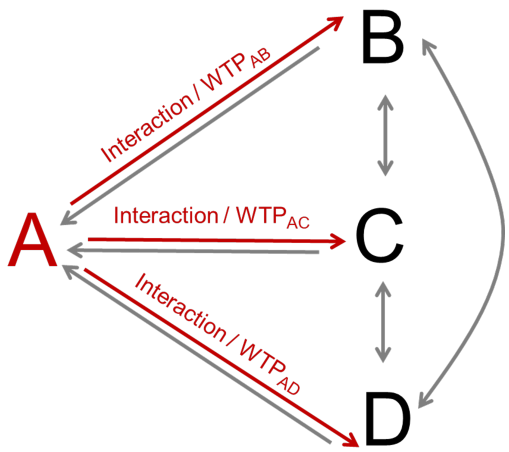


Table 1: Descriptive Statistics

	Women			Men		
	N	Mean	(SD)	N	Mean	(SD)
Panel A: Baseline Characteristics						
High School Math Grade	238	4.54	(0.79)	323	4.53	(0.86)
High School Language Grade	238	4.91	(0.54)	323	4.65	(0.57)
High School in German	241	0.85	(0.35)	325	0.86	(0.35)
Expected Grade	241	4.60	(0.49)	325	4.74	(0.51)
Expected Probability Passing First Year	241	75.3	(16.4)	325	76.6	(19.1)
Expected Highest Degree is MA	241	0.83	(0.38)	325	0.76	(0.43)
Expected Highest Degree is PhD	241	0.09	(0.28)	325	0.12	(0.33)
Overall Educational Expectation	241	-0.09	(0.94)	325	0.07	(1.04)
Panel B: Endline Characteristics						
Expected Grade	187	4.35	(0.66)	247	4.53	(0.61)
Expected Probability 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)

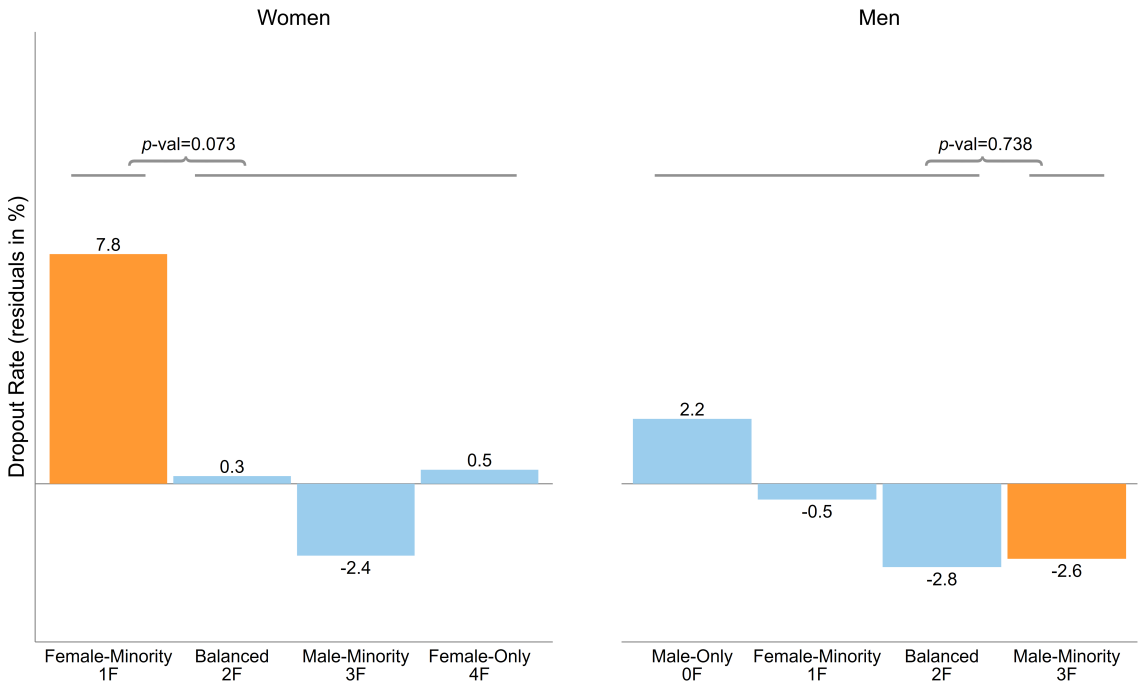
Note: *Overall educational expectation* is a combined index for answers to three questions: (1) the expected grade for the course, (2) the expected probability of passing the first year at the university, and (3) the expected highest academic degree. The index is created as follows: standardize the answer to each sub-question, 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?

	Women		Men	
	Minority Status	(clustered s.e.)	Minority Status	(clustered s.e.)
High School Math Grade	-0.078	(0.081)	0.106	(0.163)
High School Language Grade	-0.021	(0.052)	-0.135	(0.048)
High School in German	-0.071	(0.035)	0.044	(0.050)
Conscientiousness	0.082	(0.072)	-0.017	(0.276)
Extraversion	-0.283	(0.041)**	-0.008	(0.168)
Agreeableness	0.043	(0.070)	0.055	(0.053)
Openness	0.040	(0.156)	0.057	(0.185)
Neuroticism	0.135	(0.093)	0.126	(0.081)
Course Retaking	0.033	(0.025)	-0.071	(0.159)
Baseline Educational Expectation	-0.094	(0.051)	0.078	(0.052)
Baseline Survey Completion	0.026	(0.013)	0.035	(0.018)
Endline Survey Completion	-0.037	(0.075)	0.020	(0.104)

Note: The table tests whether minority status predicts baseline characteristics and survey responses of women and men. I regress each characteristic on the minority indicator, using OLS regressions. All regressions control for the fixed effects of study program and experiment year. For baseline characteristics and survey response, I cluster standard errors at the study program level because students have not interacted with their group members yet. For endline survey response, I cluster standard errors at the study group level to capture potential within-group correlations. For each outcome, I report the estimated coefficient of minority status and the clustered standard errors. $*p < .1, **p < .05, ***p < .01$.

Figure 3: Group Gender Composition and Dropout



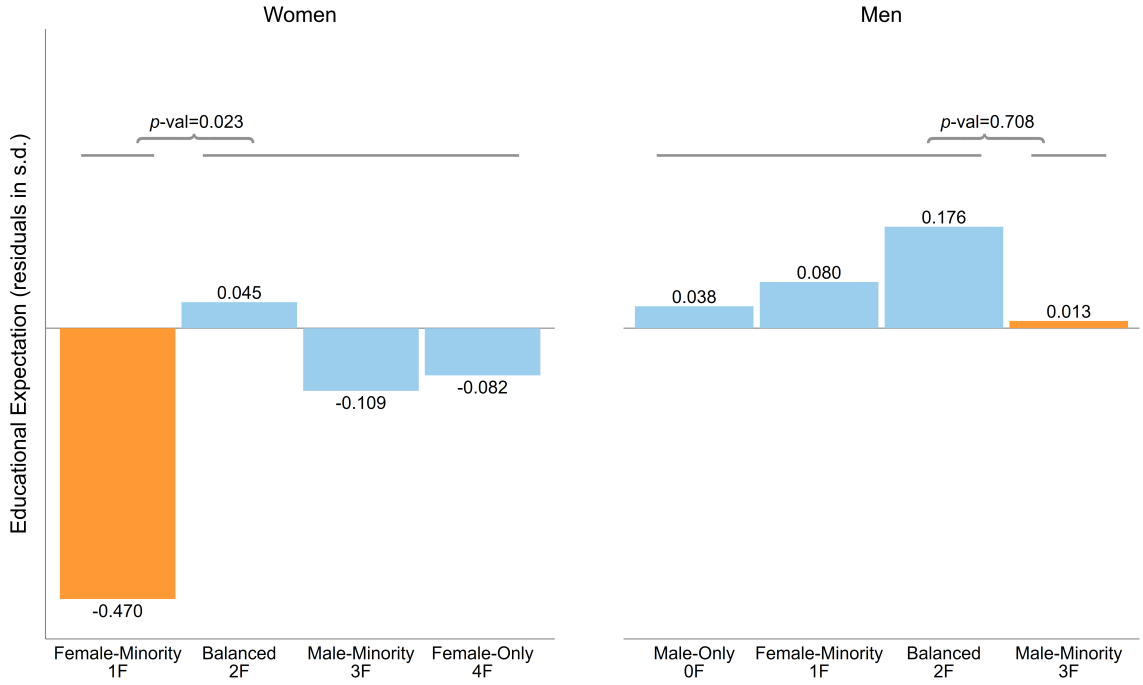
Note: The figure shows the residuals of dropout rate after controlling for assignment strata (year and study program) fixed effects by student gender and group gender composition. The gap between women in female-minority groups and women in other groups is about 8 percentage points. The raw dropout rate is 13% for women and 9% for men. Randomization inference p-values in the graph are derived from 1,000 replications of study group assignment within randomization strata.

Table 3: The Impact of Minority Status on Dropout

	(1)	(2)	(3)	(4)	(5)	(6)
		Women			Men	
DV: Dropout						
Panel A: OLS regressions						
Minority Status	0.101	0.096	0.105	-0.015	-0.011	-0.010
{RI p-value}	{0.068}*	{0.058}*	{0.038}**	{0.818}	{0.894}	{0.891}
(clustered s.e.)	(0.057)*	(0.057)*	(0.058)*	(0.061)	(0.065)	(0.067)
R-squared	0.031	0.082	0.114	0.029	0.046	0.062
Panel B: Probit regressions						
Minority Status	0.097	0.094	0.095	-0.013	-0.002	0.005
{RI p-value}	{0.056}*	{0.043}**	{0.036}**	{0.835}	{0.968}	{0.925}
(clustered s.e.)	(0.048)**	(0.042)**	(0.041)**	(0.054)	(0.058)	(0.057)
Pseudo R-squared	0.041	0.121	0.172	0.046	0.083	0.115
Basic controls	Y	Y	Y	Y	Y	Y
High school controls	N	Y	Y	N	Y	Y
Personality traits	N	N	Y	N	N	Y
Observations	259	238	235	361	323	319
R-squared	0.031	0.082	0.114	0.029	0.046	0.062
Mean of DV	0.127	0.113	0.111	0.094	0.081	0.082
SD of DV	0.334	0.318	0.314	0.292	0.272	0.274

Note: The table examines the impact of minority status on the dropout behavior of women and men. The dependent variable is an indicator for dropout. Panel A uses OLS regressions. Panel B uses probit regressions and reports the average 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—conscientiousness, extraversion, agreeableness, openness, and neuroticism. Clustered standard errors are in parentheses. Randomization Inference (RI) p-values in braces are derived from 1,000 replications of study group assignment within randomization strata (year and study program). $*p < .1, **p < .05, ***p < .01$.

