Team Production, Leadership and Gender*

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

Francesca Gioia
(School of Economics, University of Edinburgh)

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 depends on the gender of the team leader. In the experiment we involved about 430 students from an Italian University who took an intermediate exam that was partly evaluated on the basis of team work. Students were randomly matched in teams of three and in each team the leader was randomly chosen. In this way, we have male and female leaders and different composition of team by gender. We find a positive and significant effect of female leadership on team performance. This effect is mainly driven by the higher performance of team members in female led teams rather than due to the leader performance. In fact, male and female leader show a similar individual performance. We also find that in spite of the higher performance of female led teams, female and male leaders are evaluated as equally effective.

JEL classification: J16, M12;M54; C93.

Keywords: Team; Leadership; Gender; Randomized Experiment

* 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, School of Economics, University of Edinburgh, 30 Buccleuch Place, Edinburgh, EH8 9JT, UK. E-mail: f.gioia@sms.ed.ac.uk; 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.
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…, and seminar participants at the University of Calabria for useful comments and suggestions. Francesca Gioia gratefully acknowledges financial support from Economic and Social Research Council.

1. Introduction

Women have made progress in many social and economic dimensions, but they are still heavily underrepresented in the apical roles. In 2017, women hold around 20% of the seats in the U.S. Congress. Similarly, the proportion of women in the national Parliaments of many European countries is well below the equalitarian share (France 25%, UK 30%, Italy 31%, Germany 37%)[1]. Moreover, in the US women represent only 4% of CEOs and hold only 19% of board seats. Likewise, only 21% of board members of the largest publicly listed companies in the EU are women.[2] The so called glass ceiling, preventing women from rising to top positions, has long attracted the attention of economists, sociologists and experts of organizations and management. The related literature has offered a number of explanations based on differences in productivity, discrimination and differences in psychological attitudes (such as risk aversion and competitiveness). An alternative explanation focuses on gender differences in leadership ability: individuals holding top positions are often involved in team working and have to deal with problems such as free riding and coordination failure for which leadership can represent a viable solution. Then, the gender gap might depend on the fact that men are more effective leaders than women. It could be, for instance, that women are less credible, have weaker communication skills or are less able to inspire and head up people.

In this paper we investigate if men and women differ in their leadership skills. By means of a field experiment, we study the causal effect of leader's gender on team performance in a real life environment represented by the preparation of a university exam. We also investigate whether male and female leaders are evaluated differently by their team members and the dimensions along which their team's activity differs.

Our experiment has involved students enrolled at four different economic courses in an Italian University. They were offered the possibility to sit to an intermediate exam of these courses according to an alternative examination scheme introducing team work. A total of 538 (out of 743) students joined the experiment. They were randomly assigned to teams composed by three members and, within each team, a randomly selected member was appointed the role of leader, in order to exclude selection issues. The leader had the task of contacting team members (by e-mail or phone calls) in order to define team meetings and to organize team studying activities. These activities consisted in solving one set of exercises assigned during teaching classes (the leader was responsible to deliver the solved exercises to the course's instructor) and preparing together the part of the exam evaluated on the basis of team performance. As a compensation for such additional activities, the leaders could receive a fixed 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). The composition of the teams, team leaders and the parts of the program evaluated on the basis of team performance were soon communicated to students. After taking the test, students were asked to fill out a final questionnaire containing questions on the activity of the team and individual's evaluations of team effectiveness and leadership activity.

We find that teams with a randomly selected female leader significantly outperform teams

[1] http://www.ipu.org/wmn-e/classif.htm.
[2] http://ec.europa.eu/justice/gender-equality/files/womenonboards/factsheet_women_on_boards_web_2015-10_en.pdf

coordinated by a male leader: the performance of students in female led teams is on average 0.759 points higher than that of students in male led teams. Such effect corresponds to 0.23 SD of the dependent variable and is robust to the inclusion of several control variables. When digging deeper into this effect we find that women appointed as leaders devote altruistically energies to improve the performance of the whole team and this slightly weakens their own performance.

Using data from the assigned homework and the final questionnaire, we analyze if team members' evaluation of leader's effort and effectiveness and several dimensions of team activity vary with the leader's gender. We find that, in spite of a better performance of female led teams, female leaders do not obtain better evaluations from team members compared to male leaders. As regards team activity, we find that female led teams are significantly more likely to submit their homework and spend on average more hours working together. These results suggest that female leaders have dedicated more effort than males in calling team members' meetings and make the team members study together. The additional studying activity has however focused only on the team part and on the homework, while no effects are found for studying on the individual part or for leisure time.

The economic literature investigating leadership mainly relies on laboratory experiments using minimum effort coordination games in which the lowest performing player determines group performance. These studies show that, in the absence of communication, coordination failure can be very common, while costless, non-binding pre-play communication between players can improve coordination and efficiency (Van Huyck et al., 1990; Blume and Ortmann, 2007; Devetag and Ortmann, 2007). Moreover, a number of works show that, when communication between all workers is unfeasible, leaders can work as a "coordination device" and improve coordination using one-way communication to convince other team members that everyone will exert high effort (Weber et al. 2001; Brandts and Cooper, 2007; Kriss and Eil, 2012; Sahin et al. 2015). Leader effectiveness can be enhanced through mechanisms that strength legitimacy such as democratic election (Brandts et al., 2015) and through different communication modes (Kriss and Eil, 2012).

Only few papers consider gender differences in leader effectiveness finding ambiguous results. 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. They show that male leaders have a greater impact of followers' decisions compared to female leaders. Moreover, followers are less likely to both attribute success to female leaders and reward them generously. Reuben and Timko (2017) extend the work of Grossman et al. (2016) by considering gender differences between elected and randomly-selected leaders and by analyzing gender differences in re-election. They find evidence of gender difference in the effectiveness of leaders only for elected leaders, while no difference emerges for randomly-selected ones. They also find that unsuccessful female leaders are re-elected at considerable lower rates than unsuccessful male leaders. No gender differences in performance with randomly selected leaders are also highlighted by Timko (2017), while weak differences are reported by Dufwenberg and Gneezy (2005) who, using a minimum effort coordination game, investigate the effect of team composition on team performance

(differences in the fraction of men and women in a team only slightly affect coordination effectiveness).

All in all these findings suggest that gender differences in leadership effectiveness might depend on the precise context in which they are studied. We enrich this literature by studying the effectiveness of female leaders in the field, in a real life environment in which they have strong incentives to perform well. 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.

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

2. Experimental Design and Data

2.1. Design and procedure

We ran a field experiment involving 433 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.[3] Courses are worth 10 credits each, corresponding to 60 hours of teaching and to nominal 250 hours of study, and are taught to students during the second semester (teaching period from February to May).

Students enrolled in the courses by filling out an online survey asking questions on family background, risk preferences, social attitudes and expectations on performance. At the beginning of the courses, students were told that they could choose whether to sit the standard final exam at the end of the course or join an alternative examination scheme. The alternative scheme consisted in two tests, a intermediate test on half of the course program to sit right after the first half of the course and a second test on the remaining part of the program to sit at the end of the course. As required by the university administration for ethical reasons, if choosing the alternative scheme, students could however change their mind and sit the standard exam.

The intermediate test was composed by two parts, one evaluated on the basis of individual performance ("the individual part") and the other evaluated on the basis of team performance ("the team part"). The individual part counted for 2/3 of the total marks (students could gain a maximum of 20 points). In the team part, instead, students could gain a maximum of 10 points (1/3 of the total marks) and their score was given by the average performance of their team. Teams were composed by three members, one of which

[3] 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.

was appointed the role of 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[4] assigned during the classes and submitting them to the course's professor. As a compensation for such additional activities, the leader could receive a fixed 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). The second test was evaluated based only on individual performance with scores ranging from 0 to 30 (as in the standard exam). The final exam grade of the alternative examination scheme was given by the average of the grades obtained in the two tests.

Students were given 5 days to choose whether to sit the standard exam or join the alternative exam scheme. Once 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: *Female Led Team* for teams with a female leader and *Male 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.[5] Within each course, all students attended the lectures in the same room, at the same time and with the same instructor and teaching material. After the first half of the teaching classes students undertook the intermediate test. All students took the test with the same questions and at the same time.

After the test students were asked to fill out an online survey available on the courses' website. Answering the final survey was not mandatory but strongly encouraged. Students could answer until the exam grades were published. The aim of such final questionnaire was to collect information on the activity of the team and individual's evaluations of team effectiveness and leadership activity. Questions were identical for leaders and members, except for that on leader's evaluation which was phrased slightly differently to elicit teammates' evaluation of the effort and effectiveness of the leader and leaders' evaluation of how much demanding they find their role.

2.2. Descriptive Statistics and Balance Checks

The design of the experiment produced three subsamples of students: those who enrolled in the courses (743), those who joined the experiment (538) and those who actually showed up at the intermediate test (433).

In Table 1 we provide descriptive statistics separately for the three subsamples of students. About 50% of students attending the courses and joining the experiment are women, while 49% of women took the test. Students in all subsamples are on average 22 years old.

[4] Solving the exercises was not mandatory for the exam. We analyse the probability of solving and submitting the homework in Section 5.

[5] We informed students that if one of the students did not show up at the intermediate exam, in order to compute the score obtained in the team part of the test by his/her teammate we would replace the absent student with a randomly selected student sitting the test.

Students enrolled in the courses and joining the experiment have an average *High School Grade* of 82.5 (*High School Grade* ranges between 60 and 100), while *High School Grade* becomes slightly higher among students taking the test (83.2), suggesting a selection of better students in the test. About 51% of students enrolled in the courses have studied in a *Lyceum*. This percentage becomes higher in the other two subsamples (about 53% and 54% for students joining the experiment and sitting the exam, respectively).

Table 1. Descriptive Statistics. Mean and SD

	Enrolled at the courses	Joining the Experiment	Taking the Intermediate test
Female	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)
Female Led Team		0.522	0.517
		(0.500)	(0.500)
Perc. Females		0.496	0.498
		(0.293)	(0.296)
Expected Grade		24.926	25.141
		(2.493)	(2.303)
Parents' Education		12.007	12.059
		(3.164)	(3.188)
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.

For students joining the experiment and taking the test we have also information on their family background, preferences and team assignment. One third of students has been assigned to the role of leader and 52% of the teams are led by a female. On average about half of the team members are female.

Students expect to score at least 25 in both subsamples and their parents have studied for 12 years on average.

Students' answers to the on-line survey filled-in when joining the experiment allow 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 euro with probability 50% or 0 euro with probability 50% (that is, when tossing a coin, head means winning 100 euro while tail means winning zero)." Students could choose their favourite option between the lottery and a certain amount in a multiple price list where the certain amount increased from 10 euro to 90 euro in steps of 10 euro. We built the variable *Risk Aversion* taking values from 0 (for students preferring the lottery to a certain amount of 90 euro) to 9 (for students preferring 10 euro with certainty to the lottery). *Risk Aversion* is on average 4.8 in both subsamples.

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 euro 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 euro, B will receive 240 euro). 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.

On average 2.6 team members show up at the intermediate test; in 68.6% of teams all members take the intermediate test, in 26.8% of teams two members show up and in only 4.6% of teams only one member takes the test. Finally, students taking the test obtain an average grade of 6 at the team part and of 11.8 at the individual part.[6]

To investigate whether the gender of the leader matters for team performance we need comparable individuals in teams led by males and by females. In Table 2 we report balance checks by studying the impact of individual characteristics on the probability of being assigned to a female led team in the subsample of students joining the experiment and in the subsample of students showing up at the intermediate test, conditional on being a female. Obviously, being a female increases the probability of having a female as a leader in both subsamples. The other characteristics are equally balanced in the two treatments except for *Parents' Education* that is negatively associated to the probability of being assigned to a female led team.

[6] In the final sample, 39.5% of students are enrolled at the course of Microeconomics, 23.1% of Macroeconomics, 25.4% of Personnel Economics and 12% study Econometrics.

Table 2. Balance Checks. The probability of being assigned to a Female Led Team as a function of individual characteristics. Dependent variable: Female Led Team

	Joining the Experiment	Taking the Intermediate test
	(1)	(2)
Female	0.326***	0.348***
	(0.043)	(0.048)
Age	-0.013	-0.011
	(0.011)	(0.013)
High School Grade	0.001	-0.001
	(0.002)	(0.002)
Lyceum	0.014	0.018
	(0.043)	(0.048)
Expected Grade	-0.004	-0.000
	(0.008)	(0.009)
Parents' Education	-0.013*	-0.016**
	(0.007)	(0.007)
Trust	0.001	0.001
	(0.001)	(0.001)
Risk Aversion	0.014	0.016
	(0.009)	(0.010)
Macro	-0.047	-0.095
	(0.055)	(0.061)
Personnel Ec.	0.043	0.032
	(0.056)	(0.063)
Econometrics	0.115	0.099
	(0.091)	(0.099)
Observations	538	433
Adjusted R-squared	0.120	0.139

Notes: OLS estimates. 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.

2.3. 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 team and to solve a number of exercises together, we are not able to directly check if this actually happened or students worked separately, disregarding the 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 team working hours of 3.87, which rises at 4 for students who show up at the intermediate test. These answers suggest that team members have effectively worked together.

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 absence of cooperation, we should observe no correlation among teammates' performance – just as we detect no correlation among their predetermined measures of abilities, such as *High School Grades*, *Expected Grade*, and so on.

In the first three columns of Table 3 we regress team performance of student *i* on the average performance on the team part of his/her teammates (*Team Grade Teammates (avg)*). We control for courses dummies, *Female*, *High School Grade*, *Age,* and other individual characteristics. We find a very 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 0.27-0.30 points with a *t*-stat of 4.47-4.37. This evidence strongly supports the idea that team members have worked together and are affected by common factors.

We find a similar effect, although smaller in magnitude, for the performance in the individual part of the exam. In columns (4)-(6) we regress student *i*'s individual grade on the average grade of his/her teammates on the individual part. An increase of 1 point in the (average) individual performance of teammates significantly increases by about 0.18-0.21 points the performance of student *i*.

While the effect on the team part corresponds to about 0.08 SD, the effect on the individual part is equal to 0.03 SD. These findings suggest that students have worked together in the team part but there were spillovers also on the individual part of the exam.

Table 3. Correlations of performance among teammates

	(1)	(2)	(3)	(4)	(5)	(6)
	Grade Team part			Grade Individual part		
Team Grade Teammates (avg)	0.301***	0.279***	0.277***			
	(0.067)	(0.063)	(0.063)			
Individual Grade Teammates (avg)				0.207***	0.192***	0.183***
				(0.068)	(0.060)	(0.059)
Macro	-1.194***	-1.426***	-1.389***	0.122	-0.140	0.027
	(0.438)	(0.438)	(0.439)	(0.758)	(0.737)	(0.731)
Personnel Ec.	1.039**	0.817**	0.794**	3.352***	3.028***	3.057***
	(0.413)	(0.388)	(0.383)	(0.825)	(0.786)	(0.775)
Econometrics	-0.179	-0.643	-0.852	1.659**	1.192	1.340
	(0.454)	(0.677)	(0.677)	(0.799)	(1.155)	(1.185)
Female		-0.468	-0.431		-0.294	-0.177
		(0.295)	(0.301)		(0.527)	(0.546)
High School Grade		0.125***	0.124***		0.244***	0.233***
		(0.015)	(0.015)		(0.024)	(0.025)
Age		-0.073	-0.062		-0.228	-0.259*
		(0.090)	(0.092)		(0.141)	(0.142)
Lyceum			0.741***			0.553
			(0.285)			(0.513)
# members present			0.029			-0.464
			(0.243)			(0.470)
Expected Grade			0.076			0.228**
			(0.069)			(0.109)
Trust			-0.006			-0.000
			(0.004)			(0.007)
Risk Aversion			-0.033			-0.008
			(0.069)			(0.115)
Constant	4.253***	-4.010*	-6.047**	8.280***	-6.553*	-9.682**
	(0.516)	(2.385)	(2.970)	(0.840)	(3.837)	(4.915)
Observations	433	433	433	433	433	433
Adjusted R-squared	0.142	0.277	0.285	0.097	0.273	0.276

Notes: The dependent variable is Team Performance in columns (1)-(3) and Individual Performance 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.

3. The impact of female leadership on team performance

In this Section, we answer our main research question, that is if the gender of the leader influences the performance obtained by the team.

We do so by using data at the student level and estimating several specifications of the following OLS model:

$$[1] \qquad GradeTeamPart_i = \beta_0 + \beta_1 FemaleLedTeam_i + \beta_3 Female_i + \beta_4 X_i + \beta_5 W_i + \mu_j + \varepsilon_i$$

where the dependent variable, $GradeTeamPart_i$ is the score that student *i* obtains in the team part of the test and $FemaleLedTeam_i$ is a dummy variable for the treatment status that takes value 1 for students assigned to a team led by a female and 0 for the reference category that is a team led by a male. We control for the student's gender, a vector X_i of variables measuring the number of team members who showed up at the intermediate test, student's predetermined characteristics and individual ability (*# Members Present, Age, High School Grade, Lyceum*) and a vector W_i of variables measuring family background, expectations and preferences (*Parents' Education, Expected Grade, Risk Aversion, Trust*) and courses fixed effects μ_j (dummies for Microeconomics; Macroeconomics; Personnel Economics; Econometrics). ε_i is an error term. In this model, β_1 represents the causal effect in terms of student's performance at the team part of being assigned to a team led by a female.

We present OLS estimates of the impact of female leadership on student's team performance in Table 4. In all our regressions Standard Errors are corrected for heteroskedasticity and clustered at the team level.

The first specification controls only for the dummy *Female Led Team*. We find a positive, strong and statistically significant effect of female leadership: a student in a team led by a female obtains 0.759 points more than a student in a team led by a male. The effect corresponds to about 0.23 SD of the dependent variable.

In the second column we add the dummy *Female* to verify if the difference is due to the student's gender: we find a positive although not significant impact of *Female* on team performance[7]; more importantly, the effect of being in a team led by a woman remains positive and highly statistically significant with a small reduction in its magnitude (from 0.76 to 0.67).

In column (3) we add the vector X_i to control for the number of team members effectively present at the exam (weaker student could decide to skip the exam and this could have an impact on remaining student) and for individual characteristics and ability: *High School Grade, Lyceum, Age*. The impact of female leadership remains almost unchanged and statistically significant. *High School Grade* and *Lyceum* are strong predictors of students' academic performance.

[7] The coefficient of the gender dummy changes direction when we control for student's ability. This is not surprising as female students in our sample have on average a higher high school grade.

In column (4) we add the vector W_i including family background and risk preferences and social attitudes that we elicited with the online survey we proposed to students at the beginning of teaching classes. Final, in column (5) we also add course dummies.

In all the specifications we find that teams led by a woman tend to perform significantly better.

Table 4. The impact of female leadership on team performance. Dependent variable: Grade Team Part. OLS Estimates.

	(1)	(2)	(3)	(4)	(5)
Female Led Team	0.759**	0.676*	0.705*	0.715*	0.525*
	(0.370)	(0.398)	(0.374)	(0.364)	(0.311)
Female		0.231	-0.465	-0.557	-0.615*
		(0.377)	(0.357)	(0.366)	(0.327)
# members present			-0.257	-0.217	-0.079
			(0.300)	(0.306)	(0.272)
Age			-0.104	-0.115	-0.061
			(0.067)	(0.070)	(0.095)
High School Grade			0.126***	0.120***	0.126***
			(0.015)	(0.014)	(0.015)
Lyceum			0.657**	0.780**	0.830***
			(0.297)	(0.304)	(0.307)
Expected Grade				0.119*	0.070
				(0.068)	(0.070)
Parents' Education				-0.057	-0.055
				(0.051)	(0.049)
Trust				-0.008*	-0.008*
				(0.005)	(0.004)
Risk Aversion				-0.051	-0.044
				(0.076)	(0.073)
Macro					-1.810***
					(0.415)
Personnel Ec.					1.166***
					(0.395)
Econometrics					-0.986
					(0.688)
Constant	5.643***	5.573***	-2.012	-3.001	-4.568
	(0.277)	(0.288)	(2.127)	(2.878)	(3.560)
Observations	433	433	433	433	433
Adjusted R-squared	0.010	0.008	0.154	0.161	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.

Our results show that teams led by females perform better on average. However, it is also interesting ti understand who drives the effect: is the whole team going better? Or, is the leader himself improving his performance? Or, vice versa, are only team members who benefit from the leader's effort? Also, are these effects different based on student's gender?

We answer these questions in the estimates reported in Table 5 where we investigate if being leader affects student's own performance and whether this effect is differentiated for males and females. In column (1) we only add among controls a dummy variable for students who were leaders of their team. We find no statistically significant effect. However, when we also include among regressors the interaction term between the dummy *Lead*er and the dummy *Female*, to see whether being a leader has a differentiated effect on males

and females, we find a negative and weakly statistically significant effect suggesting that female leaders increase the performance of their team members at the cost of a slightly worse individual contribution to team performance. For instance, in estimates reported in column (5), in which we include among regressors the full set of controls, we find that the performance of a member (not leader) increases by 0.804 if the team is led by a female. We also find that male leaders tend to perform better (+0.604) than male members, even if the effect is imprecisely estimated. In contrast, female leaders tend to perform slightly worse than their teammates: -1.063+0.804=-0.259 and worse (in terms of individual performance at the team part of the exam) also compared to male leaders: +0.804-0.357-1.063=-0.616. However, these differences are not statistically significant.

All in all, our results suggest that women appointed as leaders devote altruistically energies to improve the performance of the whole team, sacrificing their own performance.

Table 5. The impact of female leadership on team performance. Dependent variable: Grade Team Part. OLS Estimates.

	(1)	(2)	(3)	(4)	(5)
Female Led Team	0.760**	0.882**	0.984**	0.989**	0.804**
	(0.371)	(0.435)	(0.412)	(0.402)	(0.372)
Leader	0.021	0.418	0.642	0.607	0.604
	(0.351)	(0.553)	(0.503)	(0.492)	(0.492)
Female		0.436	-0.203	-0.301	-0.357
		(0.412)	(0.394)	(0.404)	(0.374)
Leader*Female		-0.804	-1.082	-1.050	-1.063
		(0.759)	(0.685)	(0.674)	(0.672)
# members present			-0.262	-0.223	-0.082
			(0.301)	(0.306)	(0.281)
Age			-0.101	-0.112	-0.051
			(0.068)	(0.071)	(0.095)
High School Grade			0.128***	0.121***	0.128***
			(0.015)	(0.014)	(0.015)
Lyceum			0.635**	0.748**	0.802***
			(0.298)	(0.303)	(0.280)
Parents' Education				0.121*	0.072
				(0.069)	(0.064)
Expected Grade				-0.050	-0.048
				(0.051)	(0.048)
Trust				-0.008*	-0.008*
				(0.005)	(0.004)
Risk Aversion				-0.044	-0.038
				(0.076)	(0.074)
Macro					-1.831***
					(0.423)
Personnel Ec.					1.134***
					(0.432)
Econometrics					-1.055
					(0.707)
Constant	5.635***	5.363***	-2.461	-3.604	-4.355
	(0.306)	(0.365)	(2.118)	(2.904)	(3.145)
Observations	433	433	433	433	433
Adjusted R-squared	0.007	0.006	0.154	0.161	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.

We have also tried to understand whether female leaders produce positive effects also on the individual part of the test. At this aim, we have considered as dependent variable the *Grade Individual Part*. 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 compared to the team part. Estimates are reported in Table 6 (same specifications reported in Table 4). We find that female led teams tend to perform better also in the individual part of the exam but the effect is very imprecisely estimated and becomes very small when adding the full set of control variables.

Table 6. The impact of female leadership on student performance at the individual part. Dependent variable: Grade Individual Part. OLS Estimates.

	(1)	(2)	(3)	(4)	(5)
Female Led Team	0.596*	0.368	0.409	0.384	0.295
	(0.313)	(0.342)	(0.308)	(0.300)	(0.284)
Female		0.636*	-0.117	-0.148	-0.177
		(0.343)	(0.309)	(0.315)	(0.295)
# members present			-0.395	-0.379	-0.340
			(0.256)	(0.257)	(0.246)
Age			-0.007	-0.019	-0.139*
			(0.048)	(0.049)	(0.072)
High School Grade			0.129***	0.123***	0.116***
			(0.012)	(0.012)	(0.013)
Lyceum			0.236	0.293	0.289
			(0.260)	(0.284)	(0.274)
Expected Grade				0.136**	0.118**
				(0.060)	(0.057)
Parents' Education				-0.036	-0.029
				(0.046)	(0.044)
Trust				-0.001	-0.000
				(0.004)	(0.004)
Risk Aversion				-0.011	-0.012
				(0.063)	(0.060)
Macro					0.075
					(0.335)
Personnel Ec.					1.911***
					(0.398)
Econometrics					0.902
					(0.593)
Constant	5.590***	5.399***	-3.912**	-6.088***	-3.180
	(0.238)	(0.251)	(1.675)	(2.203)	(2.454)
Observations	433	433	433	433	433
Adjusted R-squared	0.007	0.014	0.202	0.206	0.260

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.

4. Are male and female leaders evaluated differently?

In this section, we use data from the online survey we proposed to students after they took the intermediate exam to analyze how they evaluate leader's activity and team effectiveness. All students joining the experiment, including those who did not show up at the intermediate test, were invited to fill out the survey.

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?".

The average rates for *Leader Effectiveness* and *Leader Effort* were respectively of 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 (not leader) students who took the intermediate test. These two measures of leader's performance are strongly correlated (0.86, *p*-value=0.00) suggesting that students are providing faithful answers to our questions even though they were not given any incentive.[8]

Table 7, which reports the results from a specification in which we only consider among regressors the dummy *Female Led Team* with no additional controls, shows that leader's gender does not significantly affect team members' evaluation of leader effectiveness (columns 1 and 2) and effort (columns 3 and 4). This holds both when we consider the whole sample of students joining the experiment who answer to the survey (columns 1 and 3) and when we restrict the sample to students who undertook the intermediate test (columns 2 and 4). Results do not qualitatively change when we control for students' characteristics and course dummies (results not reported).

Then, in spite of the better performance of female led teams, female leaders do not obtain better evaluations from team members. We obtain the same result also when considering students' answers to a question asking a general evaluation of team effectiveness (we asked the following question "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. Ordered probit estimates show that the leader's gender has no effect on how team members evaluate the effectiveness of their team (results not reported and available upon request).

[8] To better 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 surveys seriously and giving reliable answers.

Table 7. Team members' evaluations of their leader. OLS Estimates.

	Leader Effectiveness		Leader Effort	
	Joining the experiment	Taking the intermediate test	Joining the experiment	Taking the intermediate test
	(1)	(2)	(3)	(4)
Female Led Team	0.244	0.400	0.295	0.393
	(0.495)	(0.516)	(0.504)	(0.529)
Constant	6.171***	6.181***	6.089***	6.110***
	(0.346)	(0.361)	(0.354)	(0.375)
Observations	305	268	305	268

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.

The questions asked in the post experiment online survey allow us to understand also whether the time spent together by team members has been affected by the leader's gender. First of all, we consider the question about the number of hours spent in studying activities with team members each week (*Hours Together).* As mentioned in Section 2, students report a weekly average number of team working hours of 3.87, which rises at 4 for students who show up at the intermediate test. Table 8 reports OLS estimates considering as dependent variable *Hours Together* to investigate whether the time spent with team members has been influenced by the leader's gender. As shown in columns (1) to (3), teams led by a female leader have spent significantly more time working together. In the first two columns, where we consider the whole sample of students joining the experiment (with the exclusion of leaders), we find that female led teams have worked together on average about 53 - 60 minutes more each week compared to male led teams. Results remain qualitatively unchanged when we restrict the sample to students who have undertaken the intermediate test (column 3).

The result that female led teams have spent more time working together compared to male led teams finds support also when looking at the probability that the team leader has submitted to the course's professor the homework assigned during teaching classes. Students were strongly encouraged to work in team and to solve the exercises assigned to them as homework. However, no specific incentive (or penalty) was announced for teams doing (not doing) the task. The percentage of students who have done their homework was 81.4% and 84.1% for students joining the experiment and students showing up at the intermediate test, respectively. To investigate the effect of female leaders on the probability of doing the homework we have estimated a Probit model considering as dependent variable the dummy variable *Done Homework*, which takes the value of one for teams who have submitted their homework to the course's professor and zero otherwise. In columns (4) and (5) we report estimates for the whole sample of students joining the experiment. As shown in column (4), in which we only include among regressors *Female Led Team*, teams led by a female have a higher probability of doing their homework. The same result holds true also when we add among regressors the full set of controls (column 5). Again, qualitatively identical results are found when looking at the subsample of students undertaking the intermediate test (column 6).

Table 8. Time spent together with teammates. OLS and Probit Estimates (marginal effects)

	Hours Together			Done Homework		
	Joining the experiment		Taking the intermediate test	Joining the experiment		Taking the intermediate test
	(1)	(2)	(3)	(4)	(5)	(6)
Female Led Team	0.891*	1.010**	0.963*	0.170***	0.140***	0.141***
	(0.470)	(0.487)	(0.545)	(0.055)	(0.047)	(0.042)
Constant	3.408***	1.876	4.395			
	(0.268)	(3.481)	(3.913)			
CONTROLS	NO	YES	YES	NO	YES	YES
Observations	305	305	268	538	538	433
Adjusted R-squared	0.013	0.029	0.027	0.053	0.288	0.344

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. Columns (4) to (6) report average marginal effects of the Probit estimates.

Using questions asked in the final survey, we also enquire 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 (88%) 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 (83%).[9] To investigate the effect of leader's gender we have estimated a Probit model for the dependent variables *Leisure Together* and *Individual Part Alone*.[10] 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 same results hold true also when we restrict the sample to students who have undertaken the intermediate test (estimates not reported and available upon request).

Table 9. Time spent together with teammates. Probit Estimates (marginal effects)

	Leisure Together		Individual Part Alone	
	(1)	(2)	(3)	(4)
Female Led Team	-0.040	-0.047	0.008	0.006
	(0.047)	(0.044)	(0.037)	(0.034)
Controls	NO	YES	NO	YES
Observations	305	305	305	305
Adjusted R-squared	0.003	0.078	0.000	0.137

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.

These results suggest that female leaders have dedicated more effort than males in calling team members' meetings and making the team members study together. The additional studying activity has

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

[10] Here we consider only students who are not leaders. Results do not change if we consider also leaders.

however focused only on the team part and on the homework because the probability of studying for the individual part together does not differ according to the leader's gender. Also, the more intense interaction among team members did not spread beyond the assigned team task as social interaction in leisure time is not significantly different.

5. Concluding Remarks

Female leadership in economic and social contexts is a rather rare phenomenon. Could the cause be that women have lower abilities or attitudes to lead a team, an organization, a firm?

We have ran a field experiment with students from an Italian University to investigate this issue. We have involved about 430 students, attending four academic courses, that took a part of an intermediate exam working in team. At the beginning of the courses, we have randomly assigned students to teams of three members and, within each team, we randomly selected a leader to organize team activities. Through this procedure, we have teams – with different gender composition – led by females and teams led by males.

We preliminarily show some evidence that students effectively worked together for the team part, from the answers they gave to a post-experiment survey and, especially, from the positive high correlation emerging from the performance of a member with the performance of the other members of the team.

Using the results of the intermediate exam, we have evaluated if the academic performance of students in the team part of the exam is affected by the gender of the leader, controlling for a number of individual characteristics. We first show that female led teams perform significantly better than male led teams (an effect of about 0.20 SD of the dependent variable). Then, we find that this effect is mainly driven by the better performance of team members in female led teams, while the performance of the female leaders does not differ from the performance of male leaders. This suggests that female leaders altruistically devote more energies to organize team activities, rather than improving their own performance.

We also find a weak positive effect of female led teams on the students' performance in the individual part, implying that there were some spillovers from studying activities from the team part to the individual part.

In the second part of the paper, using the results from a post-experiment survey among participating students, we analyze if female leaders are evaluated differently from male leaders. First, we find that female led teams tend to work more hours together and to do their homework with higher probability. However, we find that – notwithstanding a better performance – members of female led teams do not evaluate better their leaders. This suggests a form of discrimination against female leaders that are evaluated with a harsh yardstick with respect to their male counterparts.

References

Blume, A., & 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.

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

Brandts, J., & 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.

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

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

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

Kriss, P. H., & 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.

Reuben, E., & 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., & Komai, M. (2015). An experimental study of leadership institutions in collective action games. *Journal of the Economic Science Association*, *1*(1), 100-113.

Van Huyck, J. B., Battalio, R. C., & 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., & Knez, M. (2001). The illusion of leadership: Misattribution of cause in coordination games. *Organization science*, *12*(5), 582-598.