Management Science

The Impact of Gender Diversity on the Performance of Business Teams: Evidence from a Field Experiment

Sander Hoogendoorn, Hessel Oosterbeek, Mirjam van Praag,

Please scroll down for article—it is on subsequent pages

The Impact of Gender Diversity on the Performance of Business Teams: Evidence from a Field Experiment

Sander Hoogendoorn

Amsterdam School of Economics, University of Amsterdam, 1018 WB Amsterdam, The Netherlands; and
Tinbergen Institute, 3062 PA Rotterdam, The Netherlands, s.m.hoogendoorn@uva.nl

Hessel Oosterbeek

Amsterdam School of Economics, University of Amsterdam, 1018 WB Amsterdam, The Netherlands,
h.oosterbeek@uva.nl

Mirjam van Praag

Amsterdam School of Economics, University of Amsterdam, 1018 WB Amsterdam, The Netherlands; and
Tinbergen Institute, 3062 PA Rotterdam, The Netherlands, c.m.vanpraag@uva.nl

This paper reports on a field experiment conducted to estimate the impact of the share of women in business teams on their performance. Teams consisting of undergraduate students in business studies start up a venture as part of their curriculum. We manipulated the gender composition of teams and assigned students randomly to teams, conditional on their gender. We find that teams with an equal gender mix perform better than male-dominated teams in terms of sales and profits. We explore various mechanisms suggested in the literature to explain this positive effect of gender diversity on performance (including complementarities, learning, monitoring, and conflicts) but find no support for them.

1. Introduction

This paper reports on a field experiment conducted to investigate the effect of gender diversity on team performance. Many decisions in organizations are nowadays assigned to teams, not to individuals. As Hamilton et al. (2003, p. 465) notes: "During the past 30 years the use of teams has become a mainstay for the organization of work." Examples include government bodies, judges in collegial courts, and company boards, but also most business start-ups are undertaken by teams (Parker 2009). A better understanding of the determinants of the effectiveness of teams has, therefore, become increasingly relevant. One of the potential determinants of the effectiveness of a team is its diversity. Although there are theoretical results about the effect of team diversity on team productivity (Kremer 1993, Prat 2002), the empirical evidence on causal effects is thin (Hamilton et al. 2003). Our study examines one particular dimension of team diversity, that is, gender diversity.

The gender mix of a team may affect its performance through various underlying mechanisms. Hamilton et al. (2003) point to the trade-off between the costs of coordination and communication and the benefits from a potentially more diverse pool of knowledge and skills. The trade-off for these factors depends on the setting. In some contexts, coordination and communication costs are likely to be very high (the army). In other contexts, diversity of knowledge and skills may be very valuable (raising children). Adams and Ferreira (2009) discuss "mutual monitoring" as a mechanism and show that more gender diverse boards are associated with more intense monitoring practices (see also Gul et al. 2011). Dufwenberg and Muren (2006) derive results from a group dictator game played in the laboratory, showing that mixed gender teams are more generous and more egalitarian. Woolley et al. (2010) show that teams with a larger percentage of women perform better because of a higher average level of social sensitivity of the group members. Pelled (1996) and Pelled et al. (1999) point to interpersonal aspects such as friendships and conflicts that vary with teams' gender composition and may affect their performance.

The teams in our field experiment consist of groups of around 12 first-year students who start and run a

business in an entrepreneurship education program in a large college in Amsterdam. The program involves taking responsibility as a group for a small-sized and short-time business, from its setting up (at the beginning of the school year) to its liquidation (at the end of the school year). Students sell stock, elect officers and divide tasks, produce and market products or services, keep records, and conduct shareholders' meetings. Students thus frequently interact, build up relationships, and create routines and processes to achieve their goals. Everything about the venture is real, including tax and social security payments. We measure team performance in terms of sales and profits.

The results of our study may be informative about the effects of gender mix in teams that operate in comparable settings. One such setting, though admittedly not entirely comparable, are corporate boards. Recently, the gender mix of these boards attracted considerable attention because of the current underrepresentation of women (see Catalyst 2010, Woods 2010) and the introduction of quota in some countries, including Norway, the Netherlands, Spain, France, and Iceland. In 2010, women held only 10% of the board seats at the top 300 European companies and just above 15% of board seats at Fortune 500 companies (Catalyst 2010, Woods 2010). A higher share of women in boards is often regarded as desirable. Commonly expressed arguments in the popular press in favor of more gender diversity in corporate boards include: enlargement of the pool from which talent is attracted, complementarities, and better mutual learning (Desvaux et al. 2007).[1] However, there is little empirical evidence supporting that gender diversity leads to better team performance.

Identifying the causal impact of gender diversity on the performance of teams is a challenging task. Studies based on observational data are likely to be plagued by various endogeneity issues, including reversed causality, unobserved heterogeneity, and (self-)selection into and out of teams (Adams et al. 2010, Hermalin and Weisbach 2003). We illustrate these issues by discussing some recent studies that acknowledge these measurement issues and address (part of) them.

Adams and Ferreira (2009) analyze data from U.S. firms and find that the gender diversity of boards has a positive impact on the intensity of their mutual monitoring and the performance of firms that have otherwise weak (external) governance. However, more gender diverse boards are harmful for the performance of firms with strong (external) governance, possibly due to overmonitoring. On average, the effect of gender diversity on the firm performance indicators Tobin's q and *RoA* (return on assets) turns out to be negative. To address the possible endogeneity of gender diversity, they use the fraction of male board members with connections to female directors in other board positions as instrumental variable and include firm fixed effects in their regressions. Adams and Ferreira (2009) critically argue that these choices do not solve the issues of endogeneity and reverse causality completely: Firm fixed effects only control for unobservables that are fixed over time, and the instrument used is not so strong and may cause weak instrument problems. Moreover, it is questionable that the instrument passes the exclusion restriction required for an instrumental variable to be valid.

Ahern and Dittmar (2012), Matsa and Miller (2010), and Nygaard (2011) all measure the effect of board composition on firm performance and/or governance by exploiting that publicly listed firms in Norway were forced to have at least 40% female directors by 2008. In 2006, when this law was implemented, only 9% of directors were women. Firms thus had to replace on average 30% of their board members. In a difference-in-differences framework, Ahern and Dittmar (2012) compare before–after differences between early compliers and late compliers. Matsa and Miller (2010) compare listed and unlisted companies and companies in Norway and in other Scandinavian countries in a double and triple differences framework. Nygaard (2011) measures the stock market reaction to the unanticipated announcement of the quota as an indicator of the expected impact of an increase in the percentage of female directors on firm value and conditions on firm-specific information asymmetry. The three studies draw opposing conclusions. Ahern and Dittmar (2012) conclude that the reform had a significantly negative impact on firm value because the newly added board members where younger and less experienced. Matsa and Miller (2010) are unable to distinguish between positive and negative effects on long-term profits. Nygaard (2011) finds a significantly positive effect of the announcement of the law on the cumulative abnormal returns (CAR) for firms with low information asymmetry, whereas firms with high information asymmetry experience negative but insignificant CAR.[2]

---

[1] In a laboratory experiment, Niederle and Vesterlund (2007) find that men often choose to compete even if they perform poorly, whereas women often choose not to compete even if they perform well. If reaching a corporate board seat requires one to be competitive, these findings imply that the underrepresentation of women indeed reduces the pool from which talent can be attracted. Adams and Funk (2012) show that female board directors who break the glass ceiling are indeed a specific subset of females, in some respects more similar to males than the general population.

[2] Nygaard (2011) attributes the differences between his findings and those of Ahern and Dittmar (2012) to oversampling of new firms in the latter study.

Finally, Apesteguia et al. (2012) analyze data from the 2007–2009 editions of an online business game for students to study the effect of gender diversity on team performance. Almost 38,000 students in 16,000 teams participated. Incentives are strong: teams can win substantial prizes with relatively high probabilities, and there is the possibility of being hired by the company that runs this business game. The results show that teams of only women perform worse than mixed teams or teams of only men. There is, however, no attempt (or mentioning) in this study to correct for the endogeneity of team formation. Teams that sign up for the game form themselves.[3]

Besides the studies mentioned, quite a number of field studies measure the relationship between teams' gender diversity and performance (e.g., Chowdhury 2005, Ellison et al. 2010, Farrell and Hersch 2005, Herring 2009, Horwitz and Horwitz 2007, Lee and Farh 2004, Pelled 1996, Pelled et al. 1999, Richard et al. 2004, Wegge et al. 2008). However, these contributions study correlations rather than effects and do not aim at overcoming endogeneity or selectivity issues.[4]

The effect of gender diversity in teams has also been studied in laboratory experiments (e.g., Dufwenberg and Muren 2006, Ivanova-Stenzel and Kuebler 2011, Pearsall et al. 2008). Lab experiments do not suffer from endogenous team composition, but their resemblance to real-world situations may be limited. Moreover, they typically measure short-term effects, whereas the consequences of a team's diversity in terms of, for example, coordination, communication, complementarities, and learning, are not likely to become evident instantaneously. It is thus useful to study the effects of team composition in the longer run and preferably in more realistic circumstances.

[3] Also related is the study by Hansen et al. (2006), who measured the impact of gender diversity in student groups on their grade for a group assignment that forms part of an undergraduate introductory management course. Male-dominated groups performed worse on a group-based performance measure than diverse groups and female-dominated groups. Performance in this study is academic achievement rather than business outcomes. Other studies looking at peer effects in education include Hoxby (2000), Lavy and Schlosser (2011), and Oosterbeek and Van Ewijk (2010).

[4] These studies that come from various literatures such as economics and finance (Ellison et al. 2010, Farrell and Hersch 2005); organization and management (Horwitz and Horwitz 2007, Pelled 1996, Pelled et al. 1999, Richard et al. 2004); entrepreneurship (Amason et al. 2006, Chowdhury 2005); or psychology (Lee and Farh 2004, Wegge et al. 2008) consider various levels at which diversity and performance are measured, that is, entire organizations, teams in the workplace, start-up, or top management teams (boards). Numerous other field studies document the correlation between team diversity along various other dimensions than gender and team performance (e.g., Amason et al. 2006, Ancona and Caldwell 1992, Bell 2007, Edwards et al. 2006, Ensley and Hmieleski 2005, Hamilton et al. 2012, LePine 2003).

Our study is the first to conduct a field experiment with random assignment to circumvent endogeneity of teams' gender composition. The field experiment was conducted in the context of the compulsory entrepreneurship program of undergraduate students in international business of the Amsterdam College of Applied Sciences. In teams of around 12, students start up, sell stock, and run a real company with the objective of maximizing profit and shareholders value. Students face strong incentives, both individually and as a team, to perform this substantial and truly joint task. The real-life situation reflected in the experiment shares some features of corporate boards. The team size is comparable and so are the tasks. There are also noticeable differences between the business teams in our study and corporate boards. The companies are new and they exist for just one year. Moreover, all team members are young and inexperienced. Female and male team members are of comparable quality in terms of education and experience (unlike in the Norwegian case).

Forty-three student companies are included in the experiment, with the majority of the observations with a share of women between 0.2 and 0.6. On this segment we find that teams with an equal gender mix perform better than male-dominated teams in terms of sales and profits. We do not have enough female-dominated teams to conclude that these results are symmetric. However, both the univariate and the regression analyses suggest that female-dominated teams do not perform better than gender diverse teams. We inquire various mechanisms suggested in the literature (including complementarities, learning, monitoring, and conflicts) but fail to find support for any.

The remainder of this paper is structured as follows. Section 2 gives more details of the context and design of the field experiment. Section 3 describes the data and reports results from randomization checks. Section 4 presents and discusses the empirical findings. Section 5 concludes.

2. Context and Design

2.1. Context

The program that we study is organized by the college in collaboration with the Junior Achievement Young Enterprise Start-Up Program, which is the leading entrepreneurship education program in (post-)secondary education in the United States and in Europe (see Oosterbeek et al. 2010). The program involves setting up and running a usually rather small-sized company for the duration of the program that lasts one entire academic year. Each team of students is responsible for raising capital by selling stock; assigning and dividing tasks among team members

by electing officers; producing, marketing, and selling products or services; keeping records; and conducting shareholders' meetings. Students thus interact intensively, build up relationships, and create work routines and processes to achieve their goals. The venture is real, including tax and social security payments. The program is thus not a business simulation.

Each student company is supported by a coach from the business world who shares their experience with the students. During the program the teams have to report to their professor and business coach on a regular basis. Business coaches typically oversee only one team. Each professor supervises two teams. Professors and business coaches are randomly matched to teams; none of them have a choice.[5]

Ventures generally proceed as follows. After an interim CEO is appointed, the team starts brainstorming about potential products or services. The teams have to choose the product or service themselves. Market research is then conducted to further analyze the business ideas that survived this process. Next, the core business activity is defined (the college puts no restrictions on the type of business activity). Table 1 lists the resulting products or services sold by all 45 teams and reports some team characteristics including their assigned gender mix.

Next, positions are defined and allocated to team members. The team elects approximately half of its members in specific management positions including the CEO and CFO. The other half of the team works for the firm in nonmanagement positions. After half a (program) year, roles are switched and the management team positions are reallocated among the nonmanaging part of the team. The relationship between the gender composition in the total team and the division of team roles across genders is discussed in §2.2.

Once the corporate plan has been finalized, students start to raise capital for their business operations by issuing shares. All shares have an initial value of 20 euros. Other sources of funding such as personal or outside loans were not allowed. Teams can start their business operations if a majority of shareholders approves the corporate plan at the first shareholders' meeting. Production and marketing of products or services then becomes the main activity of teams. At the end of the year, all ventures are liquidated and each team issues an annual report that needs approval at the final shareholders' meeting. Any profits will be proportionally divided among the shareholders.

The entrepreneurship program at the Department of International Business Studies of the Amsterdam College of Applied Sciences is compulsory, it lasts for an entire academic year and covers about one-fifth of the first-year curriculum. The Department of International Business Studies is divided into five fields of study: management, business management, financial management, trade management Asia, and business languages. The experiment reported in this paper was conducted in the academic year 2008–2009. The total number of students in that year was 550. Graduates in international business studies usually find jobs that lead to managerial positions in the corporate sector, although some will start up their own firms or work for the government sector (and around 30% will first pursue a master's degree).

2.2. Design

One week before the start of the program, we received the names of the students together with their gender and field of study. Within fields of study, we determined and varied the fractions of female students for each team and assigned male and female students randomly to these teams. We assigned 550 students to 45 teams given the predefined restriction that single-sex teams or teams with only one person of a specific sex were not appreciated by the college. We then communicated this assignment to the coordinators of the five fields of study who enforced its implementation. Students were informed about the team they belonged to. A few late applicants were randomly distributed among the existing teams and a few "no shows" were also randomly distributed across teams (as they did not know to which team they were assigned to at that stage).

The last two columns of Table 2 show the range of the share of women, overall and conditional on field of study. This reveals that there are no teams with a share of women above 0.58 outside the field of business languages, and that there are no teams with a share of women below 0.17 outside the field of trade management Asia. This means that only the range from 0.17 to 0.58 is covered by more than one field of study. Of the 11 teams from the field of business languages, only 3 have a share of women below 0.58. Hence, the relation between share of women and performance is poorly identified in the range above 0.58.

The field of study coordinators were informed about the character of the exogenous variation we imposed. We urged them not to inform professors or students. Professors only knew that a research project was conducted that required to stick to the imposed team assignment. Students were told that their program was evaluated and that they were not allowed to switch teams. Only six students switched teams during the year.

We think that it is unlikely that students noticed that there was randomization based on gender.

---

[5] The gender and prior performance of the professors and coaches are unknown to us.

Table 1   Teams

No.	Name	Female (share)	Team size	Sales (euro)	Profits (euro)	Description of product/service
1	A-Card	0.25	16	1,236.15	−848.05	Discount card Amsterdam nightlife
2	A'dam Gadgets	0.36	12	534.12	−41.40	USB hot plate for coffee, tea, etc.
3	Appie	0.90	11	454.75	149.86	Apple-shaped box to preserve apples
4	Aqua de Cocktail	0.42	12	1,130.47	−305.94	Comprehensive cocktail shaker set
5	ArtEco Bags	0.40	11	912.00	−401.69	Durable giveaway bag clothes stores
6	BubbleMania	0.18	11	503.00	−61.79	Multipurpose protective key chain
7	D'Wine	0.25	9	740.00	−55.00	Bottles of wine
8	Eastern Green	0.36	14	513.00	105.51	Engravable text bean growing a plant
9	Escapade Inc	0.67	9	592.55	−111.30	Tube clip for sealing food, toiletry, etc.
10	eyeBMA	0.38	16	557.50	124.66	Package with easy-to-use eye shadow
11	Firefly	0.50	12	2,225.65	293.62	Ascending fire lantern for celebrations
12	Fl!pthat	0.23	13	455.00	214.88	Redecorating already existing websites
13	Ginger	0.58	12	976.50	−106.81	Multipurpose solar energy charger
14	Himitsu	0.30	10	775.00	36.00	n/a
15	I-Care	0.38	15	1,204.45	477.15	Dead Sea minerals beauty products
16	iJoy	0.36	14	1,952.85	93.56	Wristband with USB capacity
17	I-Juice	0.38	13	1,255.38	−38.54	Pocket-size lightweight mobile charger
18	IMSC	0.27	11	625.00	−390.00	n/a
19	iShield	0.44	11	4,209.49	129.76	Invisible protective shield for iPhones
20	KISBag	1.00	9	205.48	−117.02	Tiny foldable bag replacing plastic bags
21	Laservibes	0.36	11	130.00	−228.90	Organizing lasershows for companies
22	Mengelmoes	0.33	10	941.50	63.14	Easy-to-wear telephone charger device
23	My-Buddy	0.17	12	297.00	−58.33	USB doll for kids reflecting emoticons
24	Nine2Five	0.73	12	235.45	−1,016.36	USB hot plate for coffee, tea, etc.
25	Picture Perfect	0.21	15	260.09	−50.87	Customized shirts for men and women
26	Pietje Plu	0.73	12	n/a	n/a	Trendy umbrellas
27	Pocket Memory	0.38	16	978.94	103.46	Business cards with USB capacity
28	Pro'Lux	0.31	14	378.25	−394.90	Promotional gifts with USB capacity
29	Qwinlok	0.31	13	340.00	34.61	Boxer shorts for female adolescents
30	Reflection	0.82	11	889.51	45.43	Cosmetics mirror including mascara clip
31	SAME	0.82	11	1,618.35	152.37	Comfortable unisex earwarmer
32	Sappho	0.50	8	980.00	n/a	n/a
33	Sharity	0.58	12	265.00	−241.12	Peace sign necklace for teenagers
34	ShoeTattoo	0.62	13	270.00	88.32	Shoe customization by graphic artists
35	Student Promotion	0.42	13	571.32	234.54	Promotional activities for companies
36	StuPill	0.38	14	731.33	−1,011.33	Convenient Indonesian anti-RSI pillow
37	Test-a-Holic	0.45	11	728.45	219.77	Alcohol breath tester for nightlife
38	We-Do Solutions	0.10	10	604.00	−266.82	Multipurpose trendy key chain
39	We 'R U	0.33	13	1,041.11	49.77	Compact wallet in several colors
40	XNG	0.50	12	1,087.50	258.31	Shirts "Chicks on Kicks" community
41	YEN Empowered	0.50	13	1,266.67	33.33	n/a
42	YET's Wear	0.53	16	789.08	−246.81	Customized shirts own YET-brand
43	YOU	0.17	12	0.00	−242.41	Hotel door hanger to store keys, etc.
44	Young Legends	0.44	9	400.00	59.00	n/a
45	YUVA	0.53	16	1,153.00	294.11	Engravable rice grain in glass covering

Notes. Share of women is based on a dummy indicator for male and female students (excluding students whose gender is unknown). The number of students whose gender is unknown amounts to 20 (out of 550 students). Team size reflects the size of teams at baseline.

We kept the apparent differences with the team assignment procedures in previous years to a minimum. Like in previous years, team composition is not self-selected by the students but enforced (by the college). Another commonality is that student teams consist of students from the same field of study. The drawback of keeping things as normal as possible is that the range in which we could manipulate the share of women per team was restricted. Based on interviews with students, we are indeed confident about their ignorance regarding the forced and varying shares of females per team.

During the year, 104 students (19%) dropped out. High dropout rates from the first year of tertiary education are common in the Netherlands, where admission based on grades or previous achievement is not allowed. This reduced the average team size from 12 at the start of the program to 10 at the end. Dropouts hardly changed the overall share of females; from 0.44 at baseline to 0.46 at the end of

Table 2   Numbers of Students and Teams, and Share of Women by Field of Study

Study field	Students	Teams	Share of women (per team)		
			Average	Minimum	Maximum
Business management	240	18	0.37	0.18	0.53
Management	60	5	0.29	0.17	0.40
Trade management Asia	105	9	0.35	0.10	0.58
Business languages	118	11	0.71	0.44	1.00
Financial management	27	2	0.40	0.38	0.42
Total	550	45	0.44	0.10	1.00

the year. Dropouts could still contaminate the design of the experiment if (i) a team's gender composition is affected or (ii) dropout rates vary across teams in relation to their gender composition. Neither is the case. The regression coefficient between the teams' share of females at baseline and at the end of the program is 1.091 (SE 0.091). Regressing students' dropout status on the share of women at baseline returns an insignificant coefficient of 0.006 (SE 0.121), indicating that the dropout decision of students is not affected by the gender composition of their team.[6] Analyses of subgroups of dropouts, such as male versus female dropouts and high-ability versus low-ability dropouts, yield similar insignificant results.

Each team was supported by a randomly assigned professor and a business coach. Our estimates could be biased if these coaches and professors would systematically treat teams with different gender compositions differently. The data show no evidence of this being the case: gender composition and the average satisfaction of team members with their coach are not correlated.

In our discussion of the results, we regard the gender mix of the entire team as the treatment variable. The reason is that descriptive statistics confirm our impression that students perceive their teams as a unit and do not distinguish strongly between the managing and working part of the team. Neither the development of entrepreneurial skills nor the development of knowledge that is relevant for entrepreneurship are different across workers and managers. We notice, however, that our results can also be interpreted as the effect of the gender mix of management teams. Each team had two different management teams: one for the first and another for the second half of the program. In principle, a mixed gender team could have had two homogeneous management teams: a female team for one half and a male team for the other half. This did, however, not occur. Regressing the share of women in the management team on the

share of women in the team returns a coefficient of 0.851 (SE 0.135) for the first half of the program and 0.860 (SE 0.109) for the second half of the program. *F*-tests acknowledge that these coefficients are not significantly different from one (*p*-value of 0.28 for the first half of the program and 0.21 for the second half of the program). Moreover, females are not significantly more or less likely to be selected in managing roles than males, in none of the semesters, neither in general nor for the specific roles of CEO and CFO.

2.3. Incentives
In this subsection, we discuss which incentives team members have to care about the business performance of their team. Incentives are strong, both individually and at the team level. As for individual incentives, students can be dismissed by team members. The decision to dismiss someone requires that two-thirds of the team agrees, together with the consent of the professor. In case of being dismissed, the student is excluded from the rest of the program and loses the 12 credit points related to the program (out of 60 credit points in the first year). Being dismissed may endanger the student's prospect of completing the bachelor program for which a minimum of 45 credit points obtained in the first year is mandatory. Dismissal of team members is not uncommon and thus a credible threat: half of the teams have experienced at least one layoff, and the average number is 0.73. The occurrence and number of layoffs are neither related to the team's business results nor to their gender composition. Analyses of subgroups of dismissed students, such as male versus female and high-ability versus low-ability, yield the same insignificant results.

Another incentive with an individual component is the grade students obtain for this program from their professor. The grade has a substantial weight in the student's grade point average (20%). Both individual and team performance determine the grade, and their weight in the total program grade is about 50/50. An indicator of the effect of individual performance is the substantial average difference between the highest and the lowest grade within a team of 1.3 (on a scale from 1 to 10). The relevance of team performance for the individual grade is indicated by the positive correlation between the grade average in the team and the team's sales and profit numbers.

The third individual incentive comes from the fact that most of the students own shares in their companies, whereas the remainder of the shareholders are often family members, friends or acquaintances. We have no exact information on the identity of the companies' shareholders. On average, half of the shares are bought by the team members themselves (approximately 50 euros per student).

Specific team incentives are provided by the formal competition among teams. At the end of the year,

---

[6] Using a quadratic specification gives coefficients of −0.303 (SE 0.352) and 0.320 (SE 0.322) for the linear and the squared term, respectively. In all these regressions, controls for field of study are included.

six selected teams present their results (along with a "business pitch") to a jury of entrepreneurs who select a winner based on the teams' business outcomes and presentations. The winning team obtains a cup, a small prize, and will represent the college in the national Young Enterprise competition. It often also gets some press attention from local and university media.

The effectiveness of these incentives is supported by the reported effort levels in terms of hours. Students spend, on average, 8.1 hours per week (SD 3.8) on the program. This is a high number given that the program counts for 20% of the students' curriculum. The average actual number of hours students in Dutch professional colleges spend on their education is 32 hours per week (Allen et al. 2009). The positive correlation between grades and sales/profits, the ownership of shares by team members, their family and friends, and the criteria used by the judges in the formal competition make it likely that the hours students spend on the project are directed toward the business outcomes we measure.

3. Data

3.1. Variables

In addition to administrative data and teams' annual reports, information was collected through three surveys. At the first day of the first week of the academic year (in September 2008), students filled out a pretreatment survey. Follow-up surveys were administered halfway (in January 2009) and at the end of the program (in May 2009). Response rates are 88% for the baseline survey, 86% for the first follow-up, and 78% for the second follow-up. The surveys provide background information about individuals and teams. We use this information to assess whether team assignment was random (conditional on gender and field of study) and to inquire possible explanations for the effect of teams' gender mix on their business performance.

The baseline survey contains questions about individual characteristics such as age, ethnicity, nationality, education, and parental background. The average age is approximately 19 years and 4 months, roughly two-thirds of the population lives with their parents, about one-third has some work experience, and over 30% have a father who is or was an entrepreneur. Twenty percent of the students are born abroad and about half of the students have at least one parent not born in the Netherlands.[7]

---

[7] We also randomized students to teams on the basis of their ethnicity. Results are reported in Hoogendoorn and van Praag (2012). Because gender diversity and ethnic diversity are orthogonal, this will not affect the results reported here. The correlation is −0.185 and not significantly different from zero.

The baseline survey also included the standard battery of questions to measure the five-factor model of personality structure known as the "big five": agreeableness, conscientiousness, extroversion, neuroticism, and openness to experience (see Goldberg 1990). This commonly used set of measures of personality has been shown to be an explanatory factor of entrepreneurship choices and outcomes (Shane et al. 2010, Zhao and Seibert 2006). Moreover, the baseline survey included statements that are combined through factor analysis into measures of entrepreneurial traits such as need for achievement, need for power, perseverance, risk aversion, self-efficacy, and social orientation. These traits are supposed to be constant over time and possibly influential for entrepreneurship decisions and outcomes (see Parker 2009, Oosterbeek et al. 2010).

Unlike these traits, entrepreneurial skills can be developed over time. Therefore, validated batteries of questions to measure the most relevant skills for entrepreneurship are included in all three surveys. The skills that are measured include analyzing skills, creativity, external orientation, flexibility, market awareness, motivating skills, networking skills, organizing skills, and proactivity (see Parker 2009). These measures are taken using the so-called Escan, a validated self-assessment test based on 114 items. The Escan is widely used in the Netherlands to determine people's entrepreneurial competencies by, for instance, the Dutch Chambers of Commerce and commercial banks. The statements load into factors (with Cronbach alpha's ranging from 0.64 to 0.79) of which the entrepreneurship literature has shown to be the most important traits and skills for successful entrepreneurship. Based on the data collected in Oosterbeek et al. (2010), the Escan has been slightly adapted to increase the validity of items when a population of students rather than entrepreneurs is involved.

Finally, all three surveys include self-assessments of the knowledge that students have in seven areas that are relevant for entrepreneurship, that is, knowledge of business, management, entrepreneurship, strategy, organization, administration, and leadership (see Karlan and Valdivia 2011, Minniti and Bygrave 2001).

To help explain possible differences in team performance, the second follow-up survey contains questions related to teams' procedures and processes. We obtain measures of the teams' atmosphere, conflicts, peer-reviewed individual effort, friendships, layoffs, satisfaction, and the existence of subgroups. Questions related to processes within the team translate into measures of group potency (De Jong et al. 2005); decision making (Oliver and Anderson 1994);

Table 3 Descriptive Statistics of Team Characteristics

	Scale	Mean	SD	Min	Max
Team characteristics					
Age	Years	19.37	0.58	18.29	20.93
Ethnicity	0–1	0.55	0.16	0.20	0.90
Nationality	0–1	0.21	0.11	0.00	0.50
Grade point average	1–10	6.46	0.07	6.33	6.64
Size (at baseline)	Persons	12.22	2.09	8.00	16.00
Conflicts	1–5	2.23	0.59	1.00	3.67
Atmosphere	1–5	3.53	0.55	2.33	4.83
Peer-reviewed efforts	1–10	6.94	0.56	6.14	9.17
Dismissals (dummy = 1 if any)	0/1	0.49	0.50	0.00	1.00
Big five characteristics	Cronbach's α				
Agreeableness	0.75	6.39	0.60	5.13	8.03
Conscientiousness	0.77	3.08	0.66	1.58	5.00
Extroversion	0.81	−1.81	0.67	−2.85	0.99
Neuroticism	0.76	5.48	0.57	4.28	6.50
Openness to experience	0.63	7.35	0.46	6.30	8.21
Team processes					
Group potency	0.87	10.85	1.51	8.29	15.17
Decision making	0.70	1.58	1.24	−0.73	4.60
Mutual monitoring	0.88	9.99	1.02	7.83	12.69
Coordination	0.80	1.83	1.11	−0.31	4.90
Credibility	0.66	2.87	0.61	1.41	4.84
Specialization	0.66	7.88	0.60	5.97	9.66

Table 4 Descriptive Statistics of Outcome Variables, in Euros

		Share of women		
	All	Low $I[\text{share} < 0.4]$	Medium $I[0.4 \geq \text{share} \leq 0.6]$	High $I[\text{share} > 0.6]$
Sales				
Mean	838	698	1,209	609
SD	707	441	1,023	506
Min	0	0	265	205
Max	4,209	1,953	4,209	1,618
Profits				
Mean	−69	−104	17	−116
SD	318	325	249	413
Min	−1,016	−1,011	−402	−1,016
Max	477	477	294	152
N	43	23	13	7

mutual monitoring (Langfred 2004); and coordination, credibility, and specialization (Lewis 2003). Table 3 reports the scales on which these variables are measured and descriptive statistics at the team level. This table shows that there is quite some variation in the scores on these variables across teams. In §4 we examine to what extent these scores are related to teams' gender composition to see whether they can potentially explain our results.

The outcome variables in our analyses are measures of teams' sales and profits. Information about these variables was retrieved from the annual reports that we managed to obtain from 43 out of 45 teams. Sales measures are straightforward and uniformly reported in these reports. Obtaining comparable profit numbers is more challenging and required careful examination of the students' financial statements. For instance, the way the wages of team members, depreciation, and the costs of unsold goods were accounted for was not always uniform. We corrected this as much as possible and are confident that the profit measure is more noisy than the sales measure but not more noisy than profit measures in other data sets.[8]

The first column of Table 4 shows descriptives of sales and profits based on the information from annual reports of all 43 teams. Average sales amount to 838 euros, with a standard deviation of 707 euros.

The worst-performing team has no sales, whereas the best-performing team sells for more than 4,000 euros. Profits are on average negative at −69 euros. The team with the lowest profits loses 1,016 euros, and the highest profits are 477 euros. The correlation between sales and profits equals 0.25 ($p = 0.11$).

The last three columns of Table 4 break the descriptives of the performance measures down by three groups of teams; teams with a low share of women (less than 0.4), teams with an intermediate share of women (between 0.4 and 0.6), and teams with a high share of women (above 0.6). The consistent picture emerging from this breakdown is that on average sales and profits are higher for teams with a balanced gender mix than for teams dominated by one sex. This alludes to the main finding of this paper, which we will present more formally in §4.

3.2. Randomization
Before we get to the main results, we first examine whether students are randomly assigned to teams of different gender composition, conditional on their gender. We test this by regressing—separately for male and female students—students' characteristics on the share of women in their team. Because, in the next section, we allow for nonmonotonic patterns in the relation between team performance and the share of women, we also do that in the randomization checks. We regress the characteristics of students on dummies for the share of women in the team being below 0.4 or above 0.6. The reference category is a balanced gender mix between 0.4 and 0.6.[9] Because randomization is conditional on field of study all regressions include controls for that.

Table 5 reports the results. Each pair of coefficients comes from a separate regression; for example, we

---

[8] Unlike in other studies using samples of firms and their profit numbers, all companies produced their financial report in the same format, and we were allowed to check each entry.

[9] Results are very similar when we employ other cutoffs (0.33 and 0.67 or 0.45 and 0.55).

Table 5 Random Assignment of Male and Female Students at the Individual Level

	Males		Females	
	I[share < 0.4]	*I*[share > 0.6]	*I*[share < 0.4]	*I*[share > 0.6]
Personal characteristics				
Age	0.166	0.783	0.315	0.652**
	(0.330)	(0.497)	(0.348)	(0.315)
Ethnicity	−0.070	−0.330**	0.045	−0.112
	(0.055)	(0.142)	(0.094)	(0.113)
Nationality	−0.026	−0.025	0.016	−0.045
	(0.060)	(0.222)	(0.062)	(0.080)
Grade point average	0.009	0.054	−0.017	0.050
	(0.026)	(0.061)	(0.056)	(0.065)
Big five characteristics				
Agreeableness	0.066	−0.671	0.008	−0.084
	(0.265)	(0.558)	(0.327)	(0.379)
Conscientiousness	0.001	−1.169**	0.224	0.665***
	(0.272)	(0.537)	(0.290)	(0.165)
Extroversion	0.423	0.139	−0.176	−0.308
	(0.335)	(0.432)	(0.270)	(0.341)
Neuroticism	−0.404	−0.034	−0.100	0.669
	(0.274)	(0.589)	(0.225)	(0.449)
Openness to experience	−0.400**	−0.008	0.197	0.112
	(0.169)	(0.392)	(0.178)	(0.381)

Notes. Pairs of coefficients come from a regression at the individual level of the row variable on dummies for share of women below 0.4 and above 0.6, separately for men and women. All regressions include controls for field of study. Robust standard errors in parentheses.

*** and ** denote significance at the 1% and 5% levels, respectively.

regressed age of male students on dummies for share of women below 0.4 and above 0.6 and found coefficients equal to 0.166 (SE 0.330) and 0.783 (SE 0.497). Ideally, none of the coefficients of the table should be significantly different from zero. There are some deviations from this ideal. Men assigned to groups with a low share of women are less open to new experience than men assigned to a team with a balanced gender mix. Likewise, men assigned to groups with a high share of women are more likely to be of Dutch origin and score lower on conscientiousness than men assigned to a team with a balanced gender mix. Finally, women assigned to groups with a high share of women are a bit older and score higher on conscientiousness than women assigned to a team with a balanced gender mix. Note that four of the five significant coefficients pertain to teams with a high share of women. As we previously mentioned, this is the segment where the effect of gender composition on team performance is poorly identified, implying that we cannot draw firm conclusions about these effects within this segment anyway.

The share of women is possibly correlated with other team characteristics. Table 6 reports results from regressions of team characteristics on dummies for low and high share of women. Team characteristics are the mean values of the individual characteristics

analyzed in Table 5. All regressions control for dummies of field of study. None of the team characteristics are significantly related to the share of women. The coefficients of a low share of women are relatively precisely estimated. The coefficients of a high share of women are not precisely measured. This is due to the small number of observations in that category.

Table 6 Regressions of (Average) Team Characteristics at Baseline on Dummies for Share of Women

	I[share < 0.4]	*I*[share > 0.6]
Personal characteristics (average)		
Age	0.233 (0.215)	0.707 (4.060)
Ethnicity	−0.030 (0.068)	−0.201 (1.006)
Nationality	−0.020 (0.039)	−0.111 (0.451)
Grade point average	0.006 (0.031)	0.048 (0.342)
Big five characteristics (average)		
Agreeableness	−0.041 (0.199)	−0.472 (1.336)
Conscientiousness	0.010 (0.238)	0.015 (2.442)
Extroversion	0.280 (0.212)	−0.055 (1.183)
Neuroticism	−0.191 (0.189)	0.356 (3.668)
Openness to experience	−0.026 (0.135)	−0.321 (2.549)
Team size	0.412 (0.721)	0.021 (8.335)

Notes. Pairs of coefficients come from a regression at the team level of the row variable on the column variables. All specifications include controls for field of study. Bootstrapped standard errors in parentheses (1,000 replications).
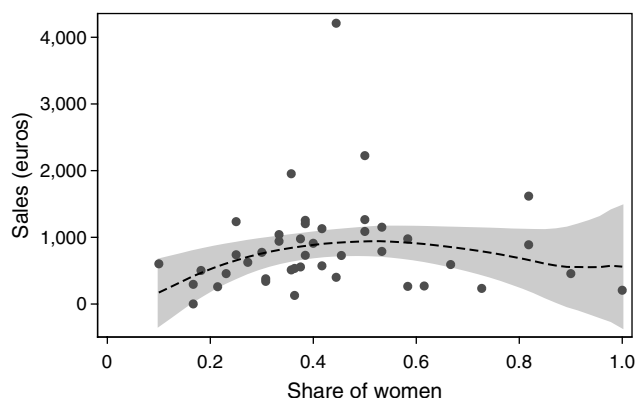
4. Results

4.1. Main Finding

Figures 1 and 2 show the relations between the share of women in a team and teams' sales and profits. The graphs are based on kernel-weighted local polynomial smoothing (details are reported below each graph). The dots represent the actual team results, and the shaded areas represent the 90% confidence intervals.

The relation between sales and share of women is inverse U-shaped. For a share of women between 0.2 and 0.5, sales increase when the share of women increases. When the share of women exceeds 0.5, sales tend to decrease when the share of women increases further. Also, profits are increasing in the share of women when the share of women is below 0.5. For higher shares of women, the relation between profits and the share of women is flat. The dots show clearly that almost all best-performing teams have an equal gender mix, and teams that perform poorly are more spread out across the distribution.

Figure 1   Relation Between Share of Women and Sales

Note. Kernel = Epanechnikov, degree = 1, bandwidth = 0.21, *p*-width = 0.32.

Figure 2   Relation Between Share of Women and Profits
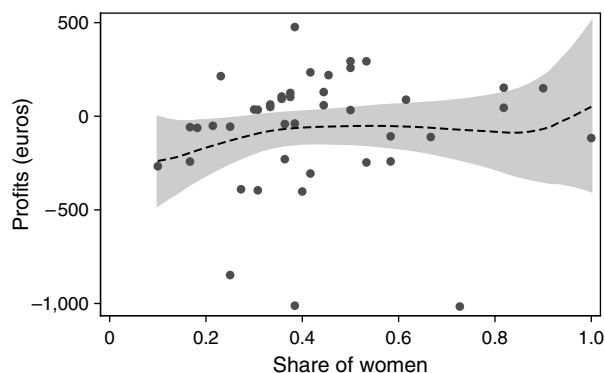
Note. Kernel = Epanechnikov, degree = 1, bandwidth = 0.19, *p*-width = 0.29.

Tables 7 and 8 report results from different regressions of sales and profits, respectively, on the share of women. Motivated by the graphs, we divide the horizontal axis into three segments: a segment with a low share of women, a segment with a balanced gender mix, and a segment with a high share of women. We present results for three pairs of cutoffs between low balanced and balanced high: 0.4 and 0.6, 0.45 and 0.55, 0.49 and 0.51.[10] In panel A, the relation between outcomes and share of women is captured in splines. This allows for different linear relations between the outcomes and the share of women on each segment. In panel B, we collapse the share of women into two dummy variables for a low and a high share of women (where a balanced share is the reference group). Columns (1), (4), and (7) report results from standard ordinary least squares (OLS) regression. Because these results may be sensitive to outliers, we also present results from median regression, in columns (2), (5), and (8), and from robust regression (using *M*-estimation), in columns (3), (6), and (9). Because the number of observations is small, analytical standard errors based on asymptotic theory may understate, and we therefore report bootstrapped standard errors based on 1,000 replications.[11]

The results in the first row of Table 7 show that sales are increasing in the share of women in the first segment. For the second and third segments, none of the estimates are significantly different from zero. Results are not very sensitive to the precise location of the first spline point. Comparing the OLS results for the first segment with the results from median and robust regression shows that the OLS estimates are substantially inflated by outliers; the estimates are more or less cut in half when we move from OLS to the other estimation methods. The result in column (2) of panel A implies that raising the share of women from 0.3 to 0.4 increases sales by 225 euros.

The first row in Table 8 shows that profits are also increasing in the share of women in the first segment. Not all estimates are, however, precise enough to reach statistical significance. The point estimates are very stable across the location of the spline point and the estimation method. Also for profits, none of the splines on the second and third segments are statistically significant.

The results from the dummy regressions in panel B of Tables 7 and 8 are in accordance with the results in panel A of Tables 7 and 8. Teams with a low

---

[10] Measures of goodness of fit of these different models are very similar. The last pair of cutoffs leaves only 3 out of 43 teams in the balanced category. Results are very similar when we choose only one breakpoint at 0.5.

[11] We note that analytical standard errors are indeed substantially smaller than the bootstrapped standard errors reported here.

Table 7   **Effect of Share of Women on Sales; Various Specifications**

	0.4 and 0.6			0.45 and 0.55			0.49 and 0.51		
	OLS (1)	Median (2)	Robust (3)	OLS (4)	Median (5)	Robust (6)	OLS (7)	Median (8)	Robust (9)
					A: Splines				
First segment	3,243.3** (1,310.4)	2,255.0* (1,343.8)	2,116.2** (1,065.4)	3,869.3** (1,829.6)	2,158.5* (1,191.9)	1,908.1** (851.5)	3,831.4** (1,833.1)	2,255.0* (1,161.8)	1,868.0** (911.9)
Second segment	1,203.4 (2,193.6)	−356.5 (2,682.5)	570.7 (5,052.0)	−2,801.0 (4,864.0)	−1,669.7 (4,270.8)	531.5 (4,205.6)	−22,604.2 (26,382.9)	−13,168.6 (20,794.8)	−846.8 (21,157.6)
Third segment	431.4 (1,691.5)	367.4 (2,969.4)	87.7 (2,640.4)	834.3 (1,592.8)	599.2 (2,027.5)	15.5 (2,832.3)	788.9 (1,682.5)	552.3 (2,011.0)	2.1 (4,652.8)
R^2	0.22	0.21	0.33	0.25	0.22	0.35	0.25	0.22	0.35
					B: Dummies				
First segment	−648.2* (347.3)	−287.0 (245.4)	−309.1 (218.4)	−323.7 (316.6)	−491.7 (310.2)	−382.4 (260.3)	−720.8** (357.8)	−507.9 (478.7)	−872.6** (428.6)
Third segment	276.9 (208.7)	189.7 (235.0)	98.8 (211.2)	−165.9 (315.4)	−290.2 (342.9)	−219.8 (289.8)	−602.8* (363.6)	−290.2 (494.5)	−603.5 (437.6)
R^2	0.23	0.19	0.27	0.10	0.17	0.26	0.14	0.18	0.34

Notes. Based on information from 43 teams. All specifications include controls for field of study. Numbers in the first row refer to cutoffs of the share of women used to create splines and dummies. OLS, median, and robust refer to the estimation method. Bootstrapped standard errors in parentheses (1,000 replications).
   ** and * denote significance at the 5% and 10% levels, respectively.

share of women have lower sales and lower profits than teams with a balanced gender mix. We also find that teams with a large share of women make lower profits than teams with a balanced gender mix. This finding is, however, sensitive to the definition of the dummy variables. It matters a lot for the estimates whether some teams with a share of women in the range between 0.55 and 0.6 are assigned to the high or balanced group.

All in all, the results presented in this subsection indicate that teams' sales and profits increase when the share of women increases from a low to an intermediate level. Our estimates lack the precision to draw firm conclusions about the effect on business performance of the share of women in a team when this share exceeds 0.6. However, the results suggest that female-dominated teams do not perform better than gender diverse teams.

4.2. Mechanisms
The studies mentioned in the introduction suggest various mechanisms as possible explanations for our findings. In what follows, we explore the potential of these mechanisms by regressing indicators of these

Table 8   **Effect of Share of Women on Profits; Various Specifications**

	0.4 and 0.6			0.45 and 0.55			0.49 and 0.51		
	OLS (1)	Median (2)	Robust (3)	OLS (4)	Median (5)	Robust (6)	OLS (7)	Median (8)	Robust (9)
					A: Splines				
First segment	750.7 (784.0)	954.2 (654.0)	682.8 (787.1)	840.7 (644.3)	814.2* (447.6)	720.3 (522.3)	845.2 (610.8)	977.8** (424.4)	731.2* (440.9)
Second segment	−292.5 (1,023.8)	−934.2 (1,313.4)	−227.7 (1,135.7)	−1,367.6 (1,815.3)	−2,495.9 (2,262.6)	−1,102.9 (2,401.7)	−8,429.3 (9,222.2)	−18,845.2 (11,476.9)	−7,245.9 (11,019.4)
Third segment	373.4 (1,132.0)	603.9 (1,563.1)	221.2 (2,596.1)	368.3 (978.2)	863.2 (1,318.0)	264.0 (2,300.8)	307.6 (1,070.2)	1,034.4 (1,360.8)	224.4 (2,141.2)
R^2	0.22	0.17	0.25	0.23	0.19	0.26	0.23	0.20	0.26
					B: Dummies				
First segment	−111.9 (92.4)	−80.0 (123.3)	−61.2 (105.4)	−160.7 (110.6)	−223.7* (122.6)	−158.3 (106.4)	−214.0** (88.8)	−223.7* (116.2)	−188.3** (84.1)
Third segment	−24.5 (201.7)	286.5 (201.5)	127.3 (186.7)	−301.3** (140.8)	−365.1** (144.1)	−237.3** (108.0)	−268.2* (161.3)	−365.1** (183.2)	−253.7* (150.9)
R^2	0.21	0.14	0.26	0.23	0.18	0.31	0.22	0.17	0.27

Notes. Based on information from 43 teams. All specifications include controls for field of study. Numbers in the first row refer to cutoffs of the share of women used to create splines and dummies. OLS, median, and robust refer to the estimation method. Bootstrapped standard errors in parentheses (1,000 replications).
   ** and * denote significance at the 5% and 10% levels, respectively.

Table 9   Mechanisms

	Complementary knowledge			Complementary skills			Learning knowledge			Learning skills			Team conflicts			Friendships		
	OLS (1)	Median (2)	Robust (3)	OLS (4)	Median (5)	Robust (6)	OLS (7)	Median (8)	Robust (9)	OLS (10)	Median (11)	Robust (12)	OLS (13)	Median (14)	Robust (15)	OLS (16)	Median (17)	Robust (18)
A: Splines																		
First segment	0.2 (1.0)	0.0 (1.7)	0.2 (1.2)	0.4 (1.1)	-0.3 (1.9)	0.4 (1.4)	-0.8 (0.9)	-1.1 (1.2)	-1.2 (0.9)	-0.0 (1.1)	-2.1 (2.1)	-0.1 (1.7)	-0.5 (1.4)	0.1 (2.1)	-0.3 (2.1)	-0.5 (1.4)	0.7 (1.9)	-0.2 (1.7)
Second segment	1.3 (2.0)	-0.0 (3.0)	1.3 (2.6)	-0.4 (1.8)	-0.3 (2.9)	-0.3 (2.6)	0.7 (1.4)	1.7 (1.6)	1.8 (1.1)	-0.3 (1.9)	0.7 (3.1)	-0.0 (2.8)	1.0 (2.5)	-0.4 (4.3)	0.8 (4.1)	4.6* (2.4)	3.3 (3.3)	4.5 (3.0)
Third segment	0.7 (1.8)	0.0 (2.3)	0.8 (6.5)	-1.9 (3.2)	-0.7 (2.5)	-1.7 (6.5)	-1.7** (0.8)	-1.6 (1.5)	-1.9 (1.4)	0.7 (1.4)	0.7 (3.6)	0.7 (3.3)	-0.3 (4.0)	-1.0 (2.0)	-0.4 (6.5)	-2.5 (1.7)	-4.0 (2.8)	-2.6 (6.8)
B: Dummies																		
First segment	-0.1 (0.2)	-0.0 (0.3)	-0.0 (0.2)	-0.0 (0.1)	0.1 (0.2)	-0.0 (0.2)	0.1 (0.1)	0.0 (0.1)	0.0 (0.1)	0.1 (0.2)	0.1 (0.2)	0.1 (0.2)	-0.1 (0.2)	-0.0 (0.3)	-0.1 (0.3)	-0.2 (0.2)	-0.3 (0.2)	-0.3 (0.2)
Third segment	0.4 (0.3)	0.0 (0.4)	0.4 (0.3)	0.0 (0.3)	-0.1 (0.4)	0.1 (0.4)	-0.2 (0.2)	-0.6* (0.3)	-0.2 (0.3)	0.1 (0.2)	-0.0 (0.3)	0.0 (0.2)	-0.1 (0.2)	0.2 (0.3)	-0.1 (0.3)	0.1 (0.7)	-0.9 (1.0)	0.1 (0.8)

	Decision making			Team atmosphere			Low/high tech	Product/service	Monitoring		
	OLS (19)	Median (20)	Robust (21)	OLS (22)	Median (23)	Robust (24)	OLS (25)	OLS (26)	OLS (27)	Median (28)	Robust (29)
A: Splines											
First segment	2.1 (2.8)	1.4 (4.2)	1.2 (3.6)	1.6 (1.3)	1.0 (1.7)	1.3 (1.7)	1.1 (1.2)	0.3 (0.9)	2.4 (2.5)	2.6 (3.4)	3.2 (3.2)
Second segment	-3.0 (5.9)	-1.7 (8.9)	-1.6 (7.9)	-0.1 (2.3)	2.0 (3.6)	0.9 (3.3)	-2.0 (2.3)	1.0 (1.1)	3.3 (3.6)	4.8 (4.8)	2.7 (4.5)
Third segment	2.9 (4.5)	2.8 (5.9)	3.3 (8.4)	1.2 (1.6)	1.1 (1.8)	1.3 (3.2)	-0.3 (1.5)	1.1 (1.2)	0.2 (3.3)	3.2 (4.3)	0.9 (5.6)
B: Dummies											
First segment	0.4 (0.5)	-0.1 (0.7)	0.3 (0.6)	-0.0 (0.2)	-0.0 (0.3)	-0.1 (0.2)	0.1 (0.2)	-0.0 (0.1)	-0.4 (0.3)	-0.5 (0.5)	-0.4 (0.5)
Third segment	0.4 (0.7)	1.1 (0.7)	0.6 (0.7)	0.2 (0.2)	0.3 (0.3)	0.3 (0.3)	0.1 (0.2)	-0.1 (0.2)	0.8 (0.6)	0.5 (0.7)	0.9 (0.7)

Notes. Based on information from 43 teams. All specifications include controls for field of study. OLS, median, and robust refer to the estimation method. Median and robust specifications for type of product are excluded since this variable is dichotomous. Bootstrapped standard errors in parentheses (1,000 replications).
**Denotes significance at the 5% level.

mechanisms at the team level on the share of women in the team, thereby using the same specifications as before (with cutoffs at 0.4 and 0.6). If the variable that indicates a certain mechanism is unrelated to the share of women, then we can safely conclude that this mechanism cannot explain our findings. If the variable that indicates a certain mechanism is related to the share of women, then this mechanism potentially contributes to the explanation of our results. For the mechanism to actually explain our findings (partially), it also needs to be true that the variable that captures the mechanism has a significant impact on sales and/or profits. Unfortunately, our research design (nor any other we know of) does not allow us to test this. The variable of interest is endogenous and we lack exogenous variation, other than the randomization of the share of women, to identify its causal impact. Including the regressor of interest as an additional control next to the share of women introduces a "bad controls" problem and renders the coefficients uninterpretable (see Angrist and Pischke 2008). Results are reported in Table 9. The remainder of this subsection describes the operationalization of the dependent variables and discusses the results.

Complementarities. Men and women in mixed teams may complement each others' skills and knowledge. We standardize the various skill and knowledge dimensions (see §3) and then compute for each skill and knowledge dimension the maximum in a team. Subsequently, we compute the minimum of the maximums of all skill dimensions and the same for all knowledge dimensions. Supposedly, if men and women complement each others skills or knowledge, these minimums are higher in mixed teams. We find no support for that; see columns (1)–(6) of Table 9.

Learning. When teams learn, mean skill and knowledge levels increase. Learning may be related to the gender composition of a team. This may be due to different initial distributions of skills and knowledge levels or due to differential team processes that may be unobserved. The team average increases in skill/knowledge levels turns out to be unrelated to teams' gender composition; see columns (7)–(12) of Table 9. There is thus not more or less learning in gender diverse teams than in other teams.

Conflicts, friendships, decision making, and atmosphere. The second follow-up survey asked to what extent there was conflict or disagreement between the team members about personal matters (that did not have anything to do with performing the tasks). Examples are social events or gossip. Respondents could give a score on a scale from 1 to 5. The average score of a team on this variable is unrelated to teams' gender composition; see columns (13)–(15) of Table 9. The same holds for agreeableness or social skills.

Moreover, in the second follow-up survey, we asked respondents whether decisions on strategies were mainly taken by a few members of the team or were generally taken by the whole team. Teams' averages of this variable are unrelated to gender diversity; see columns (19)–(21) of Table 9. The second follow-up survey also asked respondents how many team members they see on a friendly basis. We took the average of that number as indicator of friendships in a team. This measure is unrelated to gender diversity; see columns (16)–(18) of Table 9. Finally, the second follow-up survey asked respondents to rate the atmosphere within their team on a 5-points scale. The average within a team is our measure of team atmosphere. The atmosphere within teams turns out to be unrelated to gender diversity; see columns (22)–(24) of Table 9.

Type of product. We have also checked whether the products/services produced by more gender diverse teams target a more diverse and thus larger market. To do this, we have categorized the various companies and their products in various ways. The market orientation of teams is as follows: 40% focuses on the looks of customers, 33% employs high technology, 88% sells a product (rather than a service), and 23% exclusively targets female buyers. No systematic relationships were observed between the market orientation of teams and their gender mix; see columns (25) and (26) of Table 9.

Monitoring. We collected information to measure the level of mutual monitoring in both follow-up surveys. The measure of monitoring is based on four items (see Langfred 2004): (i) we check to make sure that everyone in the team continues to work; (ii) we check whether everybody is meeting their obligations to the team; (iii) we monitor each other's progress on the project; (iv) we watch to make sure that everyone in the team meets their deadlines. The Cronbach's alpha of 0.88 indicates the validity of the factor. We restrict the analyses to the level of monitoring measured in the second follow-up (in May 2009) because students are likely to have a more comprehensive overview of mutual monitoring in their teams at the end of the program. We find no evidence that monitoring varies with the share of women in the team; see columns (27)–(29) of Table 9.

5. Conclusion

The key finding of this study is that business teams with an equal gender mix perform better than male-dominated teams in terms of sales and profits. Our study also suggests that teams with an equal gender mix perform no worse than teams with a majority of females, although the distribution of our data does not allow firm conclusions about the effect of

gender mix on performance for female-dominated teams. This result is based on a field experiment in which students participating in an entrepreneurship program were randomly assigned to teams that each start and run a business during one year. The design of this field experiment combines the strong features (or avoids the weaknesses) of studies based on observational data and laboratory experiments. It shares the high internal validity of laboratory experiments with the realistic setting of data from studies based on observational data. Yet the external validity of the results is uncertain. The teams in our experiment consist of young, inexperienced people who run their business for the limited duration of one year. This may restrict the generalizability of the results.

Some important things remain unknown. Practical matters restricted the range over which we could vary the share of women. Because of this, the results are most informative about the effects of changing from a male-dominated team into a team with a balanced gender mix. The results are less informative about the effects of changing a female-dominated team into a team with a balanced gender mix. Nevertheless, the range of the gender distribution where we have sufficient variation is relevant for feeding the actual discussion of "improving" the gender balance in management teams. Although we collected detailed information about team processes, we were unable to identify any process that could potentially explain why mixed teams perform better than male-dominated teams.

The entrepreneurship program in which we were allowed to manipulate the gender composition of teams is used in many schools in many countries. This gives ample opportunity for follow-up studies. These studies should, in the first place, attempt to replicate results from the current study, preferably on a larger scale. If possible, these studies should also cover a wider range of the share of women, so that we can learn whether effects are symmetric around a balanced gender mix. Finally, with research on a larger scale, it should also be possible to uncover some of the underlying mechanisms.

Acknowledgments

References

Adams R, Ferreira D (2009) Women in the boardroom and their impact on governance and performance. *J. Financial Econom.* 94(2):291–309.

Adams R, Funk P (2012) Beyond the glass ceiling: Does gender matter? *Management Sci.* 58(2):219–235.

Adams R, Hermalin B, Weisbach M (2010) The role of boards of directors in corporate governance: A conceptual framework and survey. *J. Econom. Literature* 48(1):58–107.

Ahern K, Dittmar A (2012) The changing of the boards: The impact on firm valuation of mandated female board representation. *Quart. J. Econom.* 127(1):137–197.

Allen J, van Breugel G, Coenen J, Fouarge D, Meng C, Ramaekers G, van Dijk J, Venhorst V (2009) Afgestudeerden van het HBO tijdens een crisis: Geen verloren generatie. Technical report, HBO-raad, Den Haag, The Netherlands.

Amason A, Shrader R, Tompson G (2006) Newness and novelty: Relating top management team composition to new venture performance. *J. Bus. Venturing* 21(1):125–148.

Ancona D, Caldwell D (1992) Demography and design: Predictors of new product team performance. *Organ. Sci.* 3(3):321–341.

Angrist J, Pischke J (2008) *Mostly Harmless Econometrics: An Empiricist's Companion* (Princeton University Press, Princeton, NJ).

Apesteguia J, Azmat G, Iriberri N (2012) The impact of gender composition on team performance and decision making: Evidence from the field. *Management Sci.* 58(1):78–93.

Bell S (2007) Deep-level composition variables as predictors of team performance: A meta-analysis. *J. Appl. Psych.* 92(3):595–614.

Catalyst (2010) Catalyst census of women board directors of the Fortune 1000. Technical report, Catalyst, New York.

Chowdhury S (2005) Demographic diversity for building an effective entrepreneurial team: Is it important? *J. Bus. Venturing* 20(6):727–746.

De Jong A, de Ruyter K, Wetzels M (2005) Antecedents and consequences of group potency: A study of self-managing service teams. *Management Sci.* 51(11):1610–1625.

Desvaux G, Devillard-Hoellinger S, Baumgarten P (2007) Women matter: Gender diversity, a corporate performance driver. Technical report, McKinsey & Company, Paris.

Dufwenberg M, Muren A (2006) Gender composition in teams. *J. Econom. Behav. Organ.* 61(1):50–54.

Edwards B, Day E, Arthur Jr, W, Bell S (2006) Relationships among team ability composition, team mental models, and team performance. *J. Appl. Psych.* 91(3):727–736.

Ellison S, Greenbaum J, Mullin W (2010) Diversity, social goods provision, and performance in the firm. CESifo. Working Paper 3171, CESifo, Munich, Germany.

Ensley M, Hmieleski K (2005) A comparative study of new venture top management team composition, dynamics and performance between university-based and independent start-ups. *Res. Policy* 34(7):1091–1105.

Farrell K, Hersch P (2005) Additions to corporate boards: The effect of gender. *J. Corporate Finance* 11(2):85–106.

Goldberg L (1990) An alternative description of personality: The big five factor structure. *J. Personality Soc. Psych.* 59(6):1216–1229.

Gul F, Srinidhi B, Ng A (2011) Does board gender diversity improve the informativeness of stock prices? *J. Accounting Econom.* 51(3):314–338.

Hamilton B, Nickerson J, Owan H (2003) Team incentives and worker heterogeneity: An empirical analysis of the impact of teams on productivity and participation. *J. Political Econom.* 111(3):465–497.

Hamilton B, Nickerson J, Owan H (2012) Advances in the economic analyses of participatory and labor-managed firms. Bryson A, ed. *Diversity and Productivity in Production Teams*, Vol. 13. (Emerald Group Publishing Limited, Bingley, UK), 99–138.

Hansen Z, Owan H, Pan J (2006) The impact of group diversity on performance and knowledge spillover: An experiment in a college classroom. Working paper, National Bureau of Economic Research, Cambridge, MA.

Hermalin B, Weisbach M (2003) Boards of directors as an endogenously determined institution: A survey of the economic literature. *Econom. Policy Rev.* 9(1):7–26.

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

Hoogendoorn S, van Praag M (2012) Ethnic diversity and team performance: A field experiment. IZA Discussion Paper 6731, Institute for the Study of Labor, Bonn, Germany.

Horwitz S, Horwitz I (2007) The effects of team diversity on team outcomes: A meta-analytic review of team demography. *J. Management* 33(6):987–1015.

Hoxby C (2000) Peer effects in the classroom: Learning from gender and race variation. Working paper, National Bureau of Economic Research, Cambridge, MA.

Ivanova-Stenzel R, Kuebler D (2011) Gender differences in team work and team competition. *J. Econom. Psych.* 32(5):797–808.

Karlan D, Valdivia M (2011) Teaching entrepreneurship: Impact of business training on microfinance clients and institutions. *Rev. Econom. Statist.* 93(2):510–527.

Kremer M (1993) The O-ring theory of economic development. *Quart. J. Econom.* 108(3):551–575.

Langfred C (2004) Too much of a good thing? Negative effects of high trust and individual autonomy in self-managing teams. *Acad. Management J.* 47(3):385–399.

Lavy V, Schlosser A (2011) Mechanisms and impacts of gender peer effects at school. *Amer. Econom. J.: Appl. Econom.* 3(2):1–33.

Lee C, Farh J (2004) Joint effects of group efficacy and gender diversity on group cohesion and performance. *Appl. Psych.* 53(1):136–154.

LePine J (2003) Team adaptation and postchange performance: Effects of team composition in terms of members' cognitive ability and personality. *J. Appl. Psych.* 88(1):27–39.

Lewis K (2003) Measuring transactive memory systems in the field: Scale development and validation. *J. Appl. Psych.* 88(4):587–604.

Matsa D, Miller A (2010) A female style in corporate leadership? Evidence from quotas. Discussion paper, Institute for the Study of Labor, Bonn, Germany.

Minniti M, Bygrave W (2001) A dynamic model of entrepreneurial learning. *Entrepreneurship Theory and Practice* 25(3):5–16.

Niederle M, Vesterlund L (2007) Do women shy away from competition? Do men compete too much? *Quart. J. Econom.* 122(3):1067–1101.

Nygaard K (2011) Forced board changes: Evidence from Norway. Discussion paper, Norwegian School of Economics, Bergen, Norway.

Oliver R, Anderson E (1994) An empirical test of the consequences of behavior- and outcome-based sales control systems. *J. Marketing* 58(4):53–67.

Oosterbeek H, Van Ewijk R (2010) Gender peer effects in university: Evidence from a randomized experiment. Discussion paper, Tinbergen Institute, Amsterdam.

Oosterbeek H, Van Praag M, IJsselstein A (2010) The impact of entrepreneurship education on entrepreneurship skills and motivation. *Eur. Econom. Rev.* 54(3):442–454.

Parker S (2009) *The Economics of Entrepreneurship* (Cambridge University Press, Cambridge, UK).

Pearsall M, Ellis A, Evans J (2008) Unlocking the effects of gender faultlines on team creativity: Is activation the key? *J. Appl. Psych.* 93(1):225–234.

Pelled L (1996) Demographic diversity, conflict, and work group outcomes: An intervening process theory. *Organ. Sci.* 7(6):615–631.

Pelled L, Eisenhardt K, Xin K (1999) Exploring the black box: An analysis of work group diversity, conflict, and performance. *Admin. Sci. Quart.* 44(1):1–3.

Prat A (2002) Should a team be homogeneous? *Eur. Econom. Rev.* 46(7):1187–1207.

Richard O, Barnett T, Dwyer S, Chadwick K (2004) Cultural diversity in management, firm performance, and the moderating role of entrepreneurial orientation dimensions. *Acad. Management J.* 47(2):255–266.

Shane S, Nicolaou N, Cherkas L, Spector D (2010) Genetics, the big five, and the tendency to be self-employed. *J. Appl. Psych.* 95(6):1154–1162.

Wegge J, Roth C, Neubach B, Schmidt K, Kanfer R (2008) Age and gender diversity as determinants of performance and health in a public organization: The role of task complexity and group size. *J. Appl. Psych.* 93(6):1301–1313.

Woods D (2010) Female representation on UK boards has risen to 13.6%. *HR Magazine* (October 5), http://www.hrmagazine.co.uk/hro/news/1018516/female-representation-uk-boards-risen-136.

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

Zhao H, Seibert S (2006) The big five personality dimensions and entrepreneurial status: A meta-analytical review. *J. Appl. Psych.* 91(2):259–271.