Figure 4: Group Gender Composition and Educational Expectation at Endline



Note: The figure plots residuals of the overall educational expectation at endline after controlling for assignment strata (year and study program) fixed effects by student gender and group gender composition. Randomization inference p-values in the graph are derived from 1,000 replications of study group assignment within randomization strata.

Table 4: 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	-0.436	-0.422	-0.400	-0.032	-0.074	-0.099	-0.101
{RI p-value}	{0.028}**	{0.022}**	{0.023}**	{0.030}**	{0.896}	{0.746}	{0.659}	{0.658}
(clustered s.e.)	(0.172)**	(0.172)**	(0.169)**	(0.169)**	(0.199)	(0.195)	(0.197)	(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	179	179	177	247	234	233	231
R-squared	0.082	0.224	0.229	0.261	0.065	0.296	0.314	0.335
Mean of DV	-0.133	-0.113	-0.113	-0.108	0.101	0.121	0.122	0.124
SD of DV	0.961	0.962	0.962	0.966	1.019	1.014	1.016	1.020

Note: 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. Standard errors in parentheses are clustered at the group level. Randomization Inference (RI) p-values in braces are derived from 1,000 replications of study group assignment within randomization strata. $*p < .1, **p < .05, ***p < .01$.

Table 5: Minority Status and Different Aspects of Educational Expectations

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Grade		Prob. Pass Yr-1		Highest Degree		PhD Degree	
	Women	Men	Women	Men	Women	Men	Women	Men
Minority Status	-0.304	-0.168	-2.605	-0.832	-0.145	0.042	-0.100	-0.003
{RI p-value}	{0.020}**	{0.265}	{0.509}	{0.880}	{0.096}*	{0.774}	{0.040}**	{0.975}
(clustered s.e.)	(0.119)**	(0.102)	(3.655)	(4.706)	(0.074)*	(0.135)	(0.035)***	(0.085)
Observations	177	231	177	231	177	231	177	231
R-squared	0.293	0.273	0.191	0.234	0.293	0.228	0.245	0.223
Mean of DV	4.362	4.554	67.28	72.81	2.944	2.923	0.084	0.103
SD of DV	0.666	0.596	19.78	20.45	0.483	0.582	0.278	0.305

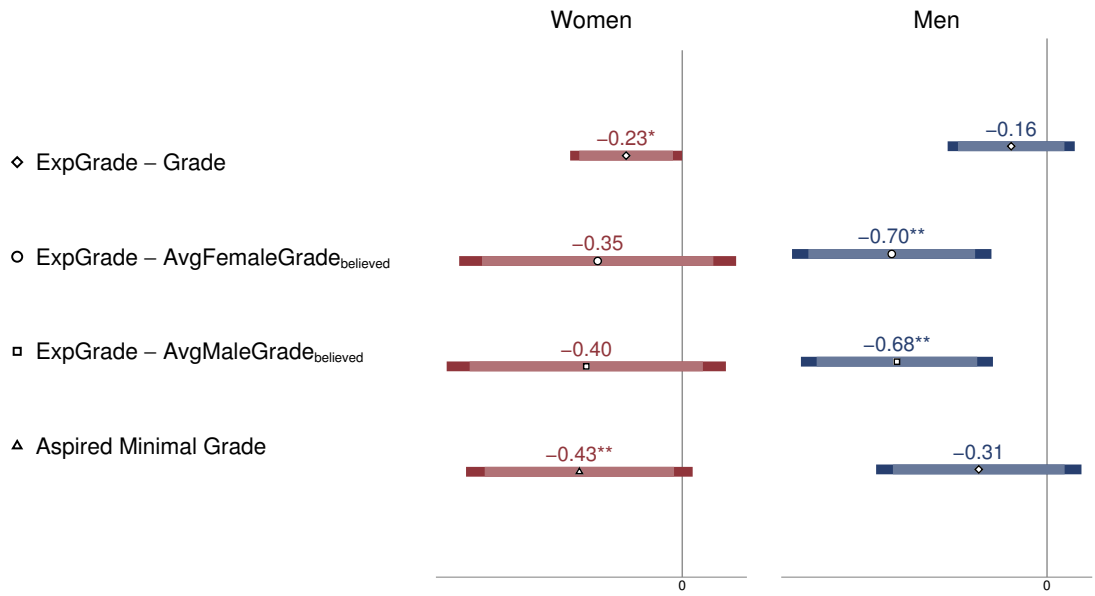
Note: The table shows the impact of minority status on different aspects of endline educational expectations, separately for women and men. The dependent variables are respectively: the expected grade for the course, the expected probability of passing the first year, the expected highest academic degree, and an indicator for whether the highest degree is a PhD. All columns use OLS regressions, and 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. Standard errors in parentheses are clustered at the group level. Randomization Inference (RI) p-values in braces are derived from 1,000 replications of study group assignment within randomization strata. $*p < .1, **p < .05, ***p < .01$.

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

	(1)	(2)	(3)	(4)	(5)	(6)
	Women			Men		
DV: the Overall Frequency of Interaction with Peers						
Minority Status	-0.540	-0.589	-0.693	-0.107	-0.127	-0.204
{RI p-value}	{0.030}**	{0.024}**	{0.011}**	{0.698}	{0.660}	{0.501}
(clustered s.e.)	(0.225)**	(0.247)**	(0.278)**	(0.175)	(0.178)	(0.190)
Basic controls	Y	Y	Y	Y	Y	Y
High school controls	N	Y	Y	N	Y	Y
Big-Five personality traits	N	N	Y	N	N	Y
Observations	146	134	131	202	191	187
R-squared	0.108	0.169	0.197	0.076	0.098	0.119
Mean of DV	-0.028	0.008	0.025	0.020	0.047	0.044
SD of DV	0.982	1.003	1.008	1.015	1.030	1.033

Note: Use OLS regressions. The DV is the standardized index for the overall interaction with peers. All regressions 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. Clustered standard errors in parentheses. Randomization Inference (RI) p-values in braces are derived from 1,000 replications of study group assignment within randomization strata. $*p < .1, **p < .05, ***p < .01$.

Figure 5: What the Impact of Minority Status on the Expected Grade Reflects



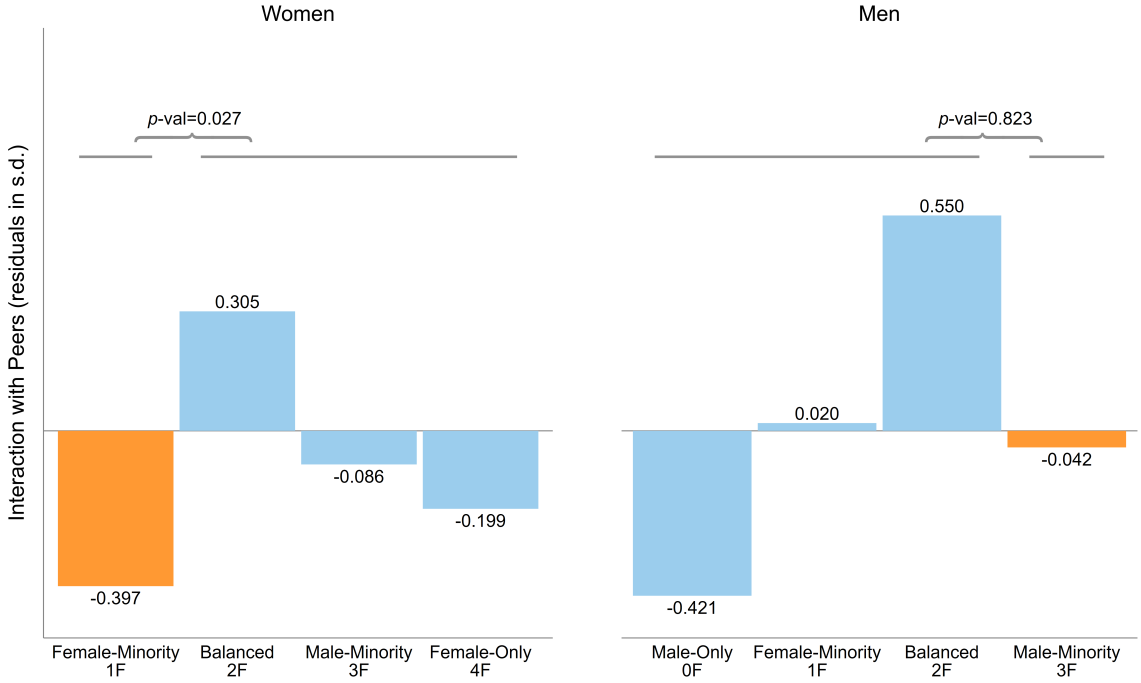
Note: The graph plots the estimated effects of minority status on: (1) the gap between the expected and realized final grade—a positive gap implies overestimation of own ability while a negative gap implies underestimation; (2) the gap between the expected grade and the believed grade for the average woman/man in the previous cohort—a positive gap implies that a student believes that he/she can achieve better than the average woman/man; and (3) the aspired minimal grade for the course, as a measurement of educational aspiration. All outcomes are sample-standardized, and measures (2)–(3) are available only in 2019. 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. *RI p <.1, **RI p <.05, ***RI p <.01. Randomization inference (RI) p-values in the graph are derived from 1,000 replications of study group assignment within randomization strata.

Table 7: Minority Status and Academic/Social Interaction with Peers

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Academic Interaction				Social Interaction			
	Women	Women	Men	Men	Women	Women	Men	Men
Minority Status	-0.498	-0.580	0.179	0.055	-0.464	-0.654	-0.370	-0.419
{RI p-value}	{0.033}**	{0.024}**	{0.550}	{0.865}	{0.090}*	{0.025}**	{0.153}	{0.157}
(clustered s.e.)	(0.234)**	(0.271)**	(0.281)	(0.289)	(0.221)**	(0.273)**	(0.179)**	(0.204)**
Additional controls	N	Y	N	Y	N	Y	N	Y
Observations	146	131	202	187	146	131	202	187
R-squared	0.068	0.115	0.043	0.105	0.109	0.267	0.097	0.116
Mean of DV	0.035	0.105	-0.026	-0.006	-0.085	-0.062	0.061	0.084
SD of DV	0.999	1.015	1.003	1.007	0.995	1.037	1.002	1.029

Note: Use OLS regressions. The dependent variable in columns (1)–(4) is the standardized index for academic interaction with peers; in columns (5)–(8), it is the index for social interaction with peers. All regressions control for student major and course-retaking status. Additional controls refer to high school grades, whether the first language is German, and the Big-Five personality traits. Clustered standard errors are in parentheses. Randomization Inference (RI) p-values in braces are derived from 1,000 replications of study group assignment within randomization strata. $*p < .1, **p < .05, ***p < .01$.

Figure 6: Group Gender Composition and Interaction with Peers (in 2019)



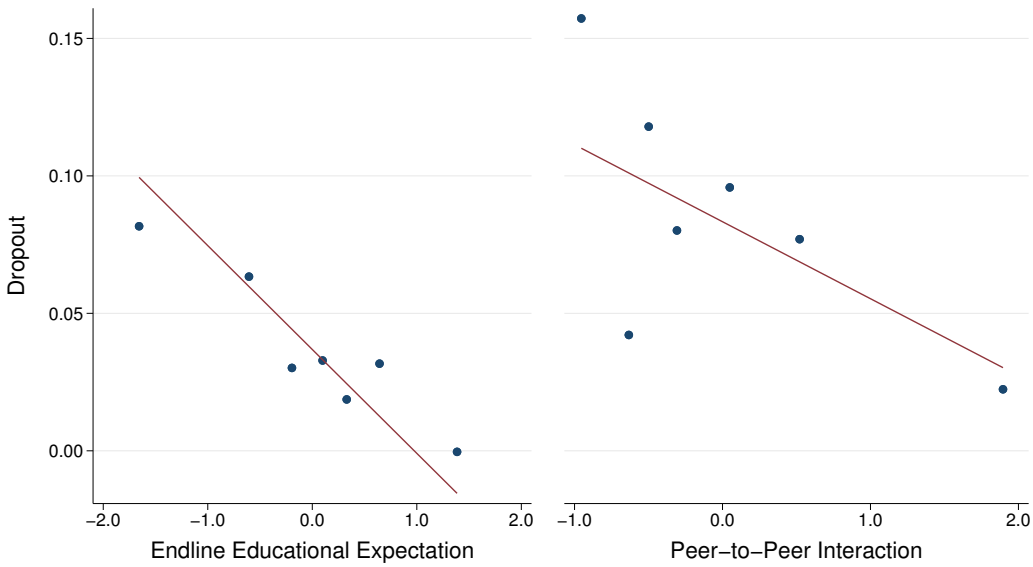
Note: The figure plots the overall intensity of interaction with peers after controlling for study program fixed effects by student gender and group gender composition. Randomization inference p-values in the graph, derived from 1,000 replications of group assignment within randomization strata.

Table 8: Gender Homophily in Pairwise Interaction and Evaluation

	(1)	(2)	(3)	(4)
	Frequency of Interaction with the Peer		Willingness to Pay for the Peer	
	Women	Men	Women	Men
Female Peer	0.016	-0.33	0.029	-0.359
	(0.066)	(0.109)***	(0.065)	(0.125)***
Pairwise observations	294	312	294	312
R-squared	0.834	0.699	0.887	0.531
Mean of DV	0.135	0.138	0.043	0.050
SD of DV	1.018	1.073	1.113	0.964

Note: The table examines how students interact with and evaluate female and male peers differently. The dependent variable in columns (1)–(2) is the standardized frequency of interaction with peers; in columns (3)–(4), the dependent variable is the standardized willingness to pay. All columns use pairwise observations and simple OLS regressions, and control for individual fixed effects. Standard errors are clustered at the individual level and are shown in parentheses. $*p < .1, **p < .05, ***p < .01$.

Figure 7: Correlations between Dropout and Intermediate Outcomes



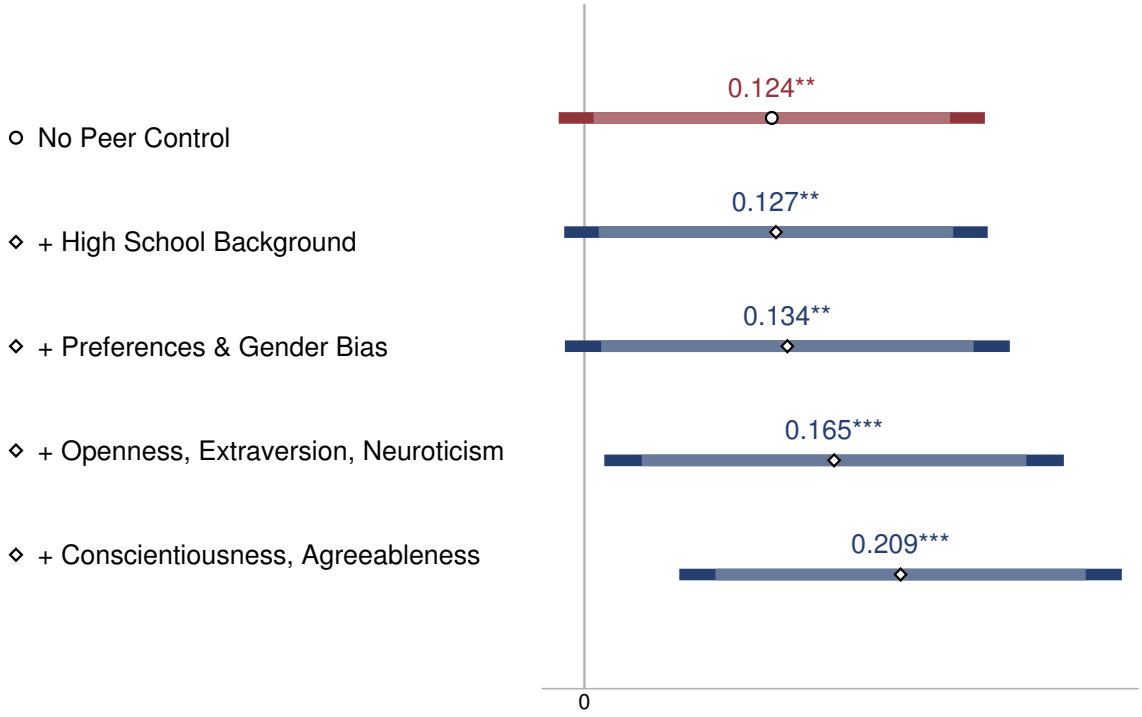
Note: The binned scatter plots show how dropout correlates with the two intermediate outcomes: the overall educational expectation at endline and the intensity of interaction with peers. Both plots control for gender, major, course-taking status, and year fixed effects. The slopes in the two sub-figures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

Table 9: Group Registration and Women's Dropout and Educational Expectations

	(1)	(2)	(3)	(4)
Panel A: Women's Dropout (Probit)				
Having A Study Group	-0.054**	-0.031*	-0.038**	-0.038**
	(0.023)	(0.013)	(0.017)	(0.018)
Observations	943	749	742	738
Pseudo R-squared	0.055	0.108	0.132	0.137
Panel B: Women's Educational Expectation at Endline (OLS)				
Having A Study Group	0.126*	0.142*	0.149***	0.126**
	(0.047)	(0.057)	(0.013)	(0.038)
Observations	586	518	515	513
R-squared	0.029	0.094	0.161	0.318
High school controls	N	Y	Y	Y
Big-Five personality traits	N	N	Y	Y
Additional individual controls	N	N	N	Y

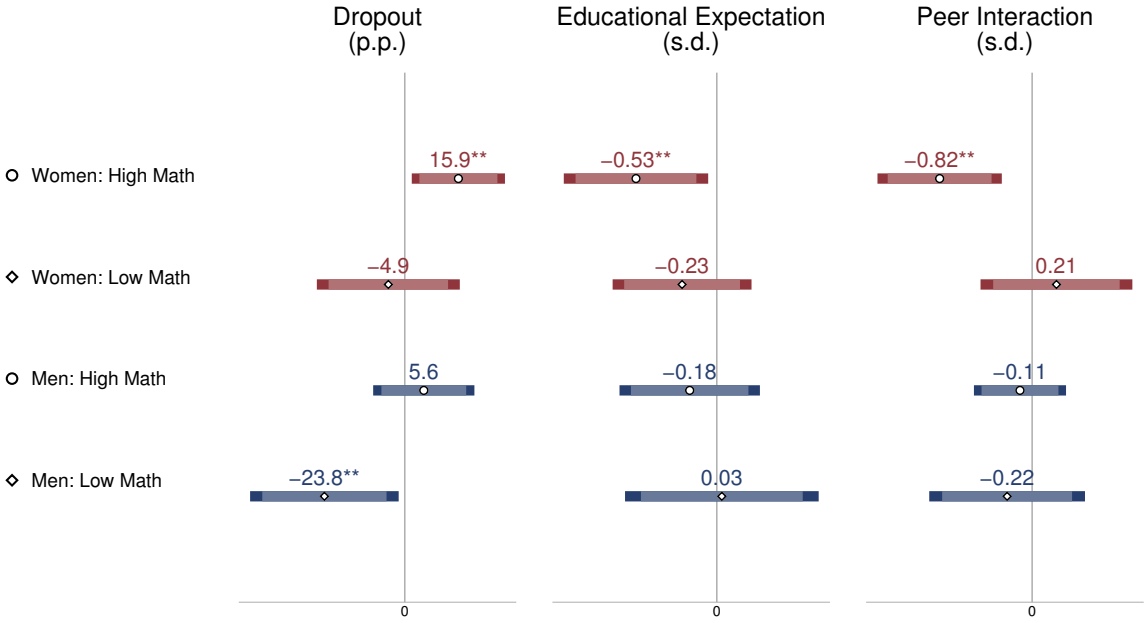
Note: The table examines whether women in study groups differ from those without study groups. All regressions control for major, course-retaking status, and year fixed effects. High school controls include high school math and language grades, and whether the first language is German. Additional individual controls include risk, patience, competitiveness, gender bias, and baseline expectation. Standard errors in parentheses are clustered at the major level. Panel A uses probit regressions and reports the average marginal effects. Panel B uses OLS regressions. $*p < .1, **p < .05, ***p < .01$.

Figure 8: The Impact of Minority Status on Female Dropout With/Without Peer Controls



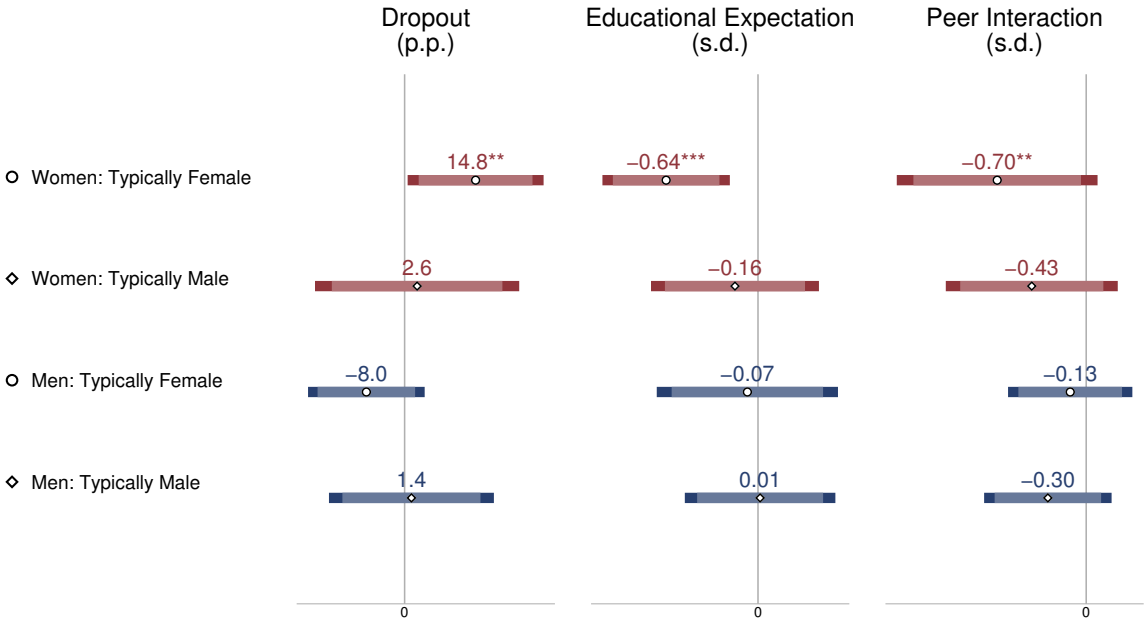
Note: The figure plots the estimated impact of minority status on women's dropout when controlling for different peer traits. The top row includes no peer traits; then I gradually include each of the following sets of peer traits: (1) high school grades and whether the first language is German, (2) risk preference, patience, competitiveness, and gender bias, (3) openness, extraversion, and neuroticism, and (4) agreeableness and conscientiousness. All regressions control for major, retaking status, and year fixed effects. 95% and 90% confidence intervals are based on standard errors clustered at the group level. *RI p <.1, **RI p <.05, ***RI p <.01. Randomization inference (RI) p-values in the graph are derived from 1,000 replications of study group assignment within randomization strata.

Figure 9: Heterogeneous Impact of Minority Status by High School Math Grade



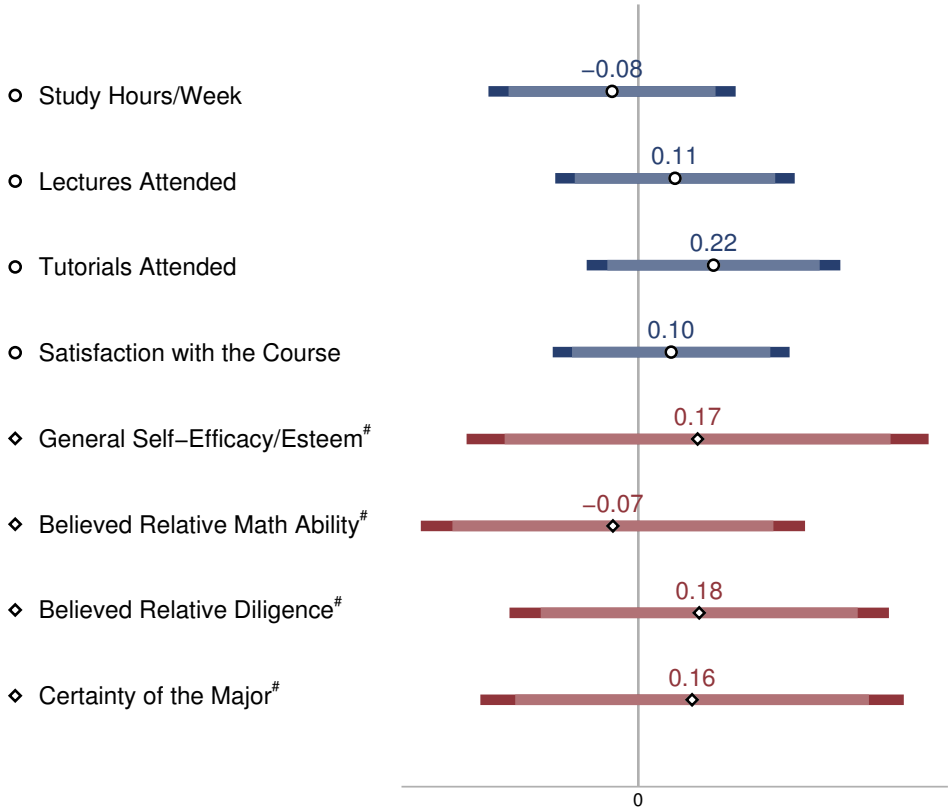
Note: The graph shows the estimated effects of minority status on three outcomes: dropout, endline educational expectation, and peer-to-peer interaction. *High Math* refers to math grades equal to and above 4.5 (the median), while *Low Math* refers to grades below 4.5. See Appendix Figure B5 for the 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. *RI p <.1, **RI p <.05, ***RI p <.01. Randomization inference (RI) p-values in the graph are derived from 1,000 replications of study group assignment within randomization strata.

Figure 10: Heterogeneous Impact of Minority Status by Gender Typicality



Note: 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. *RI p <.1, **RI p <.05, ***RI p <.01. Randomization inference (RI) p-values in the graph are derived from 1,000 replications of study group assignment within randomization strata.

Figure 11: Does Minority Status Affect Female Dropout Through Other Channels?



Note: 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.

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 $A^g \geq a^g$.[23] 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

[23] 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: $\Delta A^f_M < \Delta A^m_M \leq 0$.

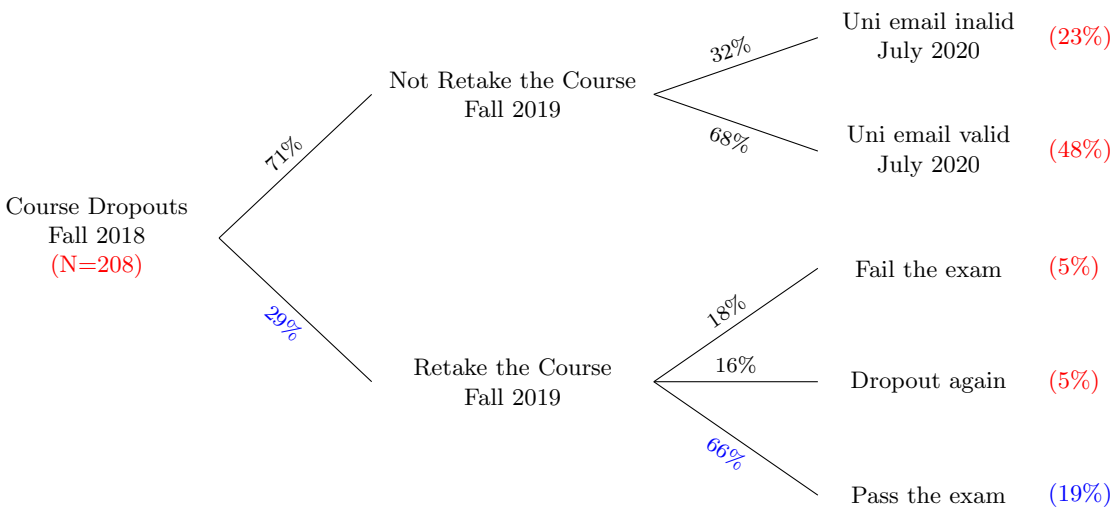
Proposition 2. *Given Proposition 1 and Hypotheses 1–2, minority status can increase the probability of dropout from college:*

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

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: $\Delta \Pr(\mathrm{d}^f)_M > \Pr(\mathrm{d}^m)_M \geq 0$.

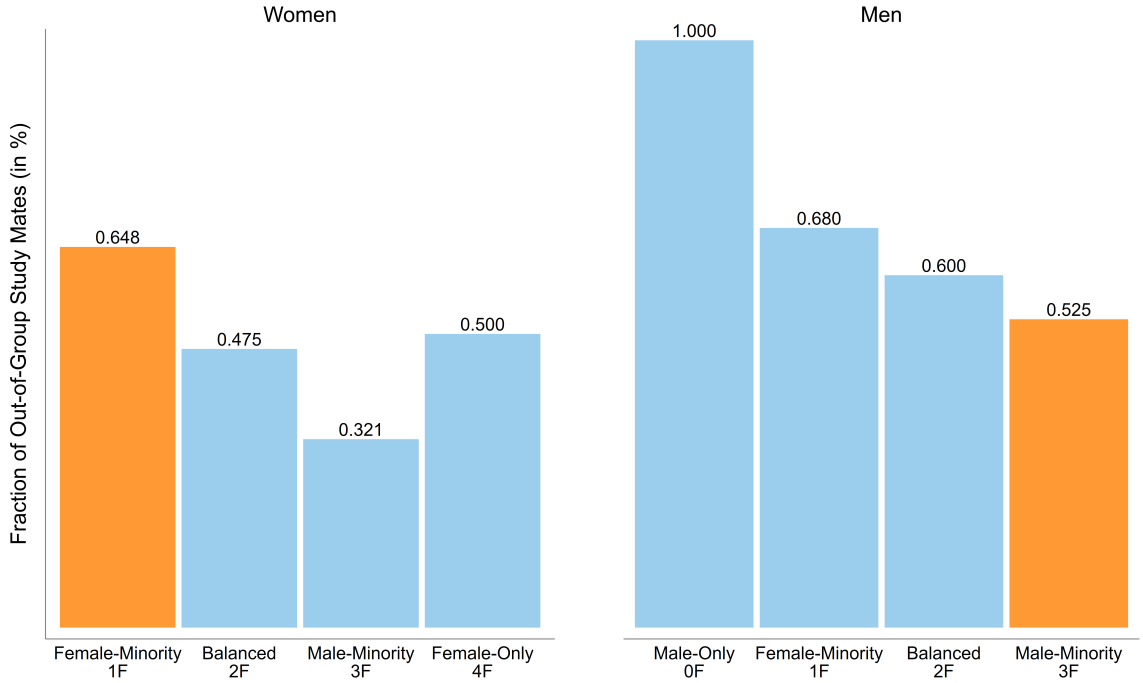
B Additional Figures and Tables

Figure B1: What Students Do After Dropping out of the Economics Course



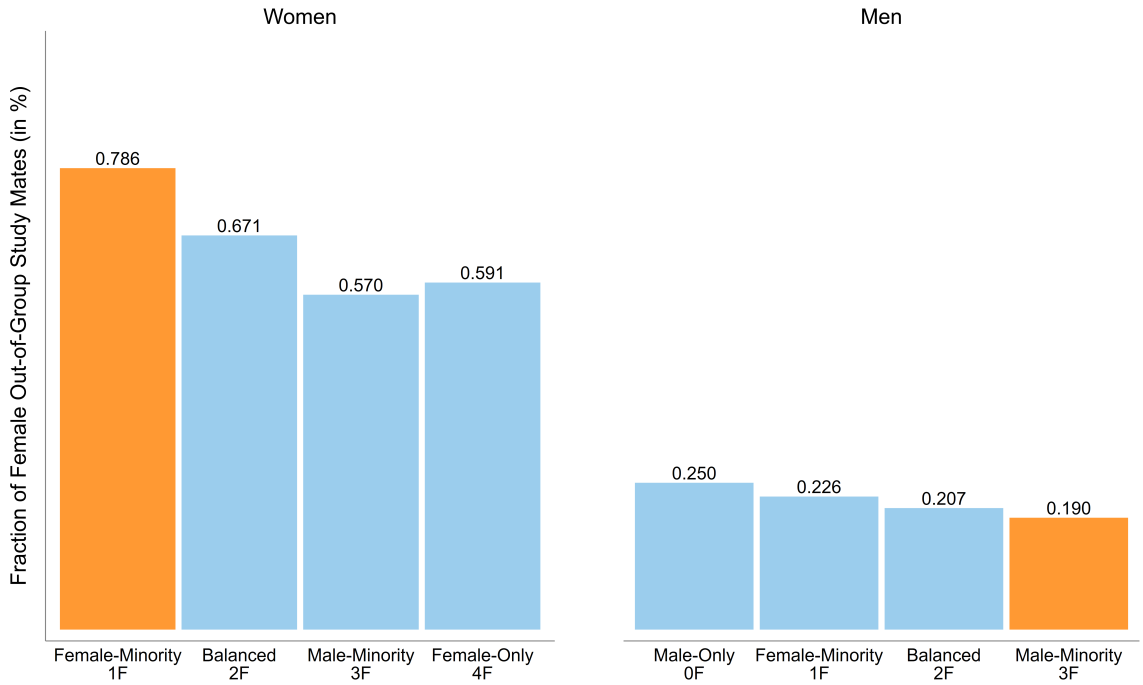
Note: The figure shows the different pathways that students typically take after dropping out of the economics course. I follow all dropout observations from the 2018 cohort: 208 students in total. Among them, 71% did not retake the course in 2019, suggesting that they almost permanently quit the course and the previously chosen majors or minors. By July 2020, 32% 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 two thirds of them manage to pass the exam. In the end, only 19% of the dropouts successfully passed the course, but at the cost of one additional year in college.

Figure B2: Gender Composition and Fraction of Out-of-Group Study Mates



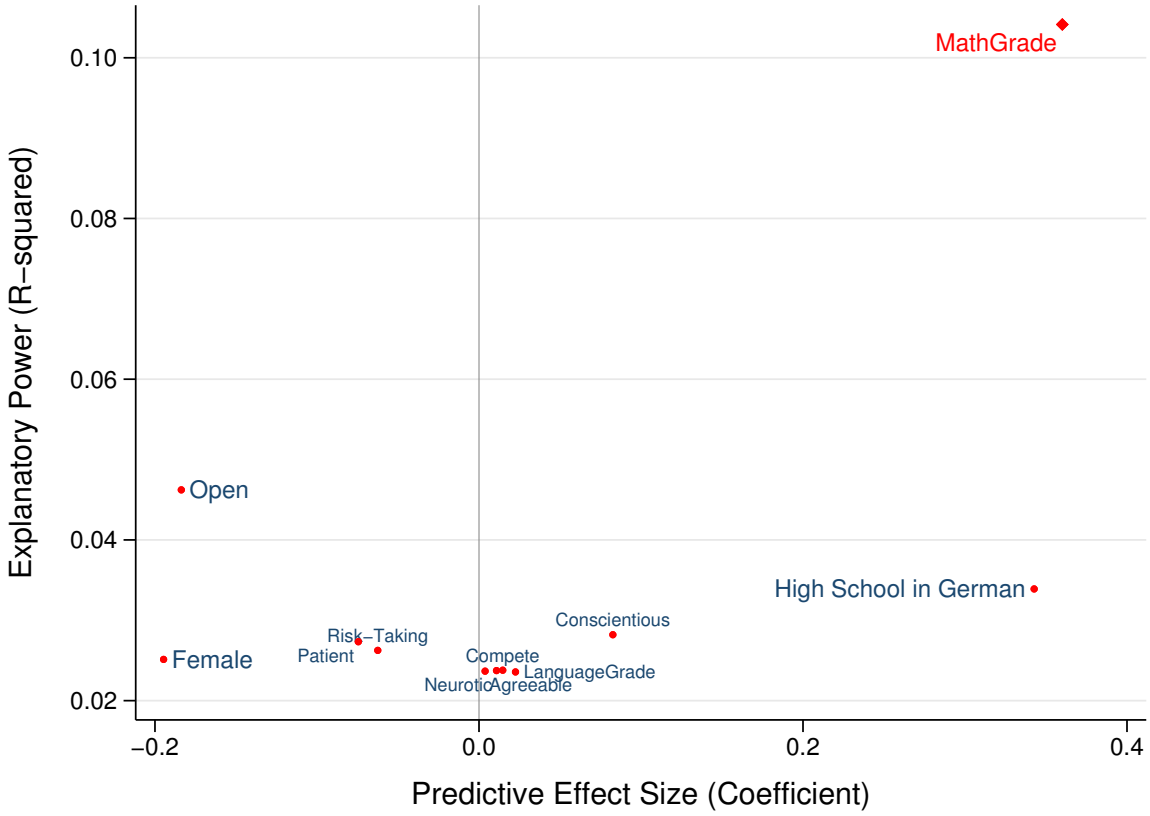
Note: In the 2018 endline survey, students reported up to five study mates that they studied most frequently with. The named students could be assigned peers from the study group or endogenous peers from out of the study group. The figure shows how the share of out-of-group study mates varies with group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

Figure B3: Gender Composition and Fraction of Female Out-of-Group Study Mates



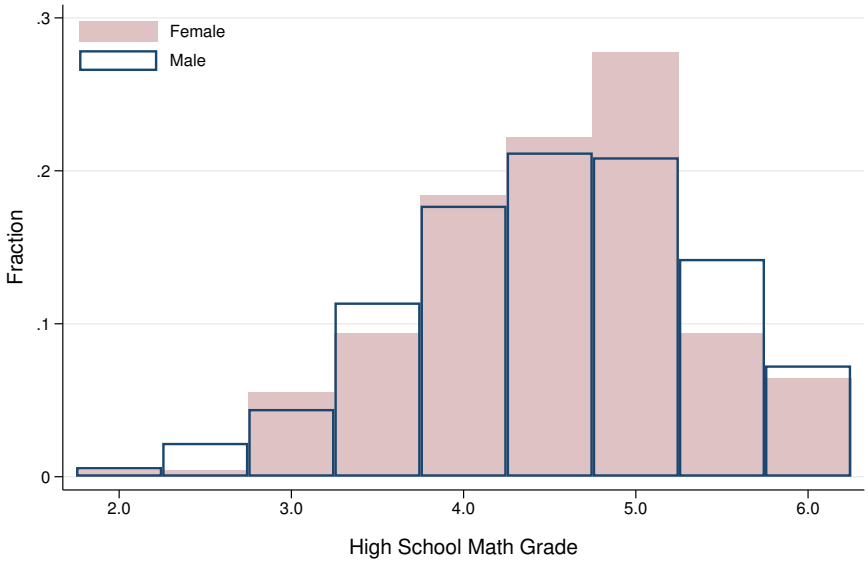
Note: In 2018, students reported up-to-five frequently-interacted study mates from within or out of the assigned group. In 2019, students report up-to-two study mates from out of the assigned group. I calculate the share of female peers among the reported out-of-group study mates, and plot the share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

Figure B4: How Different Characteristics Predict Course Grade



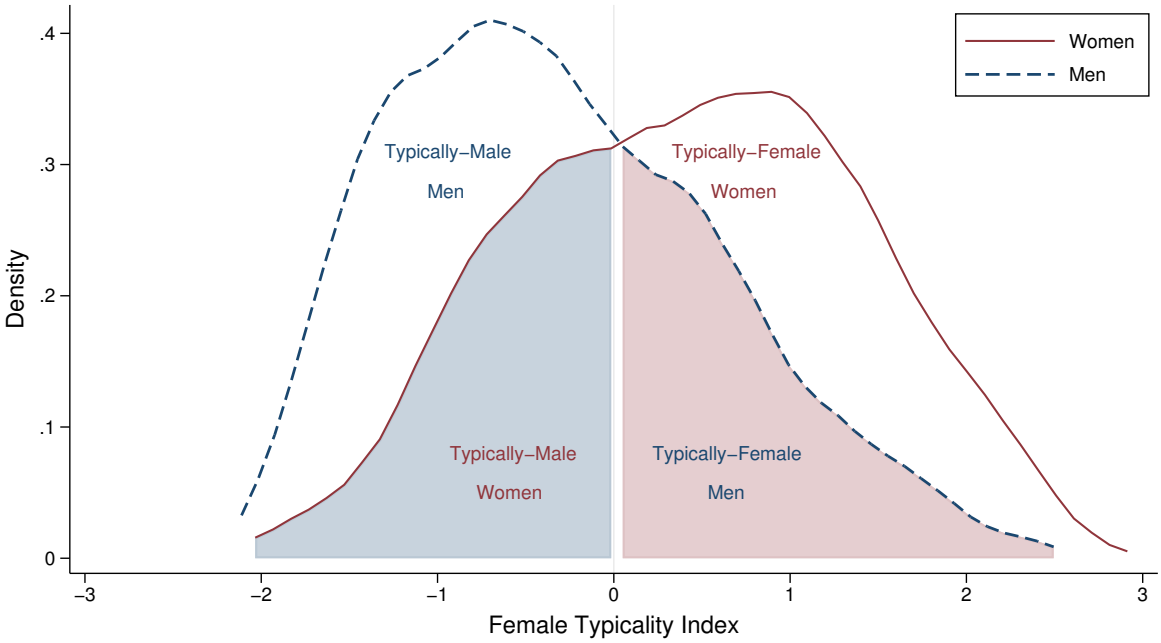
Note: The figure shows how the final grade is predicted by baseline characteristics, including gender, high school background, personality traits, gender bias, general competitiveness, risk preference, and patience. Gender bias is measured with gender role questions from the World Values Survey. General competitiveness is measured with the sing-item question from Buser et al. (2020). Risk preference and patience are measured with single-item questions asking how generally risk-taking and patient a student perceives himself/herself. I regress grade on each characteristic (standardized value) separately, controlling for major, course retaking, and year fixed effects, and plot the corresponding coefficient and R-squared.

Figure B5: The Distribution of High School Math Grade by Gender



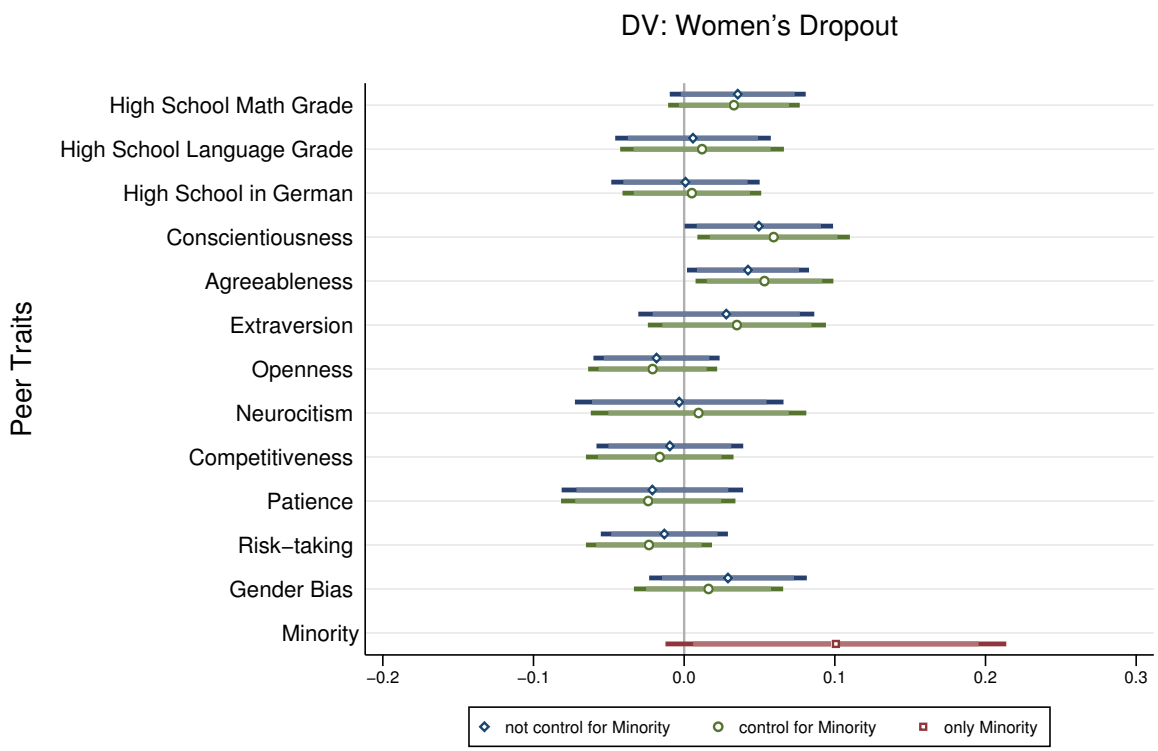
Note: The grade plots the distribution of self-reported high school math grade by gender. The median grade for both female and male students is *4.5*.

Figure B6: The Distribution of Female Typicality by Gender



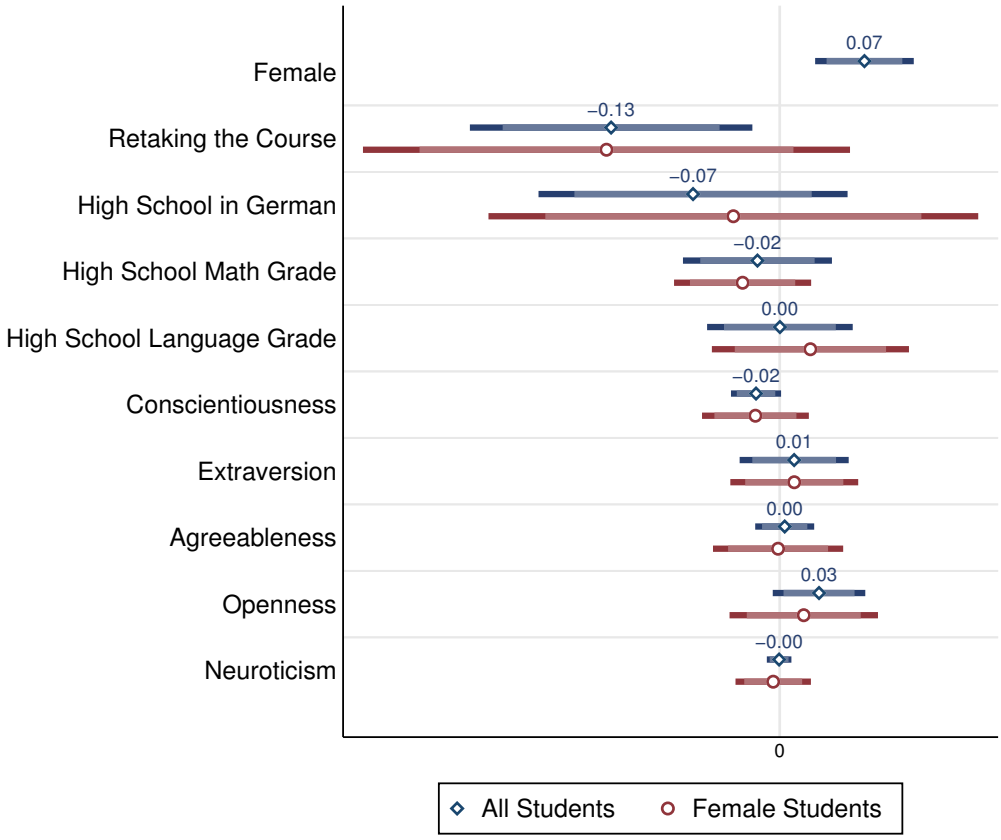
Note: The figure plots the distribution of the female typicality index by gender. Female typicality is the predicted value of female as a linear combination of the Big-Five personality traits, risk preference, patience, competitiveness, and gender bias. Students with an above-zero female typicality index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

Figure B7: The Impact of Peer Traits on Women's Dropout



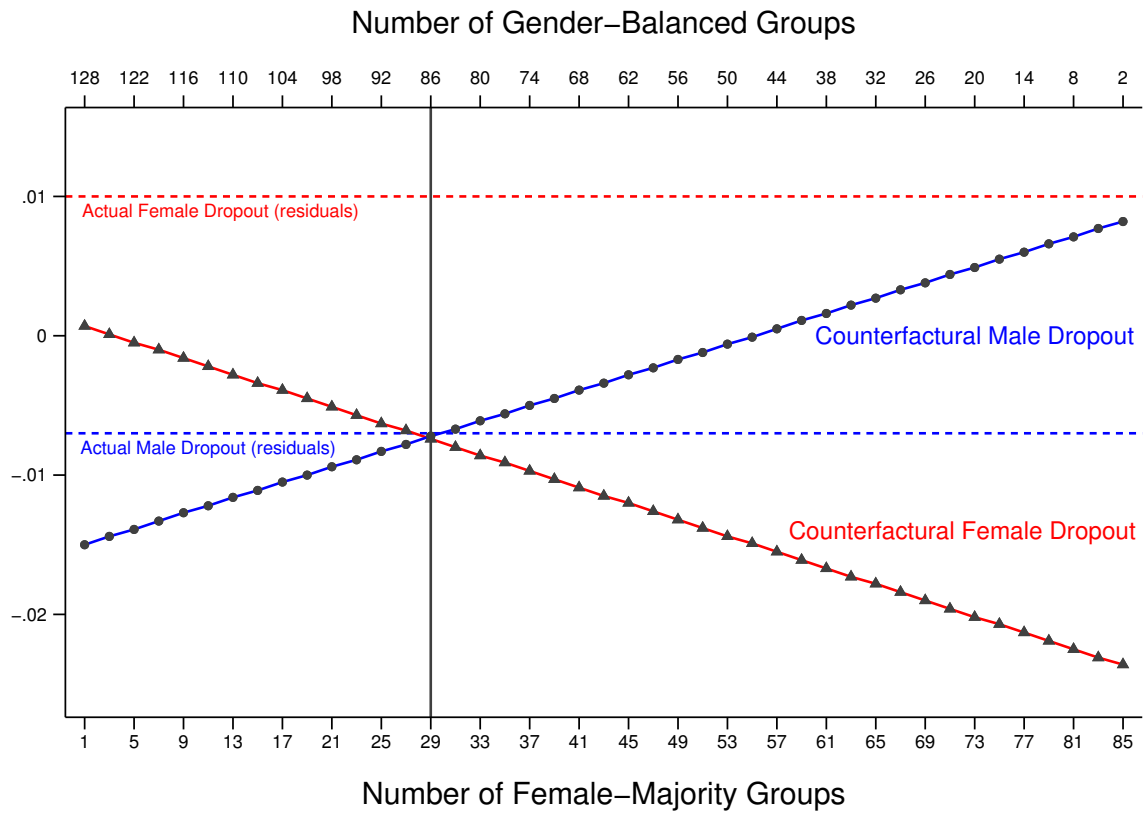
Note: The figure plots the estimated coefficients of different peer traits with or without controlling for minority status, and the coefficient of minority status, with 95% and 90% confidence intervals. Each point corresponds to one OLS regression, where the dependent variable is the dropout indicator for women, and controls include major, course retaking, 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. Cluster S.E. at the study group level.

Figure B8: What Kind of Students Register for Study Groups?



Note: 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 B9: Counterfactual Dropout Rate with Gender-Balanced and Female-Majority Groups



Note: 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, 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 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

Note: 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: Group Gender Composition and Women's Dropout

	(1)	(2)	(3)	(4)	(5)
	Marginal Effect	Compared to Female-Minority Groups			
DV: Women's Dropout					
Number of Women	-0.040				
{RI p-value}	{0.123}				
(clustered s.e.)	(0.024)*				
Gender-Balanced		-0.073			-0.085
{RI p-value}		{0.224}			{0.053}*
(clustered s.e.)		(0.062)			(0.061)
Male-Minority			-0.137		-0.124
{RI p-value}			{0.037}**		{0.006}**
(clustered s.e.)			(0.068)**		(0.062)**
Female-Only				-0.090	-0.099
{RI p-value}				{0.565}	{0.169}
(clustered s.e.)				(0.096)	(0.073)
Observations	259	147	137	81	259

Note: Column (1) examines the marginal effect of one female peer in a study group on women's dropout. Columns (2)–(4) compare female-minority groups separately to gender-balanced, male-minority, and female-only groups. Column (5) includes all types of gender compositions. Note that column (1) is the traditional model estimating gender peer effects: regressing the outcomes on the share/number of female students/peers. It is also the main regression model in my pre-analysis plan registered in the AEA RCT Registry. Given the strong non-linearity as implied by results in columns (2)–(5), I divert from the pre-analysis plan by focusing on the difference between women in female-minority groups and other women, even though the estimated linear effect is also significant. All columns use OLS regressions and control for major, course retaking, and year fixed effects. Results using probit regressions are very similar. Standard errors in parentheses are clustered at the group level. Randomization inference (RI) p-values are derived from 1,000 replications of study group assignment within randomization strata. $*p < .1, **p < .05, ***p < .01$.

Table B3: The Impact of Minority Status on the Final Grade

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Women				Men			
DV: Final Grade								
Panel A: All students								
Minority Status	-0.346	-0.297	-0.327	-0.788	-0.137	-0.118	-0.116	-0.719
{RI p-value}	{0.109}	{0.159}	{0.111}	{0.171}	{0.638}	{0.649}	{0.664}	{0.410}
(clustered s.e.)	(0.231)	(0.223)	(0.222)	(0.336)**	(0.259)	(0.277)	(0.288)	(0.526)
Observations	259	238	235	235	361	323	319	319
R-squared	0.025	0.158	0.192	0.173	0.022	0.109	0.128	0.191
Panel B: Conditional on taking the exam								
Minority Status	-0.014	0.038	0.022	0.013	-0.194	-0.117	-0.105	-0.115
{RI p-value}	{0.916}	{0.800}	{0.883}	{0.926}	{0.295}	{0.528}	{0.580}	{0.542}
(clustered s.e.)	(0.161)	(0.155)	(0.156)	(0.154)	(0.173)	(0.157)	(0.169)	(0.166)
Observations	226	211	209	209	327	297	293	293
R-squared	0.049	0.187	0.219	0.227	0.040	0.186	0.209	0.208
High school controls	N	Y	Y	Y	N	Y	Y	Y
Personality traits	N	N	Y	Y	N	N	Y	Y
Weighted	N	N	N	Y	N	N	N	Y

Note: The table examines the impact of minority status on students' final grades ranging from point 1 to 6. Panel A include all students, while Panel B only include students who have taken the final exam. The average grade is 3.7 among women and 4.0 among men. The standard deviation of the final grade is 1.3 for women and 1.2 for men. All columns use OLS regressions, and include the basic controls: major, course retaking, 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. Columns (4) and (8) use inverse probability weighted regression where the weight the predicted probability of taking the final exam based on the basic controls, high school controls, and personality traits. Standard errors in parentheses are clustered at the group level. Randomization inference (RI) p-values are derived from 1,000 replications of study group assignment within randomization strata. $*p < .1, **p < .05, ***p < .01$.

Table B4: Group Gender Composition and Women's Educational Expectation

	(1)	(2)	(3)	(4)	(5)
	Marginal Effect	Compared to Female-Minority Groups			
DV: Women's Educational Expectation					
Number of Women	0.151				
{RI p-value}	{0.069}*				
(clustered s.e.)	(0.102)				
Gender-Balanced		0.421			0.422
{RI p-value}		{0.031}**			{0.006}***
(clustered s.e.)		(0.192)**			(0.184)**
Male-Minority			0.251		0.322
{RI p-value}			{0.278}		{0.030}**
(clustered s.e.)			(0.192)		(0.178)*
Female-Only				0.990	0.599
{RI p-value}				{0.122}	{0.010}***
(clustered s.e.)				(0.361)***	(0.425)
Observations	177	103	92	52	177
R-squared	0.267	0.356	0.248	0.483	0.253
Mean of DV	-0.133	-0.125	-0.296	-0.374	-0.133
SD of DV	0.961	0.946	0.887	1.179	0.961

Note: Use OLS regressions. Column (1) examines the marginal effect of one female peer in a study group on women's expectation. Columns (2) to (4) compares women in female-minority groups separately to women in gender-balanced, male-minority, and female-only groups. Column (5) includes the indicators of all gender compositions. All regressions control for major, course retaking, year fixed effects, the baseline level of the overall educational expectation, high school grades, whether the first language is German, and the Big-Five personality traits. Clustered standard errors at the group level are in parentheses. Randomization inference (RI) p-values are derived from 1,000 replications of random assignment of study groups within randomization strata. $*p < .1, **p < .05, ***p < .01$.

Table B5: Minority Status and the Number of Peers with Frequent Interaction

	(1)	(2)	(3)	(4)
	N(Peers) Studying Together > Once/Month		N(Peers) with Joint Social Events ≥ Once/Month	
	Women	Men	Women	Men
Minority Status	-0.580	0.243	-0.276	-0.184
{RI p-value}	{0.016}**	{0.408}	{0.100}	{0.303}
(Clustered S.E.)	(0.236)**	(0.351)	(0.139)*	(0.079)**
Observations	146	202	146	202
R-squared	0.061	0.037	0.127	0.076
Mean of DV	0.603	0.545	0.199	0.292
SD of DV	1.034	0.925	0.628	0.683

Note: The table presents the impact of minority status on the self-reported number of peers with intensive academic or social interaction. The dependent variable in columns (1)–(2) is the number of peers studying together more than once per month. In columns (3)–(4), the dependent variable is the number of peers with social events together at least once per month. All columns use OLS regressions and control for course retaking and major fixed effects. Results further controlling for high school background and personality traits are very similar. Randomization inference (RI) p-values are derived from 1,000 replications of random assignment of study groups within randomization strata. Clustered standard errors in parentheses. $*p < .1, **p < .05, ***p < .01$.

Table B6: 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)	(0.090)	(0.126)*	(0.127)***	(0.132)***
Pairwise observations	179	177	147	187	174	150
R-squared	0.758	0.923	0.918	0.519	0.618	0.607
Group FEs	Y	N	N	Y	N	N
Individual FEs	N	Y	Y	N	Y	Y

Note: 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 B7: Effects of Minority Status Using Lee Bounds

	(1)	(2)	(3)
DV: Endline Educational Expectation			
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: 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

Note: 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.

Table B8: Group Registration and Dropout/Expectation

	(1)	(2)	(3)	(4)
Panel A: Women's Dropout (OLS)				
Having A Study Group	-0.056	-0.033	-0.040*	-0.042*
	(0.028)	(0.019)	(0.018)	(0.019)
Observations	943	749	742	738
R-squared	0.049	0.084	0.101	0.105
Panel B: Men's Dropout (Probit)				
Having A Study Group	-0.016	0.021	0.017	0.014
	(0.026)	(0.023)	(0.021)	(0.021)
Observations	1,356	943	933	923
Pseudo R-squared	0.0176	0.0654	0.0891	0.111
Panel C: Men's Dropout (OLS)				
Having A Study Group	-0.015	0.018	0.013	0.011
	(0.023)	(0.020)	(0.019)	(0.021)
Observations	1,356	943	933	923
R-squared	0.013	0.036	0.049	0.063
Panel D: Men's Educational Expectation at Endline (OLS)				
Having A Study Group	0.054	0.045	0.050	0.054
	(0.077)	(0.071)	(0.076)	(0.088)
Observations	763	641	637	632
R-squared	0.021	0.079	0.108	0.263
High school controls	N	Y	Y	Y
Personality traits	N	N	Y	Y
Additional individual controls	N	N	N	Y

Note: The table examines whether students with study groups differ from those without study groups in terms of dropout behavior and educational expectations. All columns control for major, course retaking, and year fixed effects. High school controls include high school math and language grades, and whether the first language is German. Personality traits are the Big-Five personality traits. Additional individual controls include risk, patience, competitiveness, gender bias, and baseline expectation. Standard errors are in parentheses and clustered at the major level. $*p < .1, **p < .05, ***p < .01$.

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 approach (Romano and Wolf, 2005a,b), which corrects the p-values based on a stepdown resampling method. Table C1 focuses on the results for women and presents the clustered p-values, the resampling p-values, and the corrected p-values. I include all the primary and secondary outcome variables analyzed in this paper. 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 Romano-Wolf correction, all outcome variables expect for the expected probability of passing the first year are significantly affected by minority status.[24]

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

Outcome Variable	Clustered p-val	Resample p-val	Romano-Wolf p-val
Dropout	0.0805	0.028	0.082
Final Grade	0.1371	0.050	0.097
Overall Educational Expectation	0.0188	0.007	0.028
Expected Grade	0.0069	0.003	0.015
Expected Probability Pass	0.5377	0.404	0.404
Expected Degree	0.1044	0.034	0.097
Overall Peer Interaction	0.0191	0.003	0.028
Academic Interaction	0.0369	0.008	0.049
Social Interaction	0.0393	0.006	0.049
N(Peers) Study > Once/Month	0.0166	0.002	0.028
N(Peers) Social ≥ Once/Month	0.0514	0.015	0.059

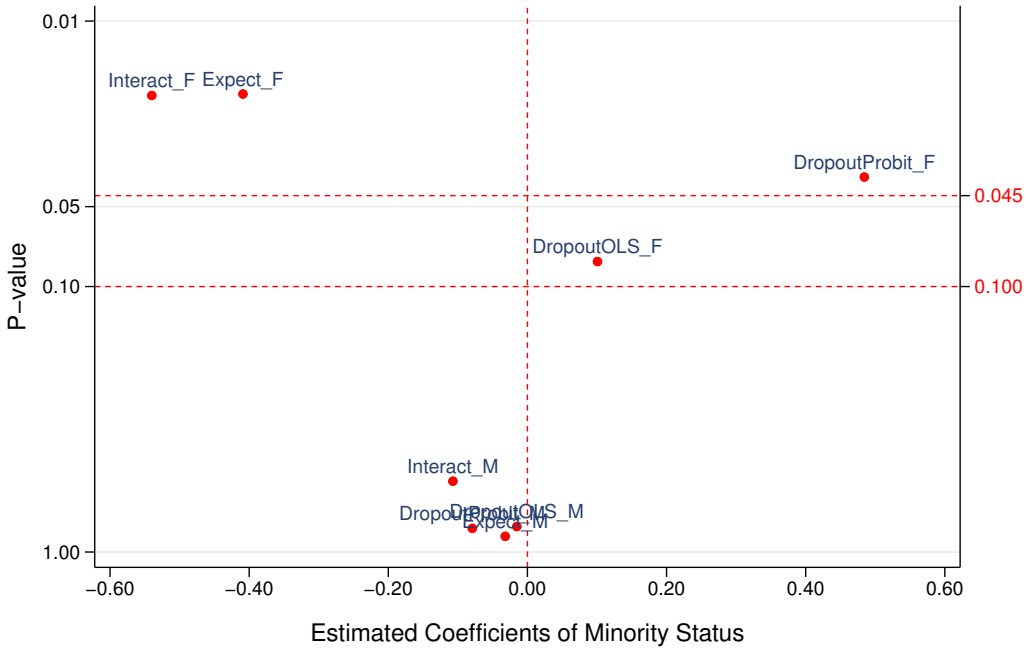
Note: The table presents the original (clustered) p-values and the corrected p-values using the Romano-Wolf (RW) multiple hypothesis correction. The RW approach controls the family-wise error rate (FWER), and corrects the p-values with a resampling-based stepdown method.

As an alternative test, I control the False Discovery Rate (FDR) for all the main results – dropout (using OLS and probit regressions), the overall educational expectation and peer-to-peer interaction – for men and women. Figure C1 shows the coefficients of minority status, the clustered p-values, and the corrected significance threshold. Without multiple-test correction, all the main results for women are significant at the 10% level, and all results for men are not

[24] The results for men are not presented here, but they all remain statistically insignificant after the Romano-Wolf correction.

statistically significant. I derive the corrected critical p-value for statistical significance (0.045 in this case), based on the step-up FDR procedure of Benjamini et al. (2006). 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 (p-values against coefficients), and the corrected critical p-value indicating the significance level—the *smile plot* based on the `multproc` command in `STATA` (Newson and Team, 2003). Each estimation dot is labeled with the outcome variable, the gender (F or M), and the regression model for dropout (OLS vs. Probit). 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 and 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

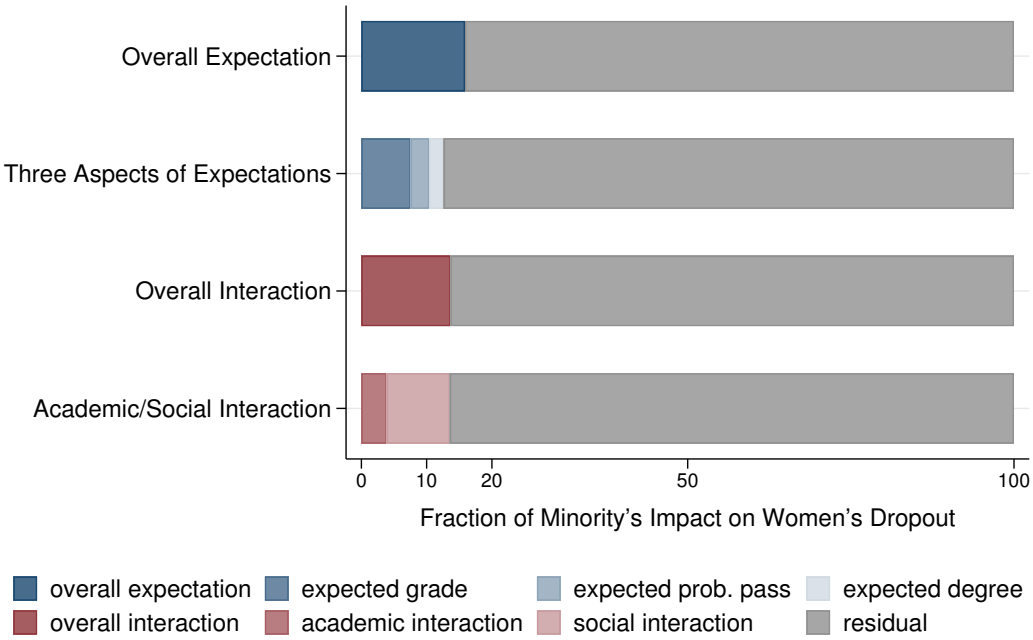
Note: All columns use OLS regressions, include both male and female observations, and control for gender, major, course-retaking, as well as baseline expectation and year fixed effects if applicable. Regressions for peer-to-peer interaction use only observations in 2019. Columns (1)–(2) examine the correlation 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 and Shaw, 2007; Rath and 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 and Rose (2018) and Bostwick and Weinberg (2018) show that women are more likely to persist in male-dominated settings if they have more female peers. However, Zölitz and Feld (2020) and Brenøe and 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 and Van Ewijk (2014) and Zölitz and 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 and Van Ewijk (2014), Huntington-Klein and Rose (2018), and Zölitz and 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

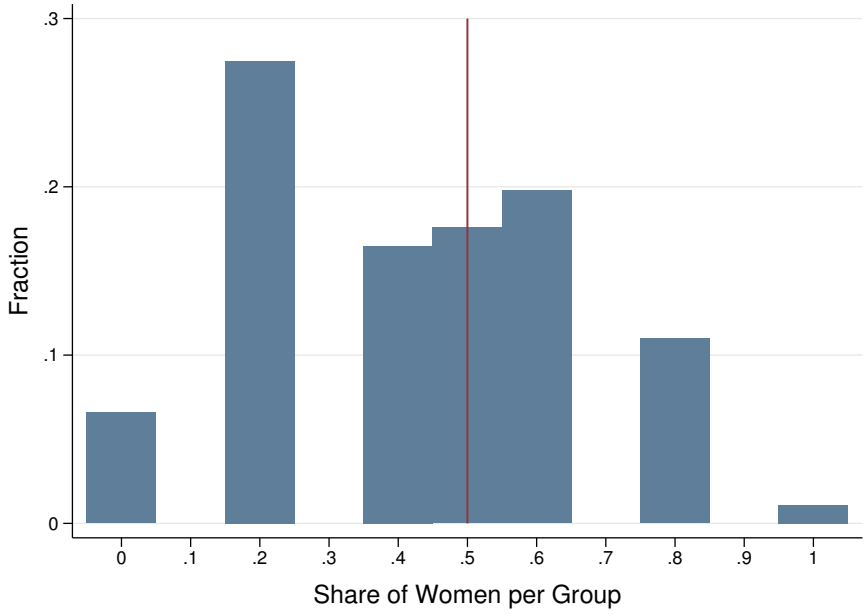
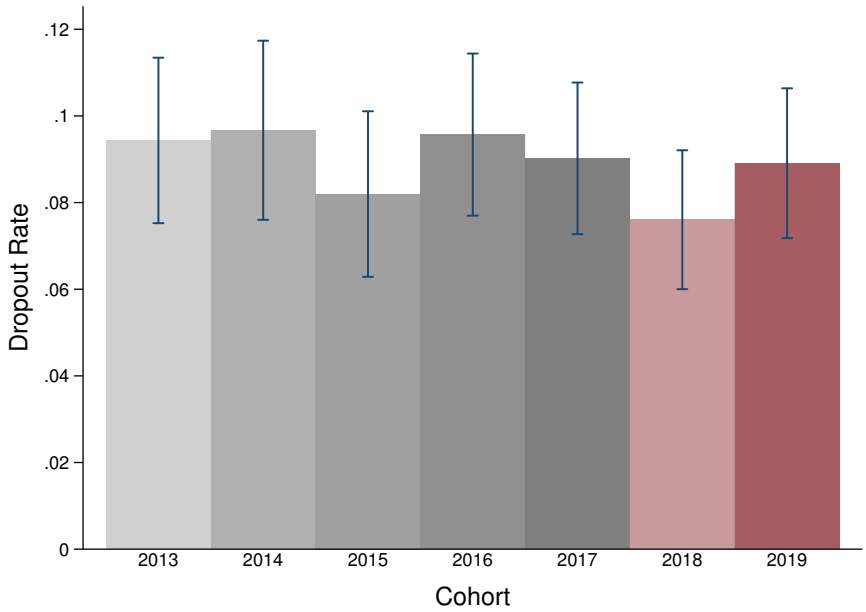


Figure C4: The Overall Dropout Rate by Cohort



Note: 95% confidence intervals. My experiment took place among the cohorts of 2018 and 2019.