The impact of gender diversity on the performance of business teams: Evidence from a field experiment*

Sander Hoogendoorn Hessel Oosterbeek Mirjam van Praag

Abstract

This paper reports about a field experiment conducted to estimate the impact of the share of women in management teams on their business performance. As part of their curriculum, undergraduate students in business studies start up a team venture. We manipulated the gender composition of teams and assigned students, conditional on their gender, randomly to teams. We find that management teams with an equal gender mix perform better than male-dominated and female-dominated teams in terms of sales, profits and earnings per share. Our results indicate that this is due to more intensive monitoring and more equalized learning in mixed teams than in homogeneous teams.

JEL-codes: J16, L25, L26, M13, C93
Keywords: Gender diversity, team performance, entrepreneurship, field experiment, entrepreneurship education, board effectiveness

1 Introduction

Women are under-represented in corporate boards. In 2010, women held only 10 percent of the board seats at the top 300 European companies and just above 15 percent of board seats at Fortune 500 companies (Woods, 2010; Catalyst, 2010). A higher share of women in boards is often regarded as desirable. Given the current low share, increasing

*This version: February 2011. We are grateful to the Department of International Business Studies of the Amsterdam College of Applied Sciences for their support in carrying out this research. We thank Kristina Dahlin and seminar participants in Amsterdam, Lisbon, London and Maastricht for their comments and suggestions. We thank the Dutch Ministry of Economic Affairs for financial support and Stibbe for their sponsoring. The usual disclaimer applies. All authors are affiliated with the University of Amsterdam, Amsterdam School of Economics. Emails: s.m.hoogendoorn@uva.nl; h.oosterbeek@uva.nl; c.m.vanpraag@uva.nl.

the share of females leads to increased gender diversity. Commonly expressed arguments in the popular press in favor of more gender diversity are abundant and include: the need to destruct the glass ceiling, enlargement of the pool from which talent is attracted, improvement of firms' corporate identity and brand image, better stakeholder relationships, higher employee motivation, higher customer satisfaction, more complementarities and better mutual learning (Desvaux et al., 2007). Increases in female labor force participation have strengthened the policy relevance of more gender diversity in management boards. Some countries (Norway, The Netherlands, Spain, France, Iceland) even (plan to) enforce a higher representation of women by setting minimum quota of board seats for female directors.

Economic theory has remained relatively silent about the optimal share of women in (board) teams and the mechanisms that cause diverse boards to perform differently. Hamilton et al. (2003) point to the trade-off between the higher costs of coordination and communication due to more diversity and the benefits of a potentially more diverse pool of knowledge and skills and the accruing possibilities for (mutual) learning. Also among empiricists gender diversity has drawn little attention. Few empirical studies find evidence of mechanisms that may explain performance differences between gender diverse and homogeneous boards. Adams and Ferreira (2009) discuss 'mutual monitoring' as a mechanism and show that more gender diverse boards are associated with more intense monitoring practices. Dufwenberg and Muren (2006) derive results from a group dictator game played in the laboratory showing that gender diverse teams are more generous and more egalitarian.[1]

Despite the pleas and implemented policies for a more equal representation of women, there is little empirical evidence supporting that gender diversity leads to better team performance. On the contrary, four recent studies inquiring the impact of gender diversity on the performance of (board) teams find that a larger share of women does not have a positive impact on performance and may even be harmful.

Adams and Ferreira (2009) analyze data with characteristics of directors and boards from almost 2000 firms in the United States for the period 1996-2003. To address the possible endogeneity of the fraction of female directors, they use the fraction of male board members with connections to female directors in other board positions as instrumental variable. In addition they include firm fixed effects. Besides their finding that gender diverse boards monitor directors more intensively, they also find that diversity

[1]The theoretical perspective on the effect of team diversity on performance is also shaped by the management, sociology and psychology literature. For instance, Pelled (1996) and Pelled et al. (1999) argue there is a relationship between team performance and conflicts, which may be positive or negative, dependent on the character of the conflict, whereas Pelled (1996) adds that the characteristic features of diversity that shape the relationship with performance are their job-relatedness and visibility. Pelled argues that more visible and less job related diversity such as gender diversity may harm the productivity of the team.

has a positive impact on performance in 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. These findings are consistent with each other if internal and external monitoring are substitutes. On average, the effect of gender diversity on firm performance turns out to be negative. This leads Adams and Ferreira to conclude that mandating gender quotas for directors can reduce the value of well-governed firms.

Ahern and Dittmar (2010) and Matsa and Miller (2010) both measure the effect of board composition on firm performance by exploiting that publicly listed firms in Norway were forced to have at least 40 percent female directors by 2008. In 2006, when this law was implemented, only 9 percent of directors were women. Firms thus had to replace on average 30 percent of their board members. In a difference-in-differences framework, Ahern and Dittmar compare before-after differences between early compliers and late compliers. Matsa and Miller compare listed and unlisted companies and companies in Norway and in other Scandinavian countries in a double and triple differences framework. Both papers conclude that the forced replacement of board members resulted in a significantly negative impact on firm value and profit.[2]

Finally, Apesteguia et al. (2010) 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.[3]

While all four studies express awareness of endogeneity issues and attempt to address these, it is uncertain to what extent they achieved this. Adams and Ferreira use the fraction of male board members with connections to female directors as instrumental variable for the share of female board members. It is unclear whether this variable really passes the exclusion restriction required for an instrumental variable to be valid. Moreover, the first stage effect is significant, but the instrument barely passes the rule of thumb of an F-

[2]Ahern and Dittmar emphasize that board members' gender has no impact on firms' value once they control for other demographic characteristics such as age and experience, whereas Matsa and Miller focus on differences between male and female leadership styles (and those of females are more costly). Both of these explanations actually imply that the female directors added to the board due to the law perform worse in terms of financial outcomes. They do not relate this to differences in team dynamics.

[3]Related is also Hansen et al. (2006) which looks at the impact of gender diversity on the performance of groups on the group assignments of an undergraduate introductory management course. Male dominated groups perform 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 (2010) and Oosterbeek and Van Ewijk (2010).

statistic being at least equal to 10. As Ahern and Dittmar admit it is questionable whether the common trend assumption for early and late compliers in their study holds. They do not provide supportive evidence for this assumption by showing that early and late compliers experienced common trends before the law was adopted. Finally, Apesteguia et al. argue that reversed causality can be excluded because teams are formed before they have to perform, and because teams remain fixed over the entire game. This ignores, however, that participants may self-select into teams of different gender composition on the basis of (unobserved) characteristics that are correlated with performance.

The goal of this study is to provide uncontaminated evidence on the impact of the share of women on the financial performance of business teams. We do this by means of a field experiment in which we manipulated the share of women in business teams and assigned participants - conditional on their gender - randomly to teams. This is arguably the most convincing way to address issues of self-selection, omitted variable bias and reversed causality.

The field experiment was conducted in the context of the mandatory 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 a profit objective for the duration of one year. Students spend approximately one day a week on it. Incentives are strong: students can be 'fired' by team members and will not be graded in that case, whereas the grade has a substantial weight in the student's GPA and determines whether students may participate in the second year program. In addition, there is competition between teams to win the contest of showing the best business results in terms of sales, profits and shareholder value. The strong incentives to perform a substantial and truly joint task of setting up and running a company with the objective of maximizing profit and shareholder value are factors that contribute to the external validity of our results.[4]

Forty-five of these student companies are included in our experiment, with the share of women varying between 0.1 and 1.0. Using various performance measures and specifications, we consistently find an inverse u-shaped relation between teams' business performance and their share of women. Performance peaks when the share of women is between 0.5 and 0.6. In search of mechanisms underlying this positive effect of gender diversity on the team's business performance, we provide evidence that mutual monitoring and more equalized learning occur more often in mixed gender teams than in more homogeneous teams and that both factors have a positive impact on companies' performance.

The remainder of this paper is structured as follows. Section 2 gives more details of

[4]The size of the team is also realistic: the average European board is composed of 11.7 people (cf. Woods, 2010)

the context and design of the field experiment. Section 3 describes the data and reports results from randomization checks. Section 4 presents the empirical findings. Section 5 discusses and concludes.

2 Context and design

2.1 Context

The program that we study is organized 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 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 common goal. Each student company is supported by one or two advisers coming from the business world and sharing their experience with the students (European Commission, 2006). During the program the teams have to report to their teacher, business coach and accountant on a regular basis. The program is not a business simulation because everything about the venture is real, including tax and social security payments.

Ventures generally proceed as follows. After an interim CEO is appointed, the team starts brainstorming about potential products or services. Market research is then conducted to further analyze the business ideas that survived this process. Next, the core business activity and allocation of positions are defined.[5] Once the corporate plan has been finished, the students start raising capital and organize a shareholders' meeting. The teams can start their business operations if the majority of shareholders approves the corporate plan. From then on producing and marketing of products or services is the main activity of the team. Subsequently, all ventures are liquidated and each team has to write an annual report that needs approval of the final shareholders' meeting. Any profit will be proportionally divided among the shareholders.

The entrepreneurship program at the international business department of the Amsterdam College of Applied Sciences is mandatory, it lasts for an entire academic year and covers about one-fifth of students' first-year bachelor curriculum. The international business department is divided into five fields of study: management, business management,

[5]Table A1 in the Appendix lists all 45 teams and reports some characteristics including their gender mix and the product or service they sell.

Table 1. Numbers of students and teams, and share of women by field of study

Study field	Students	Teams	Female
Management	60	5	0.29
Business management	240	18	0.37
Financial management	27	2	0.40
Trade management Asia	105	9	0.35
Business languages	118	11	0.71
Total	550	45	0.44

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. Within study fields, students were assigned to 45 teams, giving an average team size of 12. Table 1 shows the numbers of students and teams formed by study field. It also shows that the average share of female students is 0.44, and that women are only overrepresented in the field of business languages.

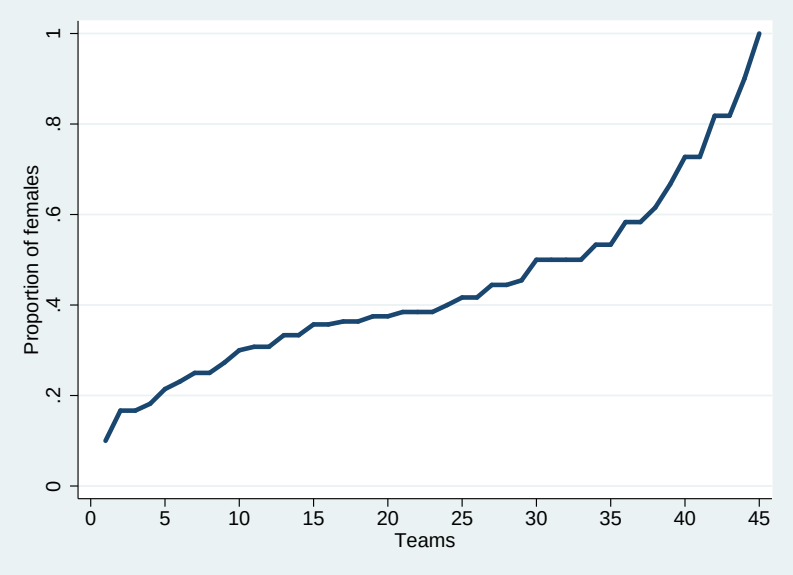
2.2 Design

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. Single-sex teams or teams with only one person of a specific sex were not appreciated by the college. We assigned 550 students to 45 teams and 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 whereas a few 'no shows' were also randomly distributed across teams (as they didn't know to which team they were assigned to at that stage).

Figure 1 shows the share of females per team with teams ordered from low to high share of females at the time the teams actually started, the baseline. This shows that there is substantial gender variation across teams. The share of females varies from 0.1 to 1.0. Only three teams ended up having a share of females below 0.2 or above 0.8.

The field of study coordinators were informed about the character of the exogenous variation we imposed. We urged them not to inform teachers. Teachers only knew that a research project was conducted which required to stick to the imposed team assignment. Coordinators and teachers were urged not to inform students. Based on interviews with students, we are confident about their uninformedness. 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.

Figure 1. Share of females per team: ordered from low to high



During the year, 104 students (19%) dropped out. 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 the year. Dropouts could still contaminate the design of the experiment if teams' gender composition is affected or if dropout rates vary across teams in relation to their gender composition. Neither is the case. The correlation coefficient between the teams' share of females at baseline and at the end of the program is 0.92. Regressing students' dropout status on the share of women at baseline and its square returns coefficients of -0.131 (s.e. 0.365) and 0.089 (s.e. 0.342), showing that students' dropout decision is not affected by the gender composition of their team.[6]

3 Data

3.1 Variables

In addition to administrative data and teams' annual reports, information was collected through three extensive 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. This information is required to assess whether team assignment was random,

[6] A linear specification gives a coefficient of -0.040 (s.e. 0.090)

given the gender distribution and to measure possible mechanisms explaining the effect of teams' gender composition 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, more than 60% of the students participated in higher general secondary education, roughly two-thirds of the population lives with their parents, about one-third has some work experience, and over 30% has a father who is or was an entrepreneur. Excluding non-Dutch exchange students (about 20%), nearly half of the remaining Dutch students has at least one parent not born in the Netherlands.[7]

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 (cf. Goldberg, 1990). This commonly used set of measures of personality has been shown to be an explanatory factor of entrepreneurship choices and outcomes (Zhao and Seibert, 2006; Shane, 2010). 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, persistence, risk taking, self-efficacy and social orientation. These traits are supposed to be constant over time and possibly influential for entrepreneurship decisions and outcomes (cf. 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 pro-activity (cf. Parker, 2009).[8]

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

To explain possible differences in performance between teams, the second follow-up

[7] We also randomized students to teams on the basis of their ethnic background. Results will be reported in a companion paper. Since gender diversity and ethnic diversity are orthogonal this will not affect the results reported here. The correlation is -0.0857 and not significantly different from zero.

[8] These measures are taken using the so-called Escan (Driessen, 2005), a validated self-assessment test based on 114 items. Based on the data collected in Oosterbeek et al. (2010) it has been slightly adapted to increase the validity of items when a population of students rather than entrepreneurs is involved. 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 test results have been shown to correlate significantly with objective measures of entrepreneurial performance in terms of survival, profits, income and sales (see Driessen and Zwart, 1999). The statements load into factors (with Cronbach alpha's ranging from 0.64 to 0.79) that the entrepreneurship literature has shown to be the most important traits and skills for successful entrepreneurship.

Table 2. Team characteristics and processes, measured at 2nd follow-up

	Scale	Mean	SD	Min	Max
Team characteristics					
Atmosphere	1-5	3.525	0.546	2.333	4.833
Conflicts	1-5	2.231	0.585	1.000	3.667
Friends	nr	2.313	0.646	1.000	3.750
Layoffs (dummy = 1 if any)	0/1	0.490	0.500	0.000	1.000
Peer-reviewed efforts	1-10	6.938	0.555	6.139	9.167
Satisfaction with coach	1-5	3.015	0.478	1.750	4.250
Satisfaction with results	1-5	3.438	0.493	2.500	4.500
Subgroups (dummy = 1 if any)	0/1	0.458	0.498	0.000	1.000
Processes	Cronbach alpha				
Group potency	0.87	10.849	1.513	8.294	15.166
Decision making	0.70	1.583	1.242	-0.729	4.595
Mutual monitoring	0.88	9.985	1.018	7.832	12.690
Coordination	0.80	1.831	1.113	-0.314	4.899
Credibility	0.66	2.870	0.610	1.412	4.839
Specialization	0.66	7.878	0.602	5.968	9.656

Note: Based on information from 45 teams.

survey contains questions concerning the development of the team and processes that took place within the teams. Questions related to teams' development address atmosphere, conflicts, peer-reviewed individual effort, friendships, layoffs, perceived diversity levels, satisfaction with the teacher, 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), mutual monitoring (Langfred, 2004) and coordination, credibility and specialization (Lewis, 2003). Table 2 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 Section 4 we will examine to what extent the scores on these variables are related to teams' performance on the one hand and to the teams' gender composition on the other.

The outcome variables in our analyses are measures of teams' business performance. Information about business performance was retrieved from teams' annual reports. We distinguish the following performance measures: sales (in euros), profits (in euros), a binary indicator for positive profits, and profits per share (in euros). Based on a careful analysis of the annual reports, we were left with the impression that some of these were put together without careful application of a minimal level of accounting rules. Taxes, labor incomes, depreciation and the costs of unsold goods were accounted for in various ways. However, as the bottom line is whether or not to disappoint the shareholders, the sign of the resulting profit level counts more heavily, is rather visible and therefore likely to be more reliable. We therefore add the binary indicator for positive profits to the

Table 3. Descriptive statistics of outcome variables

	Mean	SD	Min	Max	Corr(Sales)
Sales (euro)	841.2	699.0	0	4209.5	1.00***
Profit (euro)	-69.2	317.8	-1016.4	477.2	0.26*
Pr(Profit)>0	0.51	0.25	0.00	1.00	0.36**
Profit/share (euro)	-0.51	6.42	-15.48	15.64	0.27*

Note: Based on information from 43 teams. ***/**/* denotes significance at the 1%/5%/10%-level.

usual set of performance measures. Table 3 shows descriptives of the four performance measures based on the information from annual reports of 43 teams. Average sales amount to 841 euros, with a standard deviation of 699 euros. The worst performing team has no sales, while the best performing team sells for more than 4000 euros. Profits are on average negative at -69 euros. The team with the lowest profit loses 1016 euros, while the highest profit is 477 euros. 22 teams make positive profit, while 21 teams run a loss. Expressed per share, profits vary between -15 and +15 euros. The final column in the table shows that the three profit measures are positively and significantly correlated with sales.

3.2 Randomization

This subsection first examines whether students are, conditional on their gender, randomly assigned to teams of different gender composition. This boils down to regressing – separately for male and female students – students' characteristics on the share of women in their team. The first four columns of Table 4 report the results. Male students who are assigned to teams with many women are not different in terms of personal characteristics or personality traits from male students who are assigned to teams with few women. The same holds for female students assigned to teams with different fractions of women. There are some – but not systematic – differences with respect to entrepreneurial knowledge and skills. There are, however, clear differences with regard to field of study. This is unavoidable given that the share of women varies across fields of study and given that teams had to be formed within fields of study.

The last four columns in Table 4 report results in which the share of women is replaced by a gender diversity index (the share of women times the share of men). No systematic differences are observed in the characteristics of males and females between more and less gender diverse teams. The field of study 'Financial Management' forms the exception. Based on the results shown in this table we will perform various robustness checks in Section 4.

Teams' share of women or teams' gender diversity may correlate with other team characteristics. Table 5 reports these correlations. In line with expectations, the share of

Table 4. Random assignment of male and female students at the individual level

		Share of women				Diversity index			
		Males		Females		Males		Females	
Personal characteristics	Age	0.006	(0.005)	0.002	(0.005)	−0.001	(0.002)	−0.002	(0.002)
	Ethnicity	−0.020	(0.019)	−0.031	(0.026)	0.003	(0.005)	0.008	(0.010)
	Nationality	0.002	(0.022)	−0.030	(0.030)	0.000	(0.009)	0.008	(0.008)
	Mathematics grade	0.022	(0.015)	0.026	(0.025)	0.014∗	(0.007)	−0.017	(0.016)
	Grade point average	0.032	(0.029)	0.031	(0.045)	0.014	(0.009)	−0.039	(0.037)
Field of study	Business Management	−0.040	(0.042)	−0.187∗∗∗	(0.052)	0.012	(0.016)	0.051∗	(0.029)
	Management	−0.075	(0.047)	−0.156∗∗∗	(0.048)	−0.013	(0.023)	−0.008	(0.026)
	Trade Management Asia	−0.063	(0.055)	−0.115∗	(0.060)	−0.025	(0.026)	0.034	(0.024)
	Business Languages	0.282∗∗∗	(0.044)	0.335∗∗∗	(0.044)	0.012	(0.016)	−0.078∗	(0.039)
	Financial Management	0.029	(0.027)	−0.119∗∗∗	(0.038)	0.031∗∗∗	(0.010)	0.043∗∗	(0.020)
Personality traits	Agreeableness	0.007	(0.007)	0.005	(0.007)	0.001	(0.002)	−0.001	(0.002)
	Conscientiousness	0.002	(0.006)	0.005	(0.007)	0.002	(0.002)	−0.005	(0.003)
	Extroversion	−0.004	(0.004)	−0.002	(0.006)	0.000	(0.002)	0.001	(0.002)
	Neuroticism	0.001	(0.006)	−0.004	(0.008)	0.001	(0.002)	0.003	(0.004)
	Openness to experience	0.002	(0.007)	−0.008	(0.009)	0.000	(0.003)	0.007	(0.005)
Knowledge	Business	−0.015	(0.010)	−0.023	(0.015)	0.004	(0.004)	−0.004	(0.007)
	Management	−0.013	(0.011)	−0.025∗∗	(0.010)	0.001	(0.003)	0.004	(0.004)
	Entrepreneurship	−0.008	(0.010)	−0.005	(0.012)	0.003	(0.003)	0.000	(0.004)
	Strategy	0.001	(0.010)	0.003	(0.014)	0.003	(0.004)	−0.007∗	(0.004)
	Organization	−0.001	(0.009)	0.003	(0.017)	0.001	(0.003)	−0.013	(0.010)
	Administration	0.003	(0.010)	−0.023∗∗	(0.010)	0.004	(0.004)	−0.006	(0.004)
	Leadership	−0.005	(0.009)	−0.006	(0.013)	−0.002	(0.002)	−0.001	(0.004)
Skills	Analyzing	0.002	(0.007)	−0.006	(0.011)	0.001	(0.002)	−0.002	(0.004)
	Creativity	0.002	(0.006)	−0.013	(0.012)	0.001	(0.002)	0.008	(0.008)
	External orientation	0.005	(0.008)	0.001	(0.010)	−0.003	(0.002)	−0.004	(0.006)
	Flexibility	0.008	(0.008)	0.002	(0.009)	0.002	(0.003)	0.003	(0.004)
	Market awareness	0.006	(0.005)	−0.009	(0.010)	0.003	(0.002)	0.009	(0.007)
	Motivating	0.012∗	(0.006)	−0.020∗	(0.012)	0.005∗	(0.002)	0.010	(0.008)
	Networking	0.009	(0.006)	−0.001	(0.008)	0.002	(0.002)	−0.005	(0.004)
	Organizing	0.000	(0.005)	−0.011	(0.010)	0.001	(0.002)	−0.005	(0.005)
	Pro-activity	0.003	(0.006)	−0.018∗∗	(0.008)	0.000	(0.002)	0.007	(0.004)

Note: Each coefficients comes from a regression at the individual level of the row variable on the column variable, separately for men and women. Standard errors in parentheses. ***/**/* denotes significance at the 1%/5%/10%-level.

Table 5. Correlations between share of women/gender diversity and other (average) team characteristics at baseline

	Share of women		Diversity index	
Individual characteristics (averages)				
Age	0.023	(0.042)	−0.013	(0.014)
Ethnicity	−0.105	(0.167)	0.016	(0.070)
Nationality	−0.127	(0.262)	0.022	(0.082)
Mathematics grade	0.149∗∗	(0.068)	−0.019	(0.047)
Grade point average	0.235	(0.145)	−0.111	(0.118)
Field of study				
Business Management	−0.123∗∗	(0.052)	0.038	(0.023)
Management	−0.146∗∗∗	(0.051)	−0.012	(0.023)
Trade Management Asia	−0.117∗	(0.062)	0.002	(0.028)
Business Languages	0.359∗∗∗	(0.056)	−0.055	(0.043)
Financial Management	−0.048	(0.040)	0.048∗∗∗	(0.016)
Big five characteristics (average)				
Agreeableness	0.100∗∗	(0.042)	−0.004	(0.015)
Conscientiousness	0.066	(0.051)	−0.015	(0.023)
Extroversion	−0.071∗	(0.039)	0.010	(0.015)
Neuroticism	−0.067	(0.053)	0.028	(0.029)
Openness to experience	−0.150∗	(0.083)	0.059	(0.045)
Entrepreneurial knowledge (average)				
Business	−0.231∗∗	(0.088)	0.012	(0.036)
Management	−0.183∗∗∗	(0.062)	0.025	(0.028)
Entrepreneurship	−0.095	(0.084)	0.026	(0.026)
Strategy	−0.154	(0.099)	0.009	(0.037)
Organization	−0.029	(0.100)	−0.065	(0.056)
Administration	−0.128	(0.106)	−0.009	(0.045)
Leadership	−0.136	(0.131)	0.001	(0.039)
Entrepreneurial skills (average)				
Analyzing	−0.094	(0.084)	0.009	(0.029)
Creativity	−0.069	(0.072)	0.044	(0.035)
External orientation	−0.003	(0.065)	−0.036	(0.029)
Flexibility	0.056	(0.064)	0.024	(0.023)
Market awareness	−0.070	(0.065)	0.057	(0.037)
Motivating	−0.077	(0.090)	0.096∗	(0.053)
Networking	0.087	(0.068)	−0.026	(0.030)
Organizing	0.047	(0.079)	−0.032	(0.042)
Pro-activity	−0.107	(0.096)	0.059	(0.036)

Note: Team size does not significantly co-vary with the percentage of females in the team. ***/**/* denotes significance at the 1%/5%/10%-level.

women does co-vary with some of the team characteristics. It is positively correlated with team-average mathematics grade in secondary school and with team-average agreeableness. It is negatively correlated with team-average knowledge of business and management. These significant correlations are unavoidable if women and men differ systematically in these characteristics unless we would attempt to balance these other characteristics. Such balancing would necessarily lead to women/men with certain characteristics having a higher probability to be assigned to a team with a high share of women.[9]

A teams' gender diversity index does not correlate significantly with team-average variables at the baseline measurement. Apparently, the average gender diverse team is comparable to a cross-section of female dominated and male dominated teams. Again, like in Table 4, the only exception is the field of study 'Financial Management'. The robustness checks in Section 4 will address the extent to which our results are driven by these gender composition related differences between teams at baseline.

4 Results

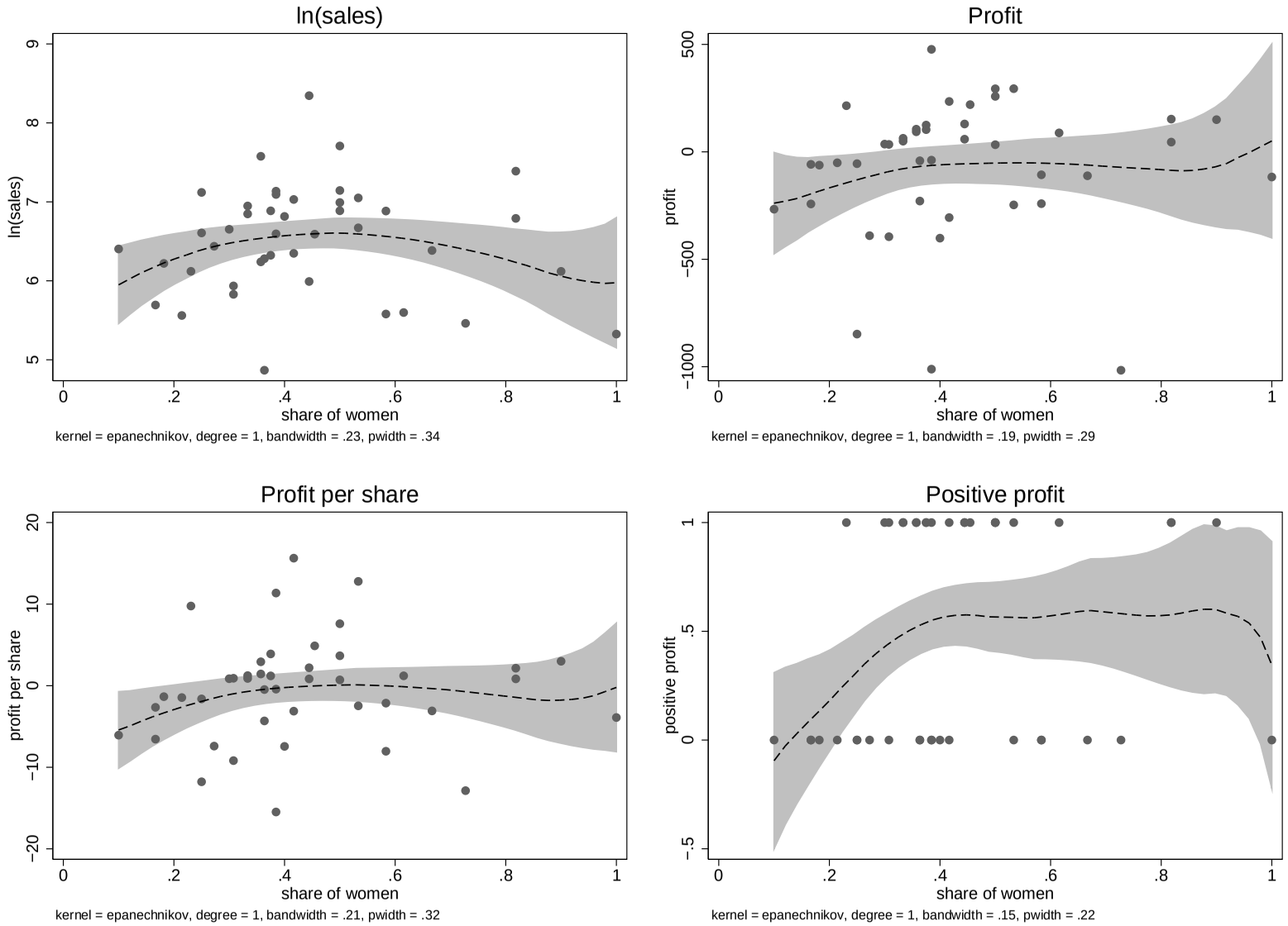
4.1 Main finding

Figure 2 shows the relation between the share of women in a team and four measures of teams' performance: (ln) sales, profit, profit per share and the probability of profits being positive. The graphs are based on kernel-weighted local polynomial smoothing (details are reported below each graph). Dots represent the actual team results and the shaded areas the 90% confidence intervals. In all graphs the relation between team performance and the share of women tends to follow an inverse u-shape. Given a low share of women in a team, team performance improves when the share of women increases up to a share of female team members of approximately 55. Beyond that percentage of female team members, further increases in the share of women tend to reduce teams' performance. This latter effect is most pronounced for (ln) sales, but also for the other performance measures there appears to be a peak around 0.55.

Further evidence of this relationship is provided in Table 6, which presents results from various regressions. In panel A performance measures are regressed on teams' share of women and its square. The first column reports results from a least squares regression in which teams' sales is the dependent variable. Both the linear term and the quadratic term are significantly different from zero with positive and negative signs, respectively. The coefficients imply that sales peak when the share of women equals 0.55. To examine whether the results in the first column are sensitive to outliers in the outcome variable, the

[9] This issue is well-known in the literature that examines the impact of gender peer effects in education (see Hoxby, 2000; Lavy and Schlosser, 2010).

Figure 2. Effect of share of women on team performance



second column reports results from a median regression. Coefficients are very similar and the share of women at which a maximum is reached is almost the same (0.54), implying that the results in the first column are not driven by outliers. As another variation on functional form, the third column reports results with the dependent variable transformed to the natural logarithm of sales. Both coefficients are different from zero at the 1%-level, the relationship is again inverse u-shaped and ln sales are maximized when the share of women equals 0.52. According to the results in the third column, an increase of the share of women from 0.3 to 0.4 increases sales by 20%. While an increase of the share of women from 0.4 to 0.5 increases sales by 7%.

Columns (4) and (5) report results from least squares and median regressions with profits as the dependent variable. In column (4) neither the linear term nor the quadratic term is significantly different from zero, although again the relation between performance and the share of women is inverse u-shaped and profits peak at a share of women equal to 0.61. The results from the median regression show that these findings are sensitive to outliers. The point estimates are slightly larger and both coefficients are now significantly different from zero at the 10%-level. The optimum share of women drops to 0.57. In the next column performance is measured as a binary indicator for profits being above zero or equal to or below zero. Both coefficients are significantly different from zero at the 5%-level, and the share of women that maximizes the probability of profits being positive equals 0.61. Finally the last two columns report results from least square and median regressions when the dependent variable is profits per share. The two sets of coefficients corroborate previous results; all coefficients are significantly different from zero, the relation is inverse u-shaped and performance is maximized at a share of women equal to 0.57 or 0.56.

The quadratic specification is not very flexible as it basically treats teams with 10% women and 90% men the same as teams with 90% women and 10% men. One might therefore worry that the inverse u-shape is mainly caused by the upward sloping concave relationship between performