UNIVERSITÀ DELLA CALABRIA | DIPARTIMENTO DI ECONOMIA, STATISTICA E FINANZA "GIOVANNI ANANIA"

Working Paper Series

WP n° 3, marzo 2021

FEMALE LEADERSHIP: EFFECTIVENESS AND PERCEPTION

Maria De Paola

University of Calabria, Department of Economics, Statistics and Finance "Giovanni Anania" – DESF
Institute for the Study of Labor (IZA), Bonn
(e-mail: maria.depaola@unical.it)

Francesca Gioia

University of Milan, Department of Law
(e-mail: francesca.gioia@unimi.it)

Vincenzo Scoppa

University of Calabria, Department of Economics, Statistics and Finance "Giovanni Anania" – DESF
Institute for the Study of Labor (IZA), Bonn
(e-mail: vincenzo.scoppa@unical.it)

Info
Via Pietro Bucci, Cubo **0/C**
87036 Rende (Cs) - Italia
tel. (+39) 0984 492434 / 492422

http://www.unical.it/desf

Female leadership: effectiveness and perception*

Maria De Paola
(Department of Economics, Statistics and Finance, University of Calabria and IZA, Bonn)

Francesca Gioia
(Department of Law, University of Milan)

Vincenzo Scoppa
(Department of Economics, Statistics and Finance, University of Calabria and IZA, Bonn)

Abstract. We ran a field experiment to investigate whether individual performance in teams depended on the gender of the leader. About 430 students from an Italian University took an intermediate exam that was partly evaluated on the basis of teamwork. Students were randomly matched in teams of three and, in each team, we randomly chose a leader entrusted the task of coordinating the work of the team. We find a positive and significant effect of female leadership on team performance. This effect is driven by the higher performance of team members in female-led teams rather than by an improvement in leader performance, suggesting that female leaders altruistically devote their energies to improving teamwork. In spite of the higher performance of female-led teams, male members tended to evaluate female leaders as less effective, whereas female members have provided more favorable judgments.

JEL classification: J16, M12, M54; C93.

Keywords: Team; Leadership; Gender; Stereotypes; Randomized Experiment.

Acknowledgements: We are grateful to the Dean of the Department of Business and Economics at the University of Calabria for allowing us to run the experiment. We would like to thank Guido de Blasio, Gabriella Conti, Marco Francesconi, Victor Lavy, Roberto Nisticò, Marco Savioli, and participants to the 2017 AIEL Conference for useful comments and suggestions.

*M. De Paola, Department of Economics, Statistics and Finance, University of Calabria, and Institute for the Study of Labor (IZA), Bonn – Address: Via Ponte Bucci, 87036 Arcavacata di Rende (CS), Italy. E-mail: m.depaola@unical.it; F. Gioia, Department of Law, University of Milan, Via Festa del Perdono 7, 20122 Milan. E-mail: francesca.gioia@unimi.it; V. Scoppa: Department of Economics, Statistics and Finance, University of Calabria, and Institute for the Study of Labor (IZA), Bonn – Address: Via Ponte Bucci, 87036 Arcavacata di Rende (CS), Italy. E-mail: v.scoppa@unical.it.

1. Introduction

Women have made progress in many social and economic spheres but they are still heavily underrepresented in leadership roles. In 2020 women held around 23-25% of the seats in the U.S. Congress and in the national Parliaments of many European countries (USA 27.5%, France 39.5%, UK 33.9%, Italy 35.7%, Germany 31.2%)[1]. Moreover, in the US, women represent only 6.6% of CEOs and hold only 25.5% of board seats. Likewise, only 26.7% of board members of the largest publicly listed companies in the EU are women (European Commission, 2019). This under-representation of women in top positions translates into a larger gender gap at the top quantiles of earnings distribution.

The low representation of women in leadership positions can be caused by several factors including differences in productivity, differences in preferences and psychological attitudes and by gender discrimination (see Eckel et al. 2020 for a review). Women might behave or share preferences and stereotypes that might affect their ability to lead and, consequently, their probability of being selected as leaders. On the other hand, women may be reluctant to accept leadership roles. Finally, despite their effectiveness as leaders and their willingness to lead, they might be discriminated against and prejudicially overlooked for these positions. Casting light on the role played by each of these causal factors is crucial as it can help in eventually formulating appropriate policies and leading to efficiency gains (Balafoutas and Sutter, 2012). Nonetheless, the economics related literature which provides evidence on these issues is scant, with a few papers focusing on self-selection (Alan et al., 2020; Born et al. 2018; Chakraborty and Serra, 2019) and a few others considering gender differences in performance (Grossman et al. 2016; Reuben and Timko, 2018; Timko, 2017a, 2017b).

With the aim of closing this gap, in the present paper we investigate whether men and women differ in their effectiveness as leaders. In our analysis, we mainly focus on the activity of organizing, motivating and coordinating the work of a team. In fact, a leader is a person typically appointed in an organization to enhance collective work through a number of activities aimed at motivating, organizing and coordinating teamwork, building mutual trust and cooperation, etc. (Yukl, 2013; Gardner, 1993). By means of a field experiment, we study the causal effect of the leader's gender on team performance in a real life environment represented by the preparation for a university exam. Since in our setting, the grade obtained in the exam was based on team performance, the leader could positively affect the grade awarded by improving coordination among team members and inducing prosocial behaviors. Women and men might perform differently in such types of tasks as they differ in a number of psychological attitudes.[2] For instance, as women tend to be more prosocial than men (see Croson and Gneezy, 2009), female leaders could be more able to induce cooperative behaviors in team members or sacrifice themselves in order to improve team performance. They might also be more capable of organizing team work or of assigning to each team member the accomplishment of his/her duties.

[1] "Women in National Parliaments", see: http://www.ipu.org/wmn-e/classif.htmandhttps://data.ipu.org/women-ranking.

[2] A large number of papers show gender differences in a range of psychological attitudes (see among the others Croson and Gneezy, 2009; Bertrand, 2011; Azmat and Petrongolo, 2014). However, women holding leadership positions might differ from the general population. For instance, Adams and Funk (2012), using a large survey of directors, show that female and male directors differ systematically in their core values and risk attitudes. Female directors are more benevolent and less power oriented than male directors. Nonetheless, in contrast to many works showing that women tend to be more risk averse, they also find that female directors are more risk loving compared to their male counterparts.

In addition to comparing male and female performance in leadership positions, we also investigate how team members react to the leader's gender and their perceptions of their leadership effectiveness. Following the sociological and psychological literature showing that the same leadership behaviors, when engaged by a female, are evaluated less favorably than they are when engaged by a male (Eagly, 1987; Eagly and Karau, 2002), we analyze whether male and female leaders are evaluated differently by their team members and the dimensions along which their team's activity differs.

Our experiment involved students enrolled on four different economics courses in an Italian University. They were offered the possibility to sit the exam according to an alternative examination scheme introducing teamwork as part of the assessment. A total of 538 students joined the experiment. They were randomly assigned to teams composed of three members and, within each team, a randomly selected member was appointed to the role of leader. The leader had the task of contacting team members (by e-mail or phone calls) in order to schedule team meetings and to organize team studying activities. These activities consisted of solving one set of exercises assigned during teaching classes (the leader was responsible for delivering the homework to the course's instructor) and preparing together the part of the exam evaluated on the basis of team performance. Even if not explicitly required, in order to carry out their task, the leaders had to motivate and coordinate team members, building mutual trust and cooperation and enhancing knowledge sharing. As a reward for the additional work involved, the leaders received 2 points to add to the final exam grade if the team's performance was above a given threshold.

After taking the test, students were asked to fill out a final questionnaire containing questions on the activity of the team and individual evaluation of team effectiveness and leadership activity.

The results of our experiment show that, controlling for leaders' and members' abilities, teams with a randomly selected female leader significantly outperform teams led by a male leader: the performance of students in female-led teams is about 0.7 points higher than that of students in male-led teams. This effect corresponds to about 0.2 SD of the dependent variable and is robust to the inclusion of several control variables. Digging deeper into this effect, we find that female members tend to react more to a female leader and that women appointed as leaders altruistically invest more energy into improving the performance of the whole team (even if this slightly weakens their own performance). This interpretation is consistent with the answers given by leaders to the post-experiment survey showing that females found the role of leader more demanding in terms of effort compared to their male counterparts. In addition, the effort demanded by the role tends to increase with the number of men in the team.

Using data from the assigned homework and the post-experiment survey, we analyze also if team activities and the evaluation of the leader's effectiveness differs based on the leader's gender. We find that female-led teams are significantly more likely to submit their homework despite spending on average the same amount of time working together. These results suggest that female leaders are more effective than male leaders in organizing team members' work. However, in spite of a better performance from female-led teams, consistently with results found by other works (for instance, Boring, 2017; Mengel et al., 2019), female leaders do not receive more enhanced evaluations from male members of their team. On the other hand, female members

tend to be more favorably disposed towards their female leaders, in that they are prone to recognize their effort, although not their effectiveness. Female leaders themselves do not seem to be aware of their better results as their level of satisfaction is similar to that of male leaders with the work done by their teams. On the other hand, we find that female leaders are stricter in the judgment they give of the effort provided by team members.

The paper is organized as follows. In Section 2 we briefly review the related literature. Section 3 describes the experiment, presents the data and reports some balance checks. In Section 4 we carry out our main empirical analysis. In Section 5 we study team members' evaluations and team activity. Section 6 concludes the paper.

2. Literature

Our research contributes to the existing literature in various ways. The economics related literature investigating leadership mainly relies on laboratory experiments that use 'minimum effort coordination games', that is games in which the lowest performing player determines the overall group performance.[3] These studies show that, in the absence of communication, coordination failure can be very common[4] and leaders can work as a "coordination device" to improve the organization using one-way communication to convince other team members that everyone will exert a high effort level (Sahin et al., 2015; Kriss and Eil, 2012; Brandts and Cooper, 2007; Weber et al., 2001). We add to these works analyzing leadership effectiveness in a real-life environment in which individuals have strong incentives to perform well; leadership is not a one-day role but a long-term task and leaders can adopt their own leadership styles instead of choosing whether to implement or not pre-determined tasks. Also, instead of considering leadership in a minimum effort coordination game, we focus on the role of leaders in organizing the work of the team when the total outcome is equally shared between team members.[5]

By focusing on gender differences in leadership efficacy, we complement the mixed evidence found by Grossman et al (2016), Reuben and Timko (2018) and Timko (2017a, 2017b). Grossman et al. (2016) run a laboratory experiment with randomly selected leaders who have to provide guidance on how to play the game to maximize group earnings and show that male leaders have a greater impact on followers' decisions compared to female leaders. Reuben and Timko (2017) extend the work of Grossman et al. (2016) by considering gender differences between elected and randomly-selected leaders. They find evidence of gender difference in the effectiveness of leaders only for elected leaders, while no difference emerges for randomly-selected leaders. Eagly (2007) and Eagly and Carli (2003) review the evidence and the causes of women's disadvantage in accessing and occupying leadership positions despite their high effectiveness.[6]

[3]The literature based on observational data mainly investigates the performance of incumbent leaders in the corporate or political arena (see for instance Ahern and Dittmar, 2012; Beaman et al., 2009; Bertrand et al., 2019; Chattopadhyay and Duflo, 2004).

[4]Costless, non-binding pre-play communication between players can improve coordination and efficiency (Blume and Ortmann, 2007; Devetag and Ortmann, 2007; Van Huyck et al., 1990).

[5]In our setting, the role of leader is exogenously defined. A number of papers investigate how leader's effectiveness can be enhanced through mechanisms that strengthen legitimacy such as democratic election (Brandts et al., 2015) and through different communication modes (Kriss and Eil, 2012).

[6]Alan et al. (2020) show that shying away from public scrutiny is a key determinant of women's tendency to avoid leadership positions. Born et al. (2018) find that women are less likely to self-select into a leadership position in male dominated

No gender difference in performance with randomly selected leaders is also highlighted by Timko (2017a, 2017b), while weak differences are reported by Dufwenberg and Gneezy (2005) who investigate the effect of team composition on team performance.[7]

While these papers report either no gender differences or a higher effectiveness of male leaders, we find a positive and significant effect of female leadership on team performance suggesting that gender differences in leadership effectiveness might depend on the specific task performed by the leader.[8] The activity of a leader is multidimensional and as argued by situational theorists of leadership (see for instance, Ayman, 2004; Chemers, 1997), the appropriateness of particular types of leader behaviors depends on the context. Our findings show that female leaders can be particularly effective in those circumstances where cooperation is important.

We also contribute to the literature investigating individual attitudes toward male and female leaders. Reuben and Timko (2017) show that unsuccessful female leaders are re-elected at considerably lower rates than unsuccessful male leaders.[9] Similar results are found by Grossman et al (2016) showing that individuals are less likely to attribute success to female leaders and also are less likely to reward them generously (see also, Eagly and Carli, 2003).[10] Consistently, we also find that women still face some difficulties in having their merits recognized, especially in teams with a majority of men. This might depend on the lack of fit between feminine qualities and leadership roles. Individuals have expectations of how leaders should behave that are not completely compatible with socially shared expectations of what is the appropriate behavior for women; this can lead to a negative evaluation of female leaders.

The concern for these distortions in perceptions of leader effectiveness may prevent women from assuming leadership positions and entering lucrative careers that they would otherwise be qualified for. In addition, women might have internalized these gender stereotyped perceptions and perceive themselves as less suitable leaders. This finds support in our results showing that female leaders themselves are not aware of their

---

environments, while Chakraborty and Serra (2019) show that women are less likely to self-select into a leadership role in a setting where leaders face the possibility of receiving angry messages from employees.

[7] A number of papers investigate how the gender composition of teams affects their economic performance and find mixed results. For instance, Apesteguia, Azmat and Iriberri (2012), considering a large business game, find that teams formed exclusively by women perform worse than all other gender combinations, while Hoogendoorn et al. (2013) show that teams with an equal gender mix perform better than male-dominated teams in terms of sales and profits. In addition, Delfgaauw et al. (2013), in a field experiment involving stores that compete in a tournament, show that stores where the store's manager and a sufficiently large fraction of the employees have the same gender, increase sales in response to the tournament.

[8] On gender differences in leadership in different domains see Grosch, Mueller, Rau, Zhurakovska (2020).

[9] Chakraborty and Serra (2017) show in a lab experiment that women tend to self-select less in leadership positions. Men and women seem to display distinct leadership styles and women seem to be more affected by distributional concerns and/or feedback.

[10] Evidence showing a gender bias against women is found also in the literature investigating students' evaluations of teaching. Boring (2017), using data from a French university, finds that male students are biased in favor of male professors. In addition, despite the fact that students appear to learn as much from women as from men, men are perceived by both male and female students as having stronger class leadership skills. Very similar results are found also by Mengel et al. (2019), who study teaching evaluations in a well-known university in the Netherlands where students are randomly assigned to sections. Funk et al. (2019) show that female students evaluate female professors more favorably (compared to male students) and are more likely to choose a female professor when the pool of professors is male dominated. A gender bias is highlighted also by MacNell et al. (2015) who run an experiment within an online course. Thanks to the peculiar features of their experiment, they are able to manipulate the information students receive about the gender of their instructor and to keep constant teaching quality and style (by deceiving students about the instructor's true gender identity). In spite of this, they find that students evaluate the male identity significantly more favorably than the female identity.

enhanced performance.

3. Experimental Design and Data

3.1. Design and Procedure

We run a field experiment involving students enrolled in the academic year 2015-2016 at the courses of Microeconomics, Macroeconomics, Econometrics and Personnel Economics offered by the First and Second Level Degree Course in Business and Administration at the University of Calabria.[11] Courses are worth 10 credits each, corresponding to 60 hours of teaching and to a nominal 250 hours of study, and are held during the second semester (from February to June).

At the beginning of the courses, students were offered the opportunity to join an alternative examination scheme instead of sitting the standard final exam at the end of the course.

The alternative exam scheme was composed of two tests, each covering half of the program, to be taken immediately after the first half of the course (intermediate test) and at the end of it (final test), respectively.

The intermediate test (on the first part of the course's program) consisted of two parts: one was evaluated on the basis of individual performance ("the individual part") and the other was evaluated on the basis of team performance ("the team part"). The two parts had similar questions and exercises but covered different parts of the program taught during the first half of the course: the first 2/3 of the program was assigned to the individual part which counted for 2/3 of the total mark (students could score a maximum of 20 points); the last 1/3 of the program taught during the first half of the course was assigned to the team part.[12] This was communicated to students in order to allow them to work together to prepare for the team part of the assessment.

In the team part, students had to answer individually a set of questions; each of them could gain a maximum of 10 points (1/3 of the total mark); the sum of the points earned by the students in a team were equally divided among team members so that the score assigned to each student was given by the average score obtained by the members of the team. Teams were composed of three members, one of whom was randomly chosen and appointed as the leader. The leader had the tasks of coordinating the team, calling team members' meetings, organizing team studying activities, guaranteeing the solving of a set of exercises assigned during the classes and submitting them to the course's professor.[13] As a reward for these additional activities, the leader received a fixed

[11] The University of Calabria is a middle-sized public university located in the South of Italy. It has currently about 30,000 students enrolled in different Degree Courses and at different levels of the Italian University system. Since the 2001 reform, the Italian University system is organized around three main levels: First Level Degrees (3 years of legal duration), Second Level Degrees (2 years more) and Ph.D. Degrees. In order to gain a First Level Degree, students have to acquire a total of 180 credits. Students who have acquired a First Level Degree can undertake a Second Level Degree (acquiring 120 more credits). After having accomplished their Second Level Degree, students can enroll in a Ph.D. degree.

[12] Even if the team part counted only for 1/3 of the total mark, this does not mean that stakes from team work are low because students could gain a maximum of 10 points in the team part and the threshold to pass the exam is 18 points. Of course, with higher stakes the effects that we find could be even larger.

[13] Solving the exercises was not mandatory for passing the exam. We analyse the probability of solving and submitting the homework in Section 4.

reward of 2 points to add to the final exam grade if the total grade of his/her team was at least equal to 18 (i.e. an average of 6).[14]

The final test, to be taken at the end of the course to complete the exam in the alternative scheme, was evaluated exclusively on the basis of individual performance with scores ranging from 0 to 30 (as in the standard exam). The final exam grade was given by the average of the grades obtained at the intermediate and final test. Students joining the alternative scheme, as required by the university administration for ethical reasons, were free to leave it at any point and to sit the standard exam instead.[15]

As in previous academic years, the standard exam was about the whole course program (without the possibility to split the course program into two tests) and students were evaluated exclusively on the basis of individual performance.

We explained to students that the aim of the experiment was to study the performance under teamwork but without mentioning the issue of gender and leadership to avoid influencing their behavior. Students were given one week to choose whether to join the alternative exam scheme or to sit the standard exam.

In Figure A1 in Appendix A we report a diagram describing the organization of the exam both under the standard scheme and in the experimental setting.

Students enrolled in the experiment by filling out an online survey asking questions on family background, risk preferences, social attitudes and expectations regarding performance. Once we obtained the list of participating students, within each course we randomly assigned them to teams of three members and, within each team, we randomly selected one member to act as the leader. We then defined two treatments on the basis of the gender of the leader: *Woman Led Team* for teams with a female leader and *Man Led Team* for teams with a male student as the leader.

Students were promptly informed of the team composition, the name of the leader and the parts of the course program assigned to teamwork and to individual work. Within each course, all students attended the lectures in the same room, at the same time and with the same instructor and teaching material. Immediately after the first half of the teaching classes, students undertook the intermediate test. Within each course, all students took the test with the same questions and at the same time.

After the intermediate test, students were asked to fill in an online survey available on the course webpages. Answering the final survey was strongly encouraged but not mandatory. Students could answer until the exam grades were published. The aim of such a final questionnaire was to collect information on team activity and individual evaluations of team effectiveness and leadership activity. All questions were identical for leaders and members, except one which was phrased slightly differently to elicit respectively teammates' evaluation of the effort and effectiveness of the leader and leaders' evaluation of the extent to which they found their role to be demanding.

[14]In the subsample of teams where all three members took the intermediate exam, 48 (17 males and 31 females) received the bonus and 51 (27 males and 24 females) did not receive it because the team performance was lower than the threshold.
[15]This implies that students could start by taking the intermediate test, observe their grade, and then shift to the standard exam.

3.2. Descriptive Statistics and Balance Checks

The design of the experiment produced three subsamples of students: those who enrolled on the courses (743), those who joined the experiment (538)[16] and those who actually showed up at the intermediate test (433).[17] On average 2.6 team members showed up at the intermediate test;[18] in 68.6% of teams all members took the intermediate test, in 26.8% of teams two members showed up and in only 4.6% of teams only one member took the test.[19]

In Table 1 we provide descriptive statistics separately for the three subsamples of students. About 50% of students are women[20]. Students in all subsamples are on average 22 years old. Students enrolled on the courses (col. 1) and joining the experiment (col. 2) have an average *High School Grade* of 82.5 (*High School Grade* ranges between 60 and 100), while *High School Grade* is slightly higher among students taking the test (83.2)(col. 3), suggesting a selection of more able students in the test. About 51% of students enrolled in the courses have studied in a *Lyceum* (col. 1). This percentage becomes higher in the other two subsamples (about 53% and 54% for students joining the experiment and sitting the exam, respectively).

Insert Table 1 here

As expected, one third of students were assigned the role of leader and 52% of the teams were led by a woman. 14% of teams who took the intermediate test were composed of three women, about 33% by two women and one man, about 40% by only one woman and two men and in about 13% of teams all the members were men.

From the online survey, we observed that students expected to obtain a grade of at least 25 in both subsamples (col. 1 and 2). Students' answers to the on-line survey filled-in when joining the experiment allowed us to build self-reported measures of risk attitudes and trust. The question we used to elicit risk attitudes is formulated as follows: "A lottery A allows you to obtain 100 euros with probability 50% or 0 euro with probability 50% (that is, when tossing a coin, head means winning 100 euros while tail means winning zero). Please, choose your favourite option between the lottery and a certain amount of 10 euros". For students choosing the certain amount there were no further questions. For the other students, we asked subsequent questions increasing the certain amount in steps of 10 euros up to 90 euros if the chosen option was the lottery. The questions ended when the chosen option was the certain amount.We built the variable *Risk Aversion* taking values

[16] In all the courses except one the number of students joining the experiment was a multiple of three. In one course 208 students joined the experiment and as a result of the randomization we ended up with two teams made by two members.

[17]80% of the sample who opted for the alternative exam actually showed up at the intermediate test. This drop-out rate is in line with the drop-out rate of other experiments involving students enrolled at the same university (De Paola, Gioia, Scoppa, 2019; De Paola, Gioia, Scoppa, 2018; De Paola and Gioia, 2016).

[18] We informed students that if one or two team members did not show up at the intermediate exam, in order to compute the score obtained in the team part of the test, we would replace the score of the absent student/s with one (two) randomly selected student/s sitting the test.

[19]In the final sample, about 40% of students are enrolled in the course of Microeconomics, 23% in Macroeconomics, 25% in Personnel Economics and 12% in Econometrics.

[20] The high percentage of women among economics students is in line with data from Italy where 43% of students enrolled in the field of economics are women (see https://anagrafe.miur.it/index.php)

from 0 (for students preferring the lottery to a certain amount of 90 euros) to 9 (for students preferring 10 euros with certainty to the lottery). *Risk Aversion* was, on average, 4.8 in both subsamples (col. 2 and 3).

The question we used to elicit trust is formulated as follows: "Consider the following situation: two subjects A and B (who do not know each other and who cannot communicate) have got 120 euros each. Subject A can transfer to subject B one of the following amounts {0, 20, 40, 60, 80, 100, 120}.The amount decided by subject A is tripled by a benefactor before being transferred to subject B (for example, if A decides to transfer 80 euros, B will receive 240 euros). B can then choose to transfer the amount that s/he prefers (also zero) to subject A. The final amount earned by each subject will be the initial endowment minus the amount transferred plus the amount received. If you were subject A, how much would you transfer to subject B?". The variable *Trust* simply reflects the values chosen by students and has an average of about 56 in both subsamples.

We measure student's performance in the individual part (*Grade Individual Part*) and the individual performance obtained in the team part (*Grade Team Part*).[21] Students taking the test obtained an average grade of 6 on the team part and of 11.8 on the individual part.

Leaving students free to choose whether to join the alternative scheme or not does not invalidate our design as long as the decision to sit the test for those enrolled is not influenced by the gender of the leader. However, selection across other dimensions may be relevant for the external validity of our results. Thus, with the aim of checking whether our design was attracting students with particular characteristics, we have conducted the Wilcoxon rank-sum test between students choosing to sit the standard exam and those joining the alternative scheme. We find that the two samples do not differ in terms of gender (p-value=0.788) or ability (*High School Grade*, p-value=0.788), but students deciding to join are significantly younger (p-value=0.000).[22] Similar results on age are found when we compare students who decided to join the experiment but then did not sit the exam with those who took the intermediate test. We also find that the latter have a significantly higher *High School Grade* (p-value=0.002); this is possibly due to less able students feeling less confident in their ability to succeed and thus avoiding to sit the test.

To investigate whether the gender of the leader matters for team performance we need comparable individuals in teams led by men and by women. In Table 2, we carry out the corresponding balance checks by studying the impact of individual characteristics on the probability of being part of a woman led team, both in the sample of students joining the experiment and in the sample of students showing up at the intermediate test, conditional on being a woman. As the samples also include leaders, being a woman increases the probability of having a woman as a leader in both samples. The other characteristics are equally balanced in the two treatments

[21] We emphasize that *Grade Team Part* is the performance of each team member and does not represent the average score within the team.

[22] The sample of students enrolled at the courses includes also students that were supposed to sit the exam the year(s) before but have failed (or never taken) it. These students usually do not attend again the lectures and this may be a reason why they prefer the standard exam.

except for *Trust* that is positively associated with the probability of being assigned to a woman led team.[23]

Insert Table 2 here

Results consistent with those shown in column (2) of Table 2 are found when we analyze the impact of having a female leader *vs* a male leader on the probability of *Taking the Intermediate Test*. As shown in Table 3, being assigned to a woman led team does not affect the probability of actually attending the intermediate test. In columns (1) and (2) we report results on the entire sample and vary the set of control variables. The coefficient returned by the dummy variable *Woman Led Team* is far from statistically significant in both specifications. This holds true both for women (col. 3) and men (col. 4). In addition, we find that, both for men and women, the probability of taking the test is positively affected by academic abilities (*High School Grade* and *Lyceum*) and negatively by *Age*.

These results are useful also to understand whether there are gender differences in the strategic behavior of leaders. In fact, to avoid free-riding, the leader can either increase effort of members or encourage lower-performing students to drop-out. Even if we are not able to understand whether students who decided to drop out have made this decision autonomously or have been induced by the team leader, we do not find any difference in the drop out behavior of students assigned to female and male leaders.

Insert Table 3 here

As our analysis focuses on students' performance at the intermediate test, we want to ensure that individuals showing up at the test and being part of women and men led teams are comparable in terms of observable characteristics. In Table 4 we report descriptive statistics of individual pre-determined characteristics for men led teams and for women led teams in the subsample of students present at the test. In the first three columns we focus on leaders while in the last columns we consider only team members. When considering leaders, we find that female leaders have on average a higher *High School Grade* than male leaders, which is not surprising as, in our sample, women have, on average, a higher *High School Grade* than men, as shown in columns (4), (5) and (6) of Table 4. To take this into account, in our analysis we control for leader's ability. Also, in Appendix B, as a robustness check, we run all our estimates on the subsample of leaders with a *High School Grade* in the interquartile range of male leaders' ability (that is higher than 70 and lower than 86).[24] The gender difference in terms of *High School Grade* remains significant also within team members; however, as it holds true both for men and woman led teams, the average ability of team members is well balanced across the two types of team.

[23]Although *Trust* seems to affect the probability of being in a *Woman Led Team*, when we run an F-test (with or without Bonferroni correction) on all the characteristics (but being woman) we find that none of the characteristics are statistically significant.

[24] We have also checked the robustness of our results on a subsample of leaders whose *High School Grade* lies on the same support (that is higher than 64 and lower than 100). Our sample reduces from 538 to 511 students joining the experiment and from 433 to 409 students taking the intermediate test and results hold and are slightly bigger in magnitude.

As regards team members (including both men and women), as shown in columns (7), (8) and (9) of Table 4, we find that all the characteristics are well balanced across the two treatment groups. Therefore, as explained above, gender difference in ability across team members is not an issue because the average ability of team members is not statistically different across treatments. We find similar results if we consider the students joining the experiment (see Table C1 in Appendix C).

Insert Table 4 here

3.3. Working in Team

Preliminarily, we try to verify if students assigned to the same team have effectively worked together. Despite asking them to work in a team and to solve a number of exercises together, we are not able to directly check if this actually happened or, whether instead, students worked separately, disregarding their assigned teammates.

In order to gather some information on the amount of time team members have worked together, we have included the following question in the post-experiment survey: "How many hours on average have you worked with your team members each week?". Students answering to the survey report a weekly average number of 3.87 hours spent working with the team[25], which increases to 4 for students who show up at the intermediate test. Leaders also report an average number of 3.96 hours of teamwork per week. These answers suggest that team members have effectively worked together.[26]

An alternative way to understand whether team members have cooperated in studying activities is to analyze the correlation of students' performance within each team. Since teams are randomly built, in the absence of cooperation, we should observe no correlation among teammates' performance.

In the first three columns of Table 5 we regress the performance in the team part of student *i* on the average performance on the team part of his/her teammates (*Team Grade Teammates (avg)*).[27] We control for course dummies, *Woman*, *High School Grade*, *Age*, leaders' predetermined ability and other individual's and group's characteristics. We find a strong effect of the performance of teammates on a student's own performance: an increase of 1 point in the average performance of teammates increases student's performance by about 0.1-0.24 points. The effect corresponds to about 0.03-0.07 SD. This evidence strongly supports the inference that team members have worked together and are affected by common factors.

Instead, when in columns (4)-(6) we regress student *i*'s individual grade on the average grade of his/her teammates on the individual part, we do not find any effect.

[25]For teams in which more than one member responded, team members were quite in agreement. In fact, the standard deviation of team members' answers ranges from 0 to 16.2 and is 0 in the first quartile, 0.7 in the second quartile, 1.7 in the third quartile and 5.6 in the 99 percentile.

[26] About 12% of students report to have studied together with the members of their team also for the individual part of the exam.

[27] We consider only teammates who show-up at the exam therefore our sample reduces because teams where only one member shows up are dropped out of this analysis.

These findings suggest that students have worked together in the team part but, overall, there have not been spillovers also on the individual part of the exam (maybe because the two parts were on different contents of the program).

On the other hand, as expected, we do not find any correlation among students' predetermined measures of abilities, such as *High School Grades* (ρ=-0.068, *p*-value=0.16), *Expected Grade* (ρ=-0.021, *p*-value=0.66), and so on.

Insert Table 5 here

In order to try to understand whether the correlation of performance within team members stems from the leaders' activity, we have investigated whether the actions taken by leaders affect the strength of within-team correlations in grades. To gather some information on team members' evaluation of the effort exerted by the leaders we have used data from the question "Using a scale going from 0 (very bad) to 10 (excellent), how do you rate the effort provided in team coordination by your leader?" asked to team members in the online survey proposed to students after they took the intermediate exam (see section 5 for more details). We find that the correlation between *Grade Team Part* and *Team Grade Teammates (avg)* is 0.295 (p-value=0.000) for team members who rate the effort exerted by their leader at a level higher than the average and 0.185 (p-value=0.050) for team members whose rating of their leader's effort is lower than the average. This difference corresponds to an increase of about 38 per cent in the correlation between the two grades with highly committed leaders. As regards the correlation between *Grade Individual Part* and *Individual Grade Teammates (avg)*, we find that it is 0.194 (*p*-value=0.020) for team members who rate the effort exerted by their leader at a level higher than the average and -0.007 (*p*-value=0.940) for the other team members. We find similar results also when we consider leaders' evaluation of how demanding was their activity (*Required effort*, see section 5 for details) and when we consider the observable part of their activity, that is, if they have delivered the assigned homework to the course's instructor.

Taken together our results suggest that students within the same team have effectively worked together and that leaders matter for team effectiveness.

4. The Impact of Female leadership on Team Performance

In this section, we investigate our main research question, that is, whether the gender of the leader influences the performance obtained by the team. We begin with a visual inspection of the data and proceed with the estimation of an econometric model.

Figure 1 shows the average grade obtained at the team part and 90 percent confidence intervals, separately by treatment. While students belonging to the *Man Led Teams* obtain on average a grade of 5.6, those assigned to *Woman Led Teams* perform better and obtain an average grade 0.76 points higher.

Insert Figure 1 here

However, the better performance of *Woman Led Teams* may depend on leaders exerting more studying effort and then performing better or on team members improving their performance thanks to the positive influence of their leader or on both. Figure 2 plots the average grade obtained on the team part separately by treatment in the subsample of leaders (a), team members that are not leaders (b) and team members that are not leaders, separately by the gender of the team member (c). Rather than an improvement in leader performance, we see evidence of a higher performance of team members, especially women, in *Woman Led Teams*.

Insert Figure 2 here

We now turn to our econometric approach and use data at the student level for the sample of students effectively taking the exam (433) to estimate several specifications of the following OLS model:

[1]

$$Grade\ Team\ Part_i = \beta_0 + \beta_1 Woman\ Led\ Team_i + \beta_2 Woman_i + \beta_3 HighSchoolGradeLeader_i + \beta_4 PercWomen_i + \beta_5 W_i + \mu_j + \varepsilon_i$$

where the dependent variable, $Grade\ Team\ Part_i$ is the score that student i obtains in the team part of the test and $WomenLedTeam_i$ is a dummy variable for the treatment status that takes value 1 for students assigned to a team led by a woman and 0 for the reference category, that is, a team led by a man. We control for the student's gender ($Woman_i$), leader's ability (*High School Grade Leader*) and the gender composition of the group (*Perc. Women*); a vector W_i of variables measuring student's predetermined characteristics as well as expectations and preferences (*Age, High School Grade, Lyceum, Expected Grade, Risk Aversion, Trust*) and the number of team members who showed up at the intermediate test (*# Members Present*); μ_j are courses fixed effects (dummies for Macroeconomics, Personnel Economics and Econometrics; Microeconomics is the reference category); ε_i is an error term. In this model, β_1 represents the causal effect of being assigned to a team led by a woman in terms of student's performance at the team part.

We present OLS estimates of the impact of female leadership on student's team performance in Table 6. In all regressions, standard errors are corrected for heteroskedasticity and clustered at the team level to take into account the influence of common factors within teams.

The first specification controls only for the dummy *Woman Led Team*. We find a positive, strong and statistically significant effect of female leadership: a student in a team led by a woman obtains 0.759 points more than a student in a team led by a man (*t*-stat=2.05). The effect corresponds to about 0.22 SD of the dependent variable, which represents an increase in the outcome variable of about 12.5 percent.

In the second column we add the dummy *Woman* and *High School Grade* to verify if the effect of a female leader is due to the student's gender and ability: we find a non-significant impact of *Woman* on team performance

and a positive and highly statistically significant effect of our measure of ability;[28] more importantly, the effect of being in a team led by a woman remains positive (0.700) and statistically significant with a small reduction in its magnitude.

As in our sample female leaders typically have higher abilities (the average *High School Grade* is of 86.6 and 79.0 respectively for female and male leaders), one might argue that our results are driven by traditional peer effects due to the fact that female leaders are of better "quality", that is, the effect we find might not be related to having a woman in charge but to having a better quality individual as leader. To deal with this issue in column (3) of Table 6 we include among controls a measure of leader quality (*High School Grade Leader*).[29] In addition, positive peer effects might also originate from an imbalanced presence of female members in groups led by men and women. Even if, as shown in Section 2, this channel should not be relevant in our setting since, conditional on the gender of the leader, the remaining gender composition of the group is balanced across women and man led teams, in column (3) we also control for the gender composition of the group. The impact of female leadership remains almost unchanged, about 0.714, statistically significant at the 10 percent level. The gender composition of the team and the leader's ability do not seem to have any effect on team performance; results do not change if instead of the percentage of women we control for dummies for having one, two or three women in the team.[30]

In column (4) of Table 6 we add some additional individual characteristics (*Age*, *Lyceum*, *Expected Grade*, *Risk Aversion*, *Trust*) and the number of team members effectively present at the exam (weaker students could decide to skip the exam and this could have an impact on remaining students).[31] Finally, in column (5) we also add course dummies. In both specifications, we show that teams led by a woman tend to perform significantly better than teams led by a man.

It is worthwhile to notice that since we could not force students to work with the assigned leader and team members and they probably worked also individually or with their traditional friends regardless of the team composition defined in the experiment, our estimates can be considered a sort of "Intention-to-Treat" effect (in which the estimated effect is diluted by partial compliance to the assigned experimental condition), while the average treatment effect is presumably somewhat larger.

Insert Table 6 here

[28] As *High School Grade* could predict performance differently for men and women we have also experimented by including among controls an interaction term between *High School Grade* and *Female* and we are not able to reject the null hypothesis of equal impact.

[29] Results remain stable if we also include the age of the leader and whether the leader has attended a *Lyceum* among controls.

[30] To be reassured that the effect we find is not driven by leader's ability we have also restricted our sample to leaders with a *High School Grade* in the interquartile range of male leaders' ability (that is higher than 70 and lower than 86). Our results continue to hold true and become larger in magnitude (results reported in Table B1 in Appendix B).

[31] The number of team members showing up at the test also represents an outcome variable that may be affected by leadership. However, as we have shown in Table 3, the probability of taking the exam does not depend on the leader's gender. This result is confirmed also when we consider as an outcome variable the number of team members showing up at the exam.

An interesting question is related to what drives the effect of the better performance of woman led teams. Is the whole team performing better? Are female leaders improving their performance? Or, does the effect derive from an increase in team members' performance? Also, are these effects different according to the student's gender?

We answer these questions in estimates reported in Table 7. First, in column (1) we try to understand how team members react to female leadership. At this aim, we exclude leaders from the sample and estimate specification (5) of Table 6, including the full set of controls. We find that female leaders increase the performance of their team members: team members' performance increases by 1.146 if the team is led by a woman.

In column (2) we try to understand whether female and male team members react differently to female leaders. For this purpose, we include among regressors an interaction term between *Woman* and *Woman Led Team*. We find that the increase in performance related to being part of a *Woman Led Team* is lower for men (+0.843; *p*-value=0.17) than for female team members (1.549=0.843+0.706, *p*-value=0.02), although the difference (0.706) is far from being statistically significant (*p*-value=0.34). Estimating separately for male and female team members, we find that in woman led teams the performance of men increases by 1.121 while the performance of women increases by 1.495 (estimates reported in Table D1 in Appendix D).

In column (3) we turn our attention to leaders and try to understand whether there is a gender difference in performance among leaders. Comparing the performance of a female leader with that of a male leader we find a small and non-significant difference against women (-0.557, *p*-value=0.39).

Insert Table 7 here

In column (4) we run our estimate on the whole sample, including among regressors the interaction term between the dummy *Woman* and the dummy *Leader* and the interaction term between *Woman* and *Woman Led Team.* We find results that are consistent with those found restricting the sample to team members and leaders, respectively. The performance of a male member (not leader) increases by 0.764 if a woman leads the team. The positive effect of a female leader is 0.726 points higher (that is, 1.49=0.764+0.726) if the member is a woman but the difference is not statistically significant. Female and male leaders do not significantly differ when we consider their own performance (-0.324=0.764-0.356+0.726-1.458, *p*-value=0.607). In addition, we find that female leaders tend to perform slightly worse than other women who are members of a woman led team (-0.854=0.604-1.458, *p*-value=0.064).

All in all, members of a team led by a woman tend to perform better (especially female members), while the leader's own performance is slightly less for women. As we will discuss below, the latter result might depend on the fact that female leaders devote a greater amount of effort in coordinating and organizing team activities and this might come at the cost of a lesser individual performance.

Results from Table 7 hold true also when we restrict the analysis to leaders with a *High School Grade* in the interquartile range of male leaders' ability (that is higher than 70 and lower than 86) in order to be reassured that our findings are not driven by gender differences in leaders' ability (see Table B2 in Appendix B).

We have also tried to understand whether the female leadership produces positive effects also on the individual part of the test. For this purpose, we have considered as dependent variable *Grade Individual Part*.[32] In order to compare the magnitude of the effects of the individual and team parts, we have divided by two the score obtained at the individual part of the test as this part was worth twice the team part. Estimates are reported in Table 8 (same specifications reported in Table 6). We find that woman led teams tend to perform better also in the individual part of the exam – suggesting that small spillovers occur from one part of the exam to the other – but the effect is very imprecisely estimated. Spillover effects from the team part to the individual part of the exam become, instead, statistically significant when we consider the subsample of leaders with a *High School Grade* in the interquartile range of male leaders' ability (see Table B3 in Appendix B).

Insert Table 8 here

4.1. Potential Drivers of the Impact of Female Leadership on Team Performance

We now inquire as to why the difference between teams led by men and by women arises. The information we have available to investigate this issue is limited; however, we can consider students' answers to the online survey aimed at evaluating team effectiveness and leader's activity. As explained above, all students joining the experiment, including those who did not show up at the intermediate test, were invited to fill out the survey. Approximately 85% of them completed the questionnaire.[33]

The questions asked in the survey allow us to understand whether the time spent together by team members has been affected by the leader's gender. As mentioned in Section 3, students report a weekly average number of team working hours (*Hours Together)* of 3.87, which increases to 4 for students who show up at the intermediate test. In Table 9, we investigate whether the leader's gender influences the time spent with team members. We report OLS estimates considering as dependent variable *Hours Together* and focusing on the whole sample of students joining the experiment (with the exclusion of leaders)[34], but similar results emerge when restricting the sample to students who have undertaken the intermediate test. As shown in column (1), the gender of the leader does not affect the time spent working together. In addition, we do not find differentiated effects according to the gender of team members (column 2). Leader's ability, on the other hand, produces a positive and statistically significant effect.

Even if the amount of time is the same, the effectiveness of the work may be different depending on the leader's gender. To investigate this issue, we look at the probability that the team leader has submitted the

[32] We have also considered the grade at the final test and the grade at the standard exam for those who decided to switch to it after the intermediate test. We find that being in a female led team does not significantly affect neither the decision to sit the final test instead of switching to the standard exam nor the grade obtained at the final test or at the standard exam.

[33] We find that, both in the sample of students joining the experiment and in the sample of those actually showing up at the test, being in a woman led team rather than in a man led team does not affect the probability of answering the survey. *High School Grade* is the only characteristic that affects this choice (positively) in both samples.

[34]When we consider the answers provided by leaders again we do not find any gender difference as regards the time they state to have worked with teammates.

homework assigned during teaching classes by the course's instructor and estimate a Linear Probability Model considering as dependent variable the dummy *Done Homework,* which takes the value of one for teams who have submitted their homework to the course's instructor and zero otherwise. Although students were strongly encouraged to work in the team and to solve the exercises assigned as homework, no specific incentive (or penalty) was announced for teams accomplishing (not accomplishing) the task. The percentage of students who have done their homework is 81.4% and 84.1% for students joining the experiment and students showing up at the intermediate test, respectively.

In our estimates, we keep one observation for each team and use team characteristics as control variables (average age of team members, average high school grade etc.).[35] We also control for our measure of leader's ability, *High School Grade Leader*. In column (3) of Table 9, we consider the whole sample of students joining the experiment, while in column (4) we focus on students taking the test. We find that teams led by a woman have a higher probability of doing their homework (+14 percentage points when we consider only students sitting the intermediate test).In addition, having a high-quality leader produces a positive and statistically significant effect.

We find similar results in the subsample of leaders with a *High School Grade* in the interquartile range of male leaders' ability (see Table B4 in Appendix B).

Insert Table 9 here

Using the questions asked in the final survey, we also investigate whether teammates have spent time together to prepare the individual part of the exam and for leisure activities. We asked students the following two questions: a) "Have you met the members of your team also for leisure?" (Students could choose Yes or No); b) "Have you studied alone or with the members of your team to prepare the individual part of the exam?" (Possible answers were: mainly alone; partly alone and partly with my teammates; mainly with my teammates). The large majority of students (83%) report that they did not spend leisure time with team members and that they have studied mainly alone to prepare for the individual part of the exam (88%).[36] We find that the leader's gender does not significantly affect the probability of spending leisure time with teammates and the probability of studying together with team members for the individual part of the exam. The effect is similar for male and female team members.

In Table 10 we analyze the link between the time spent together by team members and having submitted the homework on student performance at the exam. More precisely, we estimate equation [1] and instead of considering among regressors the dummy variable *Woman Led Team* we focus on *Hours Together* and *Homework.* In columns (1), (2) and (3) we focus on the time team members spent together. As shown in the first two columns (the first considering the whole sample and the second excluding leaders) having spent more time

[35] The effect of being in a team led by a female is statistically significant both in the whole sample of students joining the experiment and in the sample of students taking the test if, instead of keeping one observation for each team and using team characteristics as control variables, we keep all team members with individual level covariates and bootstrap (100 reps) to estimate the impact of leader's gender.

[36] We have grouped the answers to this question into two categories because students reporting to have studied mainly with teammates are only 2%.

working in team does not produce any statistically significant effect on team performance. Spending time together with teammates seems instead to be harmful for individual performance probably because the time left for studying the individual part of the exam reduces. On the other hand, we find that having submitted the homework is positively related with students' performance at the team part of the exam (columns 4 and 5). The impact on the individual part of the exam is still positive but not statistically significant.

All in all, our results suggest that, despite spending on average the same amount of time working with the team, students in teams led by women were more effective in finalizing teamwork. Then, the better results obtained by woman led teams might be due to the fact that female leaders are better at coordinating team activities.

Again, our findings do not seem to be driven by gender differences in leaders' ability as they are robust to the restriction of the sample to leaders with a *High School Grade* in the interquartile range of male leaders' ability (see Table B5 in Appendix B). The estimated effect of having submitted the homework is larger in this subsample while spending time together with teammates is no longer harmful for individual performance.

Insert Table 10 here

5. Are Male and Female Leaders Evaluated Differently?

In this Section, we use data from the online survey presented to students after they took the intermediate exam in order to analyze how they evaluate leader's activity and team effectiveness. With the aim of investigating how the leader was evaluated by team members, we asked the following two questions: a)"Using a scale going from 0 (very bad) to 10 (excellent), how do you rate the effectiveness of your leader?"; b) "Using a scale going from 0 (very bad) to 10 (excellent), how do you rate the effort provided in team coordination by your leader?".[37]

The average grades for *Leader Effectiveness* and *Leader Effort* were respectively 6.3 and 6.24 when considering the whole sample of students joining the experiment (with the exclusion of leaders). These rates slightly increase (to 6.39 and 6.32, respectively) when we focus only on students who took the intermediate test.

In Table 11 we report results considering the whole sample of students joining the experiment who answer the survey (with the exclusion of leaders). In all specifications we include the full set of controls. In columns (1) and (2) the outcome variable is *Leader Effectiveness*, while in columns (3) and (4) we consider *Leader Effort* as dependent variable.[38] We find that, on average, in spite of the higher performance levels of woman led teams, female leaders do not obtain better evaluations from team members (columns 1 and 3). This is consistent with results found by works analyzing teaching evaluations showing that women receive systematically lower teaching evaluations than their male colleagues (Boring, 2017; Mengel et al., 2019; Funk et al., 2019). Similar evidence is

[37]To investigate whether students are providing faithful answers to questions of our surveys we have regressed the *Expected Grade* on the effective *Grade* students obtained. We find a positive and high coefficient on *Grade* (0.74), with a *p*-value of 0.00, suggesting that students are taking our survey seriously and giving reliable answers.
[38]These two measures of leader's performance are strongly correlated (0.86, *p*-value=0.00).

found also by Chakrabort and Serra (2019) who, in an experimental framework, find that female managers receive more aggressive messages from employees.

We also find that female and male team members show different attitudes. In fact, when we include among our regressors the interaction term between *Woman Led Team* and *Woman* we find that while the evaluation of leader effectiveness made by male team members tends to be lower when the leader is a woman (about -1.284, statistically significant at 5 percent level), female members tend to evaluate similarly female and male leaders (+0.106=+1.178-1.284 for female leaders).[39] This holds true when we restrict the sample to students who undertook the intermediate test (results reported in Table E1 in Appendix E).[40]

As shown in column (4) similar results are found when *Leader Effort* is considered as an outcome variable. We find that women students tend to evaluate more favorably (1.457-0.969) the effort provided by female leaders compared to their male peers (-0.969). Consequently, taken together, these results suggest that male team members hardly recognize neither female leaders' effectiveness nor their effort, while female members tend to be more prone to recognize effort than effectiveness.

Insert Table 11 here

These results might be driven by men's stereotypes against female leadership. To get some additional evidence on this point we have considered the answers given by leaders to the post experiment survey and tried to understand whether female leaders working in the team composed mainly by men tend to assess differently their team. We firstly look at a question asking them to evaluate the working of their team: "Using a scale going from 0 (very bad) to 10 (excellent), how do you rate the functioning of your team?". We focus on the sample of leaders who took the intermediate test (we lose some observations because of missing answers).

As shown in Table 12, we do not find any difference based on gender in this response suggesting that female leaders are not aware of their effectiveness in coordinating and finalizing the work of their teams. In addition, team composition does not affect the answers given by leaders.[41] On the other hand, we find that female leaders are aware of the effort they have provided in accomplishing their role. In column (3), using as outcome variable an indicator of *Required Effort*, based on the answers given by leaders to a question asking them to rate on a scale going from 0 to 10 how demanding their activity was,[42] we find that on average female leaders rate their own effort higher compared to male leaders (statistically significant at 13 percent level). As shown in

[39] The interaction term *Female Led Team*Female* has *p*-value=0.106.

[40] We obtain the same result also when we consider students' answers to a question asking a general evaluation of team effectiveness ("How do you rate the effectiveness of your team? - Good, Neither Good Nor Bad; Bad)". 63.4% of students answered that they were satisfied with the work of their team, 21.4% expressed a negative evaluation and 15.2% were in the middle.

[41] Here we consider the subsample of leaders taking the test. Similar results are obtained when considering the sample of leaders joining the experiment.

[42] "Using a scale going from 0 (very bad) to 10 (excellent), how do you rate the effort you have provided in team coordination?".

column (4), this is especially true for women leading teams composed mainly of men (the interaction term *Woman*Perc. Women* is negative even if imprecisely estimated, *p*-value=0.12).[43]

In columns (5) and (6), we analyze how leaders evaluate the effort provided by team members in team activities (we have information for each of the two members[44] and consider the average value). We find that female leaders tend to evaluate as less intense the effort provided by team members compared to male leaders (-1.15 points) but the judgment does not significantly change with the gender composition of the team.

Results from Table 11 and Table 12 are similar but sometimes less precisely estimated in the subsample of leaders with a *High School Grade* in the interquartile range of male leaders' ability (see Table B6 and Table B7 in Appendix B).

These results, together with findings discussed above (female leaders are especially capable in enhancing the performance of female members; male members are less likely to evaluate positively female leaders), confirm that women might experience prejudiced evaluations especially in masculine organizational contexts (Eagly and Carli, 2003).

Insert Table 12 here

6. Concluding Remarks

Female leadership in economic and social contexts is a rather rare phenomenon. One possible explanation could be that women are less effective in leading roles. In addition, women's reputations as leaders might be affected by gender-based stereotypes which may jeopardize their success as leaders.

To shed light on gender differences in leadership effectiveness and on factors influencing leadership effectiveness, we have run a field experiment with a sample of students from an Italian University who took one part of their exam working in a team. At the beginning of the courses, we have randomly assigned students to teams of three members and, within each team, we have randomly selected a leader to organize team activities. Thus, we have built a setting in which teams are led by a woman or by a man and have different gender composition.

We evaluate if the gender of the leader affects team academic performance controlling for a number of individual characteristics and we find that woman led teams perform significantly better than man led teams (an effect of about 0.2 SD of the dependent variable). Considering that students typically could have studied also with other colleagues (who were not members of the same team), our estimates are to be interpreted as an "Intention to Treat Effect" and therefore the impact of the female leadership is likely diluted.

We also find that the positive effect of female leaders on team performance is mainly driven by the better performance of female team members. While there is no significant difference in the performance of male and

[43]Instead, when we consider the whole sample of students joining the experiment (including also those who have not sit the intermediate exam) we do not find any gender difference in the way leaders perceive the effort demanded by their role.
[44]"Using a scale going from 0 (very bad) to 10 (excellent), how do you rate the effort provided by team member *name*?".

female leaders, the latter tend to perform less effectively than other women who are members of a woman led team. This suggests that female leaders altruistically invest more energy into organizing teamwork, rather than to improving their own performance.

In order to have a more complete picture of the team activities, we use answers from a post-experiment survey together with the submission of team homework. We find that, despite the fact that teams led by men and women spend on average the same amount of time together, the effectiveness of the work is higher in woman led teams that are more likely to submit the homework to the course instructor.

When analyzing leaders' evaluation of their role, we find that female leaders are not aware of their effectiveness in coordinating and finalizing the work of their team, but are aware of the effort they have provided in accomplishing their role. In addition, they tend to evaluate as less intense the effort provided by team members compared to male leaders.

Even if we cannot say exactly which dimensions of woman's leadership style improve team performance, our results suggest that the female leaders' advantage, detected in our study, is probably due to better organizational skills, persuasiveness or authority in ensuring that everyone studies. Stereotypically feminine qualities such as cooperation, mentoring and collaboration might also have played a role in our experimental context. These qualities might turn out to be increasingly important in contemporary organizations which might benefit from prosocial behaviors, such as paying more attention to subordinates, helping others with their work or volunteering for tasks that go beyond their role description.

Nonetheless, women still suffer from prejudice and resistance when they occupy these roles; in fact, consistently with previous research (Eagly, 2007; Eagly and Carli, 2003), when investigating teammates' evaluation of their own leader, we find suggestive evidence that men tend to evaluate female leaders less favorably while women are more prone to recognize female leader's effort compared with male counterparts but less prone to recognize female leader's effectiveness. This suggests that much work still needs to be done to eradicate gender stereotypes in leadership, especially in settings where female leaders are non-traditional, in such a way as to increase both women's confidence in their capabilities as leaders and teammates' recognition of female leaders' work. Given the deep changes taking place in women's roles, more recognition of their authority and effectiveness as leaders will be an important development.

Our research contributes by informing that female and male leaders might not differ much in terms of effectiveness and, in some cases, women might well outperform men. This type of awareness is important first of all for women themselves, in order to enable them to be capable of reducing the negative effects that stereotypes can produce on their lives. It is also important in helping to change perceptions of the value of women in leadership roles and to motivate decision makers to contemplate the potential offered by more than half of the population. Despite the fact that providing leadership has historically been depicted as a masculine task, we have shown that women have a leadership advantage in contexts involving an important teamwork component, such as the one analyzed in our experiment, and consequently, they should be encouraged to take on leadership roles because they tend to work harder and their team might benefit more from their guidance.

It is worthwhile to notice that our results pertain to a specific domain of leadership activity. Investigating whether gender differences exist also along other dimensions represents a promising avenue for future research. Understanding what happens when considering, for example, the leader's ability to serve as a symbol to motivate or renew, would give a more complete picture of gender differences in leadership effectiveness and recognition.

References

Adams R. and Funk P. (2012). Beyond the Glass Ceiling: Does Gender Matter? *Management Science*, 58(2), pp. 219-235.

Ahern, K.R. and Dittmar, A.K. (2012). The changing of the boards: The impact on firm valuation of mandated female board representation. *Quarterly Journal of Economics*, 127, 137-197.

Alan, S., Ertac, S., Kubilay, E. and Loranth, G. (2020). Understanding gender differences in leadership, *Economic Journal*, 130(626), 263-289.

Apesteguia, Azmat G. and Iriberri N. (2012). The Impact of Gender Composition on Team Performance and Decision-Making: Evidence from the Field. *Management Science*, 58(1):78-9, 2012.

Avolio, B. J. (1999). *Full leadership development: Building the vital forces in organizations*. Thousand Oaks, CA: Sage.

Ayman, R. (2004). Situational and contingency approaches to leadership. In J. Antonakis, A. T. Cianciolo, and R. J. Sternberg (Eds.), *The nature of leadership* (pp. 148–170). Thousand Oaks, CA: Sage.

Azmat, G. and Petrongolo, B. (2014) Gender and the labor market: what have we learned from field and lab experiments?, *Labour Economics*, 30, pp. 32–40.

Balafoutas L. and Sutter M. (2012), Affirmative Action Policies Promote Women and Do Not Harm Efficiency in the Laboratory, *Science*, 335, pp. 579-582.

Beaman, L., Duflo, E., Pande, R. and Topalova, P. 2012. Female leadership raises aspirations and educational attainment for girls: A policy experiment in India. *Science* 335, 582-586.

Bertrand, M. (2011) New perspectives on gender, Handbook of Labor Economics, North Holland, Amsterdam.

Bertrand, M., Black, S.E., Jensen, S., and Lleras-Muney, A. (2019), Breaking the glass ceiling? The effect of board quotas on female labour market outcomes in Norway. *The Review of Economic Studies* 86, 191-239.

Blume, A., and Ortmann, A. (2007). The effects of costless pre-play communication: Experimental evidence from games with Pareto-ranked equilibria. *Journal of Economic theory*, 132(1), 274-290.

Boring A. (2017). Gender biases in student evaluations of teaching. *Journal of Public Economics.* 145, pp. 27-41.

Brandts, J., and Cooper, D. J. (2007). It's what you say, not what you pay: an experimental study of manager-employee relationships in overcoming coordination failure. *Journal of the European Economic Association*, 5(6), 1223-1268.

Brandts, J., Cooper D.J., Weber, R.A. (2015). Legitimacy, Communication, and Leadership in the Turnaround Game. *Management Science* 61(11): 2627-2645.

Chakraborty, P. and Serra, D. (2017), "Gender Differences in Leadership: An Experiment", *mimeo*.

Chakraborty, P. and Serra, D. (2019). Gender differences in top leadership roles: Does worker backlash matter?, *mimeo*.

Chattopadhyay, R. and Duflo, E. 2004. Women as policy makers: Evidence from a randomized policy experiment in India. *Econometrica*, 72, 1409-1443.

Chemers, M. M. (1997). *An integrative theory of leadership*. Mahwah, NJ: Erlbaum.

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

De Paola, M., and Gioia, F. (2016). Who performs better under time pressure? Results from a field experiment. *Journal of Economic Psychology*, *53*, 37-53.

De Paola, M., Gioia, F., & Scoppa, V. (2018). The adverse consequences of tournaments: Evidence from a field experiment. *Journal of Economic Behavior & Organization*, *151*, 1-18.

De Paola, M., Gioia, F., & Scoppa, V. (2019). Free-riding and knowledge spillovers in teams: The role of social ties. *European Economic Review*, *112*, 74-90.

Delfgaauw J., Dur A.J., Sol J. and Verbeke W. (2013). Tournament incentives in the field: Gender differences in the workplace, *Journal of Labor Economics*. 31(2), pp. 305- 326.

Devetag, G., and Ortmann, A. (2007). When and why? A critical survey on coordination failure in the laboratory. *Experimental economics*, 10(3), 331-344.

Dufwenberg, M., and Gneezy, U. (2005). Gender and coordination. In *Experimental business research* (pp. 253-262). Springer US.

Eagly, A. (2007), Female leadership advantage and disadvantage; resolving the contradictions, *Psychology of Women Quarterly*, 31, pp. 1–12.

Eagly, A. H., and Carli, L. L. (2003). The female leadership advantage: An evaluation of the evidence. *Leadership Quarterly*, 14, 807–834.

Eckel C., Gangadharan L, Grossman P. and Xue N. (2020), The Gender Leadership Gap: Insights from Experiments, Discussion number 14/20, Monash Business School.

European Commission (2019), Report on equality between women and men in the EU. ISSN 2443-5228

Funk P., Iriberri N. and Savio G. (2019).Does Scarcity of Women Instructors Create Demand for Diversity among Students? Evidence from Observational and Experimental Data, CERP discussion papers, n. 14190.

Gardner, J. (1993). *On leadership*. Simon and Schuster.

Grosch K., S. Müller, H. A. Rau, L. Zhurakhovska (2020), Selection into Leadership and Dishonest Behavior of Leaders: A Gender Experiment, Cesifo working papers, 8514

Grossman, P. J., Eckel, C., Komai, M., and Zhan, W. (2016). It Pays to Be a Man: Rewards for Leaders in a Coordination Game (No. 38-16). Monash University, Department of Economics.

Hoogendoorn S., Oosterbeek H. and van Praag M. (2012). The Impact of Gender Diversity on the Performance of Business Teams: Evidence from a Field Experiment. *Management Science*, 59(7), pp. 1479-1724.

Kriss, P. H., and Eil, D. (2012). Managers as coordination devices: Keep your door open, but your skin thick. *Essays on the Impact of Costly Communication on Coordination, Dissertation by Peter Kriss, Carnegie Mellon University, Pittsburgh*, 106-140.

MacNell, L., Driscoll A., and Hunt A. (2015). What's in a Name: Exposing Gender Bias in Student Ratings of Teaching. *Innovative Higher Education*, 40, pp. 291–303.

Mengel F., Sauermann J. and Z¨olitz U. (2019). Gender Bias in Teaching Evaluations. *Journal of the European Economic Association*, 17(2). Pp.535–566.

Miller, A. R., (2017). Women and Leadership. Oxford Handbook on Women and the Economy, eds. Susan L. Averett, Laura M. Argys and Saul D. Hoffman. New York: Oxford University Press. 2018, Forthcoming.

Reuben, E., and Timko, K. (2017). On the Effectiveness of Elected Male and Female Leaders and Team Coordination (No. 10497). IZA Discussion Papers.

Sahin, S. G., Eckel, C., and Komai, M. (2015). An experimental study of leadership institutions in collective action games. *Journal of the Economic Science Association*, 1(1), 100-113.

Timko, K. (2017a). Gender, communication styles and leader effectiveness. MPRA Paper 77021, University Library of Munich, Germany.

Timko, K. (2017b). Men and Women Are Equally Effective Leaders. MRPA Paper 77022, University Library of Munich, Germany.

Van Huyck, J. B., Battalio, R. C., and Beil, R. O. (1990). Tacit coordination games, strategic uncertainty, and coordination failure. *The American Economic Review*, 80(1), 234-248.

Weber, R., Camerer, C., Rottenstreich, Y., and Knez, M. (2001). The illusion of leadership: Misattribution of cause in coordination games. *Organization science*, 12(5), 582-598.

Yukl, G. A. (2013). *Leadership in organizations*. Pearson Education India.

Table 1. Descriptive Statistics. Mean and Standard Deviation

	Enrolled at the courses	Joining the Experiment	Taking the Intermediate test
	(1)	(2)	(3)
Woman	0.499	0.496	0.487
	(0.500)	(0.500)	(0.500)
Age	22.266	22.068	21.902
	(2.370)	(2.363)	(2.300)
High School Grade	82.495	82.530	83.236
	(10.716)	(10.736)	(10.805)
Lyceum	0.506	0.526	0.540
	(0.500)	(0.500)	(0.499)
Leader		0.335	0.346
		(0.472)	(0.476)
Woman Led Team		0.522	0.517
		(0.500)	(0.500)
High School Grade Leader		82.245	82.180
		(11.144)	(11.220)
Perc. Women		0.494	0.497
		(0.291)	(0.296)
Expected Grade		24.926	25.141
		(2.493)	(2.303)
Risk Aversion		4.866	4.838
		(2.187)	(2.152)
Trust		56.468	55.612
		(34.847)	(34.830)
Macroeconomics		0.234	0.231
		(0.424)	(0.422)
Microeconomics		0.387	0.395
		(0.487)	(0.489)
Personnel Economics		0.273	0.254
		(0.446)	(0.436)
Econometrics		0.106	0.120
		(0.308)	(0.325)
# members present			2.640
			(0.569)
Grade Team Part			6.036
			(3.487)
Grade Individual Part			11.797
			(6.212)
Observations	743	538	433

Notes: Standard Deviations are reported in parentheses.

Table 2. Balance Checks. The Probability of Being Assigned to a Woman Led Team as a Function of Individual Characteristics. Dependent Variable: Woman Led Team

	Joining the Experiment	Taking the Intermediate test
	(1)	(2)
Woman	0.335***	0.359***
	(0.043)	(0.047)
Age	-0.012	-0.012
	(0.011)	(0.013)
High School Grade	0.001	-0.001
	(0.002)	(0.002)
Lyceum	-0.004	-0.008
	(0.041)	(0.046)
Expected Grade	-0.004	0.000
	(0.008)	(0.009)
Trust	0.001*	0.001*
	(0.001)	(0.001)
Risk Aversion	0.014	0.016
	(0.009)	(0.010)
Macro	-0.048	-0.095
	(0.055)	(0.061)
Personnel Ec.	0.048	0.038
	(0.056)	(0.064)
Econometrics	0.119	0.115
	(0.092)	(0.100)
Observations	538	433
Adjusted R-squared	0.115	0.132

Notes: OLS estimates (Linear Probability Model). Dependent variable Being Assigned to a Woman Led Team. Standard errors (corrected for heteroskedasticity) are reported in parentheses. The symbols ***, **, * indicate that the coefficients are statistically significant at the 1, 5 and 10 percent level, respectively.

Table 3. The Probability of showing up at the intermediate test as a function of Woman Led Team and Individual Characteristics. Dependent Variable: Taking the Intermediate test

	Whole		Women	Men
	(1)	(2)	(3)	(4)
Woman Led Team	-0.016	-0.019	0.025	-0.044
	(0.035)	(0.036)	(0.062)	(0.057)
Macro	-0.029	0.026	0.025	0.069
	(0.045)	(0.045)	(0.068)	(0.058)
Personnel Ec.	-0.073	0.006	-0.028	0.068
	(0.045)	(0.050)	(0.073)	(0.065)
Econometrics	0.091**	0.309***	0.360***	0.234**
	(0.046)	(0.074)	(0.109)	(0.092)
Woman		-0.054		
		(0.038)		
Leader		0.051	0.019	0.023
		(0.034)	(0.056)	(0.052)
Age		-0.048***	-0.044***	-0.054***
		(0.011)	(0.017)	(0.013)
High School Grade		0.004**	0.004*	0.002
		(0.002)	(0.002)	(0.002)
Lyceum		0.032	-0.066	0.148***
		(0.034)	(0.048)	(0.048)
Expected Grade			0.028***	0.020*
			(0.011)	(0.012)
Trust			-0.000	-0.001
			(0.001)	(0.001)
Risk Aversion			-0.001	-0.001
			(0.011)	(0.010)
Constant	0.830***	1.476***	0.671	1.324***
	(0.033)	(0.285)	(0.535)	(0.430)
Observations	538	538	267	271
Adjusted R^2	0.007	0.070	0.091	0.113

Notes: OLS estimates (Linear Probability Model). Dependent variable showing up at the intermediate test. Standard errors (corrected for heteroskedasticity) are reported in parentheses. The symbols ***, **, * indicate that the coefficients are statistically significant at the 1, 5 and 10 percent level, respectively.

Table 4. Students' Characteristics across Treatment Groups separately for leaders and team members

	Leaders						Team Members		
	Woman Led Team	Man Led Team	Diff. (p-value)	Women	Men	Diff. (p-value)	Woman Led Team	Man Led Team	Diff. (p-value)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Woman							0.486	0.467	0.019
							(0.501)	(0.501)	(0.740)
Perc. Women Members							0.5	0.478	0.222
							(0.378)	(0.345)	(0.478)
Age	22.116	21.928	0.188	21.783	21.889	-0.106	21.774	21.909	-0.134
	(2.637)	(2.217)	(0.637)	(2.233)	(2.232)	(0.691)	(2.070)	(2.397)	(0.615)
High School Grade	86.671	79.054	7.617	86.341	80.730	5.611	83.000	83.852	-0.852
	(10.399)	(10.910)	(0.000)	(10.490)	(9.919)	(0.000)	(11.014)	(10.055)	(0.497)
Lyceum	0.461	0.635	-0.175	0.489	0.581	-0.092	0.554	0.519	0.036
	(0.502)	(0.485)	(0.032)	(0.502)	(0.495)	(0.121)	(0.499)	(0.502)	(0.551)
Expected Grade	25.408	24.851	0.557	25.267	25.034	0.233	25.108	25.185	-0.077
	(2.111)	(2.652)	(0.158)	(2.379)	(2.133)	(0.388)	(2.507)	(1.944)	(0.772)
Trust	51.053	57.838	-6.785	52.296	59.865	-7.569	59.595	52.593	7.002
	(35.084)	(36.350)	(0.247)	(31.479)	(36.552)	(0.062)	(34.084)	(34.446)	(0.087)
Risk Aversion	5.105	4.392	0.713	4.881	4.885	-0.004	4.959	4.800	0.159
	(2.030)	(2.195)	(0.041)	(2.282)	(2.055)	(0.989)	(2.023)	(2.311)	(0.539)
Obs.	76	74		135	148		148	135	

Notes: Standard deviations are reported in parentheses. *p*-values are reported in the columns with differences (3, 6 and 9).

Table 5. Correlations of Performance among Teammates

	Grade Team Part			Grade Individual Part		
	(1)	(2)	(3)	(4)	(5)	(6)
Team Grade Teammates (avg)	0.243***	0.237***	0.113*			
	(0.056)	(0.056)	(0.061)			
Individual Grade Teammates (avg)				0.078	0.075	-0.021
				(0.058)	(0.058)	(0.060)
Woman	-0.440	-0.495	-0.525	-0.405	-0.430	-0.385
	(0.377)	(0.381)	(0.376)	(0.696)	(0.703)	(0.710)
Age	-0.103*	-0.113*	-0.028	-0.026	-0.048	-0.251*
	(0.059)	(0.061)	(0.092)	(0.099)	(0.100)	(0.149)
High School Grade	0.131***	0.125***	0.129***	0.265***	0.253***	0.237***
	(0.015)	(0.015)	(0.015)	(0.027)	(0.027)	(0.028)
Lyceum	0.765**	0.813***	0.793***	0.657	0.681	0.586
	(0.310)	(0.309)	(0.297)	(0.552)	(0.552)	(0.535)
# members present	0.129	0.217	0.253	-0.292	-0.159	-0.362
	(0.354)	(0.356)	(0.355)	(0.663)	(0.667)	(0.648)
Perc. Women	0.180	0.229	0.073	0.975	0.991	0.668
	(0.653)	(0.659)	(0.655)	(1.174)	(1.182)	(1.184)
High School Grade Leader	0.001	-0.000	0.006	0.005	0.006	0.012
	(0.015)	(0.015)	(0.014)	(0.026)	(0.026)	(0.026)
Expected Grade		0.113	0.077		0.255**	0.229**
		(0.075)	(0.072)		(0.120)	(0.114)
Trust		-0.008*	-0.008*		-0.003	-0.002
		(0.005)	(0.004)		(0.008)	(0.008)
Risk Aversion		-0.054	-0.049		-0.064	-0.063
		(0.073)	(0.072)		(0.123)	(0.122)
Course Dummies	NO	NO	YES	NO	NO	YES
Constant	-4.775**	-6.367**	-7.043**	-11.041***	-15.832***	-9.259*
	(2.083)	(2.826)	(3.254)	(3.656)	(4.717)	(5.491)
Observations	413	413	413	413	413	413
Adjusted R-squared	0.182	0.189	0.248	0.202	0.206	0.255

Notes: OLS Estimates. The dependent variable is performance in the team part in columns (1)-(3) and performance in the individual part in columns (4)-(6). Standard errors (corrected for heteroskedasticity and allowing for clustering at the team level) are reported in parentheses. The symbols ***, **, * indicate that the coefficients are statistically significant at the 1, 5 and 10 percent level, respectively.

Table 6. The Impact of Female Leadership on Team Performance. Dependent Variable: Grade Team Part. OLS Estimates

	(1)	(2)	(3)	(4)	(5)
Woman Led Team	0.759**	0.700*	0.714*	0.741*	0.592*
	(0.370)	(0.375)	(0.443)	(0.435)	(0.367)
Woman		-0.564	-0.443	-0.447	-0.468
		(0.353)	(0.361)	(0.366)	(0.371)
High School Grade		0.125***	0.122***	0.117***	0.123***
		(0.015)	(0.015)	(0.015)	(0.016)
High School Grade Leader			0.011	0.010	0.010
			(0.016)	(0.017)	(0.015)
Perc. Women			-0.389	-0.257	-0.410
			(0.788)	(0.803)	(0.726)
Age				-0.119	-0.065
				(0.072)	(0.095)
Lyceum				0.664**	0.718**
				(0.302)	(0.291)
# members present				-0.209	-0.066
				(0.308)	(0.273)
Expected Grade				0.120*	0.071
				(0.068)	(0.070)
Trust				-0.007	-0.007*
				(0.005)	(0.004)
Risk Aversion				-0.051	-0.045
				(0.075)	(0.074)
Macro					-1.786***
					(0.420)
Personnel Ec.					1.201***
					(0.395)
Econometrics					-0.982
					(0.687)
Constant	5.643***	-4.463***	-4.953***	-4.143	-4.635
	(0.277)	(1.178)	(1.458)	(2.867)	(3.109)
Observations	433	433	433	433	433
Adjusted R-squared	0.010	0.144	0.142	0.157	0.247

Notes: The dependent variable is performance in the team part of the intermediate test. Standard errors (corrected for heteroskedasticity and allowing for clustering at the team level) are reported in parentheses. The symbols ***, **, * indicate that the coefficients are statistically significant at the 1, 5 and 10 percent level, respectively.

Table 7. The Impact of Female Leadership on Team Performance. Dependent Variable: Grade Team Part. OLS Estimates.

	Team members		**Leaders**	**Whole sample**
	(1)	(2)	(3)	(4)
Woman Led Team	1.146**	0.843		0.764
	(0.534)	(0.610)		(0.584)
Woman	0.062	-0.268	-0.557	-0.356
	(0.525)	(0.648)	(0.642)	(0.574)
Leader				0.604
				(0.541)
Leader*Woman				-1.458*
				(0.747)
Woman*Woman Led Team		0.706		0.726
		(0.737)		(0.703)
High School Grade Leader	0.001	0.002		0.009
	(0.018)	(0.018)		(0.015)
Perc. Women	-0.894	-1.007	-0.931	-0.959
	(1.071)	(1.079)	(1.273)	(0.820)
Age	-0.170	-0.177	0.089	-0.055
	(0.129)	(0.129)	(0.145)	(0.095)
High School Grade	0.113***	0.113***	0.151***	0.125***
	(0.020)	(0.020)	(0.026)	(0.015)
Lyceum	0.543	0.529	0.922*	0.675**
	(0.359)	(0.358)	(0.524)	(0.282)
# members present	-0.336	-0.350	0.366	-0.072
	(0.360)	(0.361)	(0.415)	(0.283)
Expected Grade	0.019	0.031	0.148	0.081
	(0.088)	(0.089)	(0.118)	(0.064)
Trust	-0.008	-0.008	-0.008	-0.007*
	(0.005)	(0.005)	(0.008)	(0.004)
Risk Aversion	-0.075	-0.068	0.000	-0.035
	(0.094)	(0.094)	(0.121)	(0.073)
Courses dummies	YES	YES	YES	YES
Observations	283	283	150	433
Adjusted R-squared	0.215	0.214	0.284	0.250

Notes: Standard errors (corrected for heteroskedasticity and allowing for clustering at the team level) are reported in parentheses. The symbols ***, **, * indicate that the coefficients are statistically significant at the 1, 5 and 10 percent level, respectively.

Table 8. The Impact of Female Leadership on Student Performance at the Individual Part. Dependent Variable: Grade Individual Part. OLS Estimates

	(1)	(2)	(3)	(4)	(5)
Woman Led Team	0.596*	0.393	0.297	0.297	0.304
	(0.313)	(0.307)	(0.373)	(0.371)	(0.334)
Woman		-0.192	-0.263	-0.221	-0.166
		(0.302)	(0.325)	(0.329)	(0.331)
High School Grade		0.130***	0.129***	0.122***	0.116***
		(0.012)	(0.012)	(0.012)	(0.013)
High School Grade Leader			0.290	0.367	0.018
			(0.658)	(0.676)	(0.629)
Perc. Women			0.003	0.003	0.001
			(0.013)	(0.013)	(0.012)
Age				-0.020	-0.140*
				(0.049)	(0.072)
Lyceum				0.240	0.241
				(0.267)	(0.257)
# members present				-0.381	-0.338
				(0.258)	(0.247)
Expected Grade				0.137**	0.118**
				(0.060)	(0.057)
Trust				-0.001	-0.000
				(0.004)	(0.004)
Risk Aversion				-0.009	-0.012
				(0.063)	(0.060)
Courses dummies	NO	NO	NO	NO	YES
Observations	433	433	433	433	433
Adjusted R-squared	0.007	0.201	0.197	0.203	0.257

Notes: Standard errors (corrected for heteroskedasticity and allowing for clustering at the team level) are reported in parentheses. The symbols ***, **, * indicate that the coefficients are statistically significant at the 1, 5 and 10 percent level, respectively.

Table 9. Time Spent Together with Teammates and Probability of Doing Homework.

	Hours Together		**Done Homework**	
	(1)	(2)	(3)	(4)
Woman Led Team	0.296	0.272	0.105	0.137*
	(0.624)	(0.638)	(0.074)	(0.075)
Woman	-0.159	-0.184		
	(0.368)	(0.473)		
Woman*Woman Led Team		0.051		
		(0.659)		
High School Grade Leader	0.074***	0.074***	0.006**	0.005**
	(0.020)	(0.020)	(0.003)	(0.003)
CONTROLS (FULL SET)	YES	YES	YES	YES
Observations	305	305	180	150
Adjusted R-squared	0.115	0.112	0.254	0.291

Notes: Columns (3) and (4) report LPM estimates. Standard errors (corrected for heteroskedasticity and allowing for clustering at the team level) are reported in parentheses. The symbols ***, **, * indicate that the coefficients are statistically significant at the 1, 5 and 10 percent level, respectively.

Table 10. Leader Behavior (Hours Together and Homework) and Student Performance. OLS Estimates

	(1) Grade Team Part	(2) Grade Team Part	(3) Individual Part	(4) Grade Team Part	(5) Grade Team Part	(6) Individual Part
Hours Together	-0.062	-0.075	-0.119*			
	(0.047)	(0.065)	(0.065)			
Homework				1.361***	1.203*	0.464
				(0.511)	(0.619)	(0.532)
CONTROLS (FULL SET)	YES	YES	YES	YES	YES	YES
Observations	412	269	269	433	283	283
Adjusted R^2	0.249	0.217	0.238	0.257	0.214	0.226

Notes: OLS Estimates. Standard errors (corrected for heteroskedasticity and allowing for clustering at the team level) are reported in parentheses. The symbols ***, **, * indicate that the coefficients are statistically significant at the 1, 5 and 10 percent level, respectively.

Table 11. Team Members' Evaluations of their Leader. OLS Estimates

	Leader Effectiveness		Leader Effort	
	(1)	(2)	(3)	(4)
Woman Led Team	-0.743	-1.284*	-0.301	-0.969
	(0.581)	(0.652)	(0.610)	(0.680)
Woman	-0.071	-0.642	0.116	-0.591
	(0.405)	(0.522)	(0.424)	(0.558)
Woman Led Team*Woman		1.178		1.457*
		(0.726)		(0.774)
High School Grade Leader	0.103***	0.105***	0.079***	0.080***
	(0.019)	(0.019)	(0.021)	(0.021)
CONTROLS (FULL SET)	YES	YES	YES	YES
Observations	305	305	305	305
Adjusted R-squared	0.241	0.246	0.199	0.207

Notes: Standard errors (corrected for heteroskedasticity and allowing for clustering at the team level) are reported in parentheses. The symbols ***, **, * indicate that the coefficients are statistically significant at the 1, 5 and 10 percent level, respectively.

Table 12. Leaders' Evaluations of their Team. OLS Estimates.

	Team Effectiveness		Required effort		Effort provided by team members	
	(1)	(2)	(3)	(4)	(5)	(6)
Woman	-0.003	0.092	0.709	1.906**	-1.152**	-1.044
	(0.093)	(0.195)	(0.473)	(0.809)	(0.589)	(1.166)
Woman* Perc: Women		-0.199		-2.512		-0.226
		(0.323)		(1.608)		(1.921)
CONTROLS	YES	YES	YES	YES	YES	YES
Observations	140	140	140	140	140	140
Adjusted R-squared	0.079	0.074	-0.014	-0.000	0.039	0.031

Notes: Standard errors (corrected for heteroskedasticity and allowing for clustering at the team level) are reported in parentheses. The symbols ***, **, * indicate that the coefficients are statistically significant at the 1, 5 and 10 percent level, respectively.

Figure 1. Grade Team part by treatment

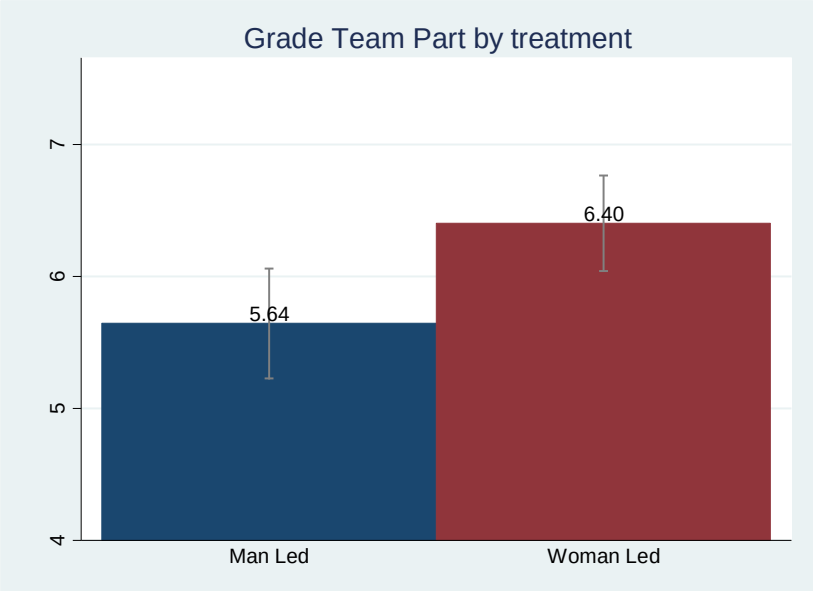
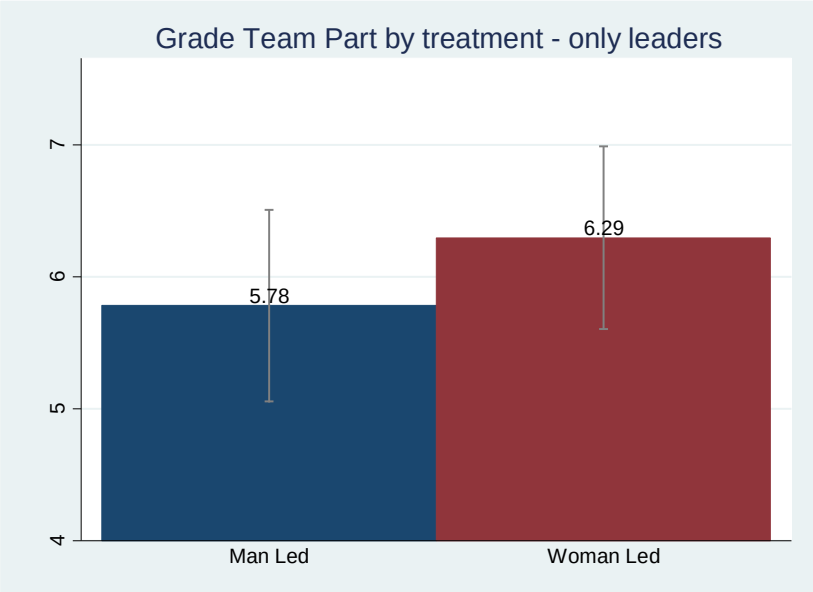
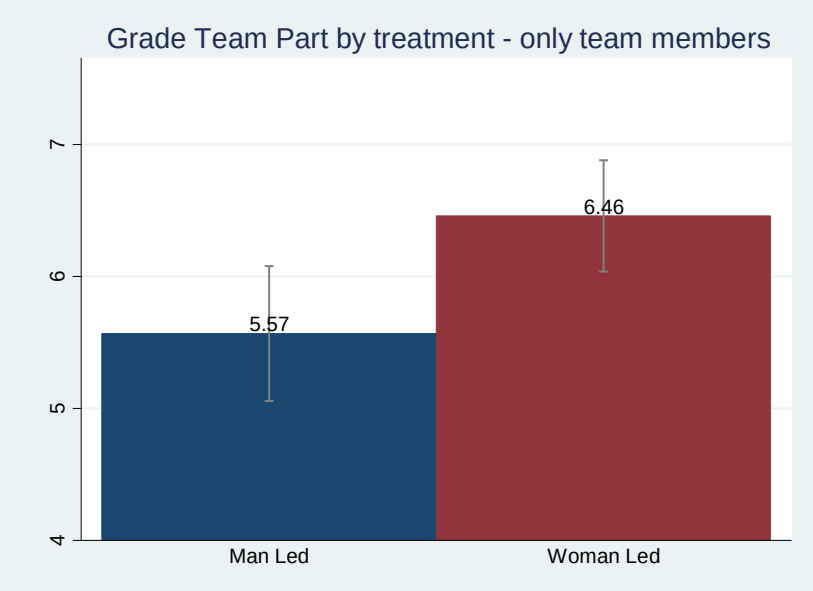


Figure 2. Grade Team part by treatment in the subsamples of leaders (a) and team members (b-c)

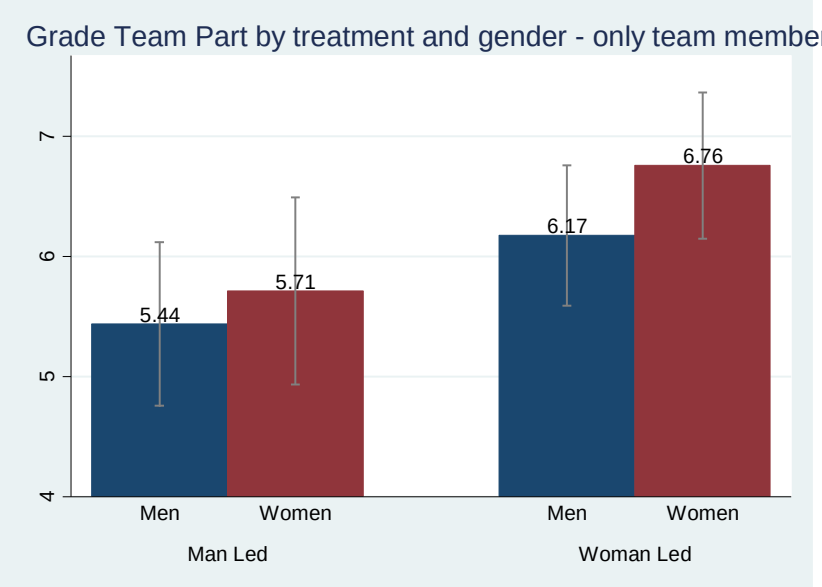
(a)



(b)


Grade Team Part by treatment - only team members
7
6
5
4
6.46
5.57
Man Led
Woman Led

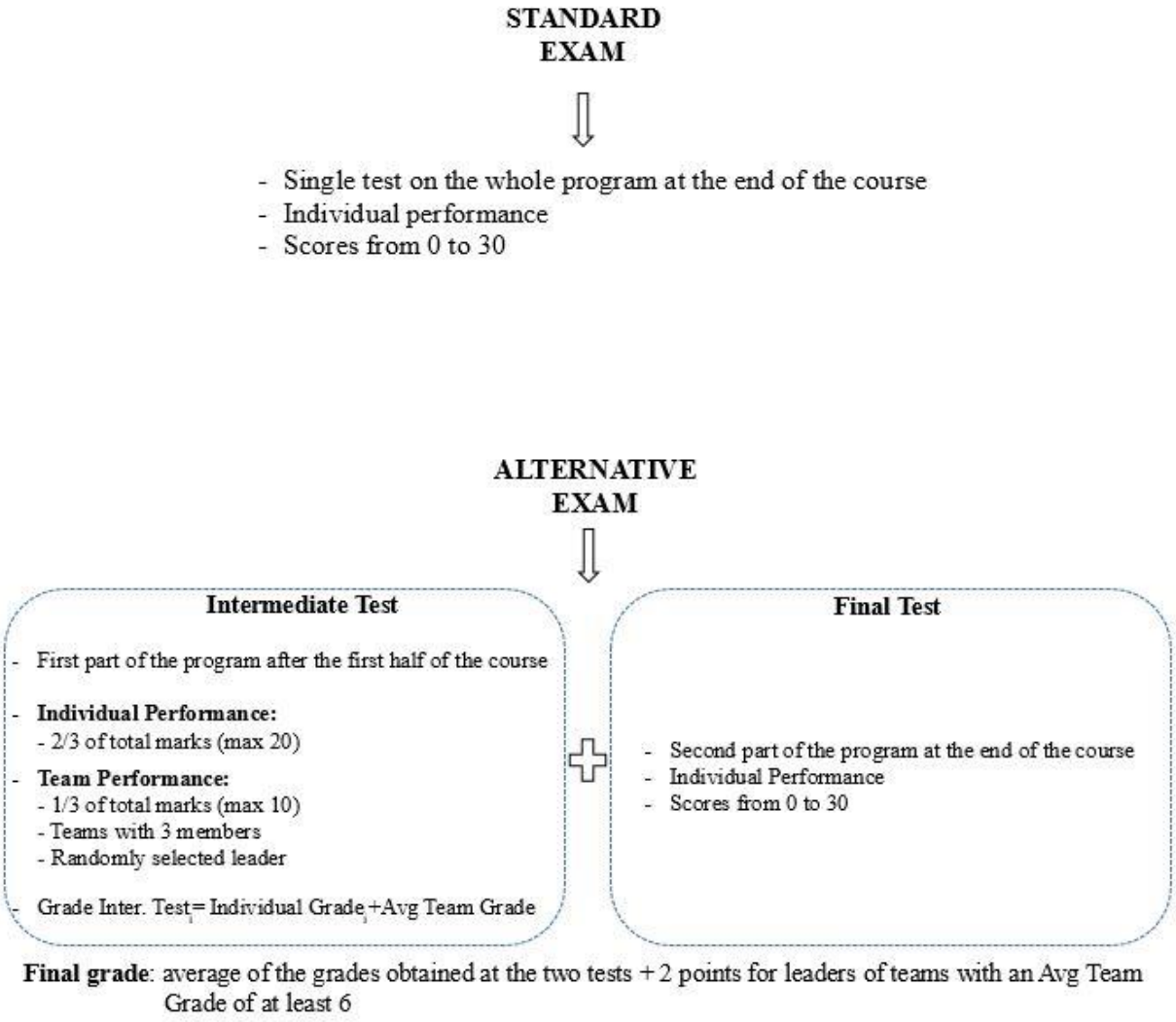
(c)


Grade Team Part by treatment and gender - only team members
7
6
5
4
6.76
6.17
5.71
5.44
Men
Women
Men
Women
Man Led
Woman Led

Appendix A

In Figure A1 we report a diagram with the organization of the exam both under the standard scheme and in the experimental setting.

Figure A1. Exam organization



Appendix B

In this appendix we run the same estimates reported in Section 3 and Section 4 on a restricted sample including leaders with a *High School Grade* in the interquartile range of male leaders' ability (that is higher than 70 and lower than 86) in order to be reassured that the effect we find is not driven by leader's ability. All in all, our results continue to hold true and become larger in magnitude.

Table B1. The Impact of Female Leadership on Team Performance. Restricted sample. Dependent Variable: Grade Team Part. OLS Estimates

	(1)	(2)	(3)	(4)	(5)
Woman Led Team	0.921*	1.264**	1.434**	1.313*	0.925*
	(0.558)	(0.595)	(0.681)	(0.666)	(0.555)
Woman		-0.721	-0.451	-0.533	-0.481
		(0.612)	(0.628)	(0.650)	(0.622)
High School Grade		0.124***	0.120***	0.116***	0.106***
		(0.026)	(0.027)	(0.028)	(0.030)
High School Grade Leader			0.071	0.089	0.084
			(0.068)	(0.078)	(0.064)
Perc. Women			-0.829	-0.679	-0.850
			(1.255)	(1.296)	(1.190)
Age				-0.121	-0.094
				(0.131)	(0.154)
Lyceum				0.386	0.533
				(0.494)	(0.473)
# members present				-0.192	-0.152
				(0.487)	(0.440)
Expected Grade				0.191	0.155
				(0.116)	(0.120)
Trust				-0.001	-0.003
				(0.007)	(0.007)
Risk Aversion				0.068	0.113
				(0.120)	(0.117)
Macro					-1.827***
					(0.693)
Personnel Ec.					0.985
					(0.726)
Econometrics					-0.570
					(1.219)
Constant	5.568***	-4.467**	-9.545*	-12.671*	-11.006
	(0.409)	(2.156)	(5.450)	(7.246)	(7.074)
Observations	182	182	182	182	182
Adjusted R-squared	0.012	0.100	0.099	0.094	0.176

Notes: Standard errors (corrected for heteroskedasticity and allowing for clustering at the team level) are reported in parentheses. The symbols ***, **, * indicate that the coefficients are statistically significant at the 1, 5 and 10 percent level, respectively.

Table B2. The Impact of Female Leadership on Team Performance. Restricted sample. Dependent Variable: Grade Team Part. OLS Estimates.

	Team members		Leaders	Whole sample
	(1)	(2)	(3)	(4)
Woman Led Team	1.790**	1.710*		1.577*
	(0.795)	(0.959)		(0.933)
Woman	0.515	0.434	-0.400	0.124
	(0.950)	(1.147)	(0.946)	(0.972)
Leader				1.541*
				(0.913)
Leader*Woman				-1.912
				(1.263)
Woman*Woman Led Team		0.160		0.106
		(1.190)		(1.181)
High School Grade Leader	0.044	0.044		0.081
	(0.080)	(0.081)		(0.071)
Perc. Women	-2.016	-1.997	-1.402	-1.603
	(1.820)	(1.834)	(1.957)	(1.327)
Age	-0.226	-0.224	0.189	-0.068
	(0.173)	(0.173)	(0.312)	(0.161)
High School Grade	0.090***	0.090***	0.295**	0.112***
	(0.034)	(0.034)	(0.114)	(0.028)
Lyceum	0.357	0.348	0.556	0.420
	(0.591)	(0.589)	(0.890)	(0.479)
# members present	-0.883	-0.886	0.975	-0.155
	(0.547)	(0.548)	(0.667)	(0.465)
Expected Grade	0.190	0.191	0.156	0.166
	(0.146)	(0.149)	(0.198)	(0.105)
Trust	-0.000	-0.000	-0.007	-0.004
	(0.008)	(0.007)	(0.012)	(0.006)
Risk Aversion	-0.000	-0.002	0.365*	0.139
	(0.149)	(0.149)	(0.186)	(0.116)
Courses dummies	YES	YES	YES	YES
Observations	119	119	63	182
Adjusted R-squared	0.224	0.216	0.096	0.182

Notes: Standard errors (corrected for heteroskedasticity and allowing for clustering at the team level) are reported in parentheses. The symbols ***, **, * indicate that the coefficients are statistically significant at the 1, 5 and 10 percent level, respectively.

Table B3. The Impact of Female Leadership on Student Performance at the Individual Part. Restricted sample. Dependent Variable: Grade Individual Part. OLS Estimates

	(1)	(2)	(3)	(4)	(5)
Woman Led Team	0.954*	1.119**	1.168*	0.990*	0.919*
	(0.503)	(0.525)	(0.590)	(0.571)	(0.522)
Woman		-0.191	-0.024	-0.053	0.029
		(0.465)	(0.481)	(0.502)	(0.502)
High School Grade		0.134***	0.128***	0.121***	0.101***
		(0.022)	(0.022)	(0.022)	(0.024)
High School Grade Leader			0.102*	0.114*	0.071
			(0.054)	(0.058)	(0.052)
Perc. Women			-0.375	-0.062	-0.565
			(1.040)	(1.072)	(1.061)
Age				-0.069	-0.237*
				(0.095)	(0.140)
Lyceum				0.191	0.208
				(0.425)	(0.404)
# members present				-0.764*	-0.744*
				(0.391)	(0.392)
Expected Grade				0.244**	0.209**
				(0.102)	(0.098)
Trust				0.000	-0.001
				(0.006)	(0.005)
Risk Aversion				0.099	0.108
				(0.096)	(0.091)
Courses dummies	NO	NO	NO	NO	YES
Observations	182	182	182	182	182
Adjusted R-squared	0.020	0.172	0.182	0.220	0.267

Notes: Standard errors (corrected for heteroskedasticity and allowing for clustering at the team level) are reported in parentheses. The symbols ***, **, * indicate that the coefficients are statistically significant at the 1, 5 and 10 percent level, respectively.

Table B4. Time Spent Together with Teammates and Probability of Doing Homework. Restricted sample.

	Hours Together		**Done Homework**	
	(1)	(2)	(3)	(4)
Woman Led Team	-1.030	-1.279	0.151	0.202*
	(0.938)	(0.931)	(0.101)	(0.101)
Woman	0.002	-0.236		
	(0.527)	(0.714)		
Woman*Woman Led Team		0.446		
		(0.946)		
High School Grade Leader	0.167	0.167	0.001	-0.001
	(0.118)	(0.119)	(0.012)	(0.012)
CONTROLS (FULL SET)	YES	YES	YES	YES
Observations	132	132	76	63
Adjusted R-squared	0.070	0.063	0.274	0.392

Notes: Columns (3) and (4) report LPM estimates. Standard errors (corrected for heteroskedasticity and allowing for clustering at the team level) are reported in parentheses. The symbols ***, **, * indicate that the coefficients are statistically significant at the 1, 5 and 10 percent level, respectively.

Table B5.Leader Behavior (Hours Together and Homework) and Student Performance. Restricted sample. OLS Estimates

	(1) Grade Team Part	(2) Grade Team Part	(3) Individual Part	(4) Grade Team Part	(5) Grade Team Part	(6) Individual Part
Hours Together	-0.106	-0.123	-0.071			
	(0.069)	(0.077)	(0.064)			
Homework				2.226***	2.122**	0.955
				(0.813)	(0.988)	(0.883)
CONTROLS (FULL SET)	YES	YES	YES	YES	YES	YES
Observations	175	116	116	182	119	119
Adjusted R^2	0.165	0.185	0.203	0.201	0.222	0.205

Notes: OLS Estimates. Standard errors (corrected for heteroskedasticity and allowing for clustering at the team level) are reported in parentheses. The symbols ***, **, * indicate that the coefficients are statistically significant at the 1, 5 and 10 percent level, respectively.

Table B6. Team Members' Evaluations of their Leader. Restricted sample. OLS Estimates

	Leader Effectiveness		**Leader Effort**	
	(1)	(2)	(3)	(4)
Woman Led Team	-1.024	-1.739	0.060	-0.612
	(1.032)	(1.197)	(1.024)	(1.196)
Woman	0.083	-0.600	0.761	0.120
	(0.788)	(0.981)	(0.781)	(0.990)
Woman Led Team*Woman		1.282		1.205
		(1.342)		(1.355)
High School Grade Leader	-0.026	-0.025	-0.121	-0.120
	(0.110)	(0.109)	(0.115)	(0.115)
CONTROLS (FULL SET)	YES	YES	YES	YES
Observations	132	132	132	132
Adjusted R-squared	0.089	0.089	0.107	0.106

Notes: Standard errors (corrected for heteroskedasticity and allowing for clustering at the team level) are reported in parentheses. The symbols ***, **, * indicate that the coefficients are statistically significant at the 1, 5 and 10 percent level, respectively.

Table B7. Leaders' Evaluations of their Team. Restricted sample. OLS Estimates.

	Team Effectiveness		**Required effort**		**Effort provided by team members**	
	(1)	(2)	(3)	(4)	(5)	(6)
Woman	0.022	0.583	1.155*	0.728	-1.791**	2.251
	(0.140)	(0.351)	(0.613)	(0.870)	(0.705)	(1.967)
Woman* Perc: Women		-1.019*		0.776		-7.339**
		(0.553)		(1.968)		(3.303)
CONTROLS	YES	YES	YES	YES	YES	YES
Observations	58	58	58	58	58	58
Adjusted R-squared	-0.035	0.012	0.070	0.053	0.061	0.160

Notes: Standard errors (corrected for heteroskedasticity and allowing for clustering at the team level) are reported in parentheses. The symbols ***, **, * indicate that the coefficients are statistically significant at the 1, 5 and 10 percent level, respectively.

Appendix C

In Table C1 we report descriptive statistics on individual pre-determined characteristics for man led teams and for woman led teams (as in Table 4) in the subsample of students joining the experiment.

Table C1. Students' Characteristics across Treatment Groups separately for leaders and team members. Students Joining the Experiment

	Leaders			Team Members					
	Woman Led Team	Man Led Team	Diff. (p-value)	Women	Men	Diff. (p-value)	Woman Led Team	Man Led Team	Diff. (p-value)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Woman							0.487	0.496	0.007
							(0.501)	(0.501)	(0.894)
Perc. Women							0.484	0.477	0.007
							(0.370)	(0.342)	(0.845)
Age	22.221	22.005	0.216	21.879	22.196	-0.317	21.909	22.189	-0.281
	(2.512)	(2.190)	(0.350)	(2.218)	(2.495)	(0.295)	(2.090)	(2.636)	(0.268)
High School Grade	85.649	78.477	7.172	85.156	80.373	4.783	82.594	82.783	-0.190
	(10.425)	(10.759)	(1.582)	(10.588)	(9.970)	(0.000)	(10.886)	(10.167)	(0.864)
Lyceum	0.457	0.581	-0.124	0.530	0.541	-0.020	0.551	0.509	0.042
	(0.501)	(0.496)	(0.074)	(0.501)	(0.500)	(0.701)	(0.499)	(0.501)	(0.428)
Expected Grade	25.106	24.709	0.397	25.058	24.810	0.247	24.877	24.988	-0.111
	(2.197)	(2.748)	(0.373)	(2.475)	(2.533)	(0.351)	(2.755)	(2.204)	(0.672)
Trust	50.851	58.372	-7.521	52.717	61.946	-9.229	60.749	53.918	6.830
	(33.909)	(36.197)	(5.241)	(31.031)	(37.366)	(0.011)	(34.164)	(35.052)	(0.063)
Risk Aversion	5.063	4.384	0.680	4.948	4.914	0.034	4.995	4.860	0.135
	(2.129)	(2.176)	(0.321)	(2.316)	(2.081)	(0.883)	(2.065)	(2.332)	(0.564)
Obs.	94	86		173	185		187	171	

Notes: Standard deviations are reported in parentheses. *p*-values are reported in the columns with differences (3, 6 and 9).

Appendix D

In Table D1 we run the same estimate as in Table 7 column (1) separately for the gender of the team members.

Table D1. The Impact of Female Leadership on Team Performance for Team Members by gender. Dependent Variable: Grade Team Part. OLS Estimates.

	Team members	
	Women	**Men**
	(1)	(2)
Woman Led Team	1.495*	1.121
	(0.858)	(0.712)
High School Grade Leader	-0.009	0.010
	(0.028)	(0.025)
Perc. Women	-0.369	-1.731
	(1.639)	(1.517)
Age	-0.202	-0.150
	(0.186)	(0.178)
High School Grade	0.105***	0.120***
	(0.029)	(0.029)
Lyceum	0.381	0.569
	(0.510)	(0.532)
# members present	-0.403	-0.322
	(0.609)	(0.457)
Expected Grade	0.039	0.050
	(0.119)	(0.143)
Trust	-0.020***	-0.000
	(0.008)	(0.007)
Risk Aversion	-0.001	-0.147
	(0.123)	(0.156)
Courses dummies	YES	YES
Observations	135	148
Adjusted R-squared	0.268	0.149

Notes: Standard errors (corrected for heteroskedasticity and allowing for clustering at the team level) are reported in parentheses. The symbols ***, **, * indicate that the coefficients are statistically significant at the 1, 5 and 10 percent level, respectively.

Appendix E

In Table E1 we run the same estimates as in Table 11 columns (2) and (4) on the sample of students who undertook the intermediate test.

Table E1. Team Members' Evaluations of their Leader. Students taking the intermediate test. OLS Estimates

	Leader Effectiveness	**Leader Effort**
	(1)	(2)
Woman Led Team	-1.355**	-1.189*
	(0.685)	(0.706)
Woman	-0.693	-0.832
	(0.574)	(0.609)
Woman Led Team*Woman	1.198	1.660**
	(0.771)	(0.830)
High School Grade Leader	0.098***	0.071***
	(0.020)	(0.022)
CONTROLS (FULL SET)	YES	YES
Observations	268	268
Adjusted R-squared	0.263	0.232

Notes: Standard errors (corrected for heteroskedasticity and allowing for clustering at the team level) are reported in parentheses. The symbols ***, **, * indicate that the coefficients are statistically significant at the 1, 5 and 10 percent level, respectively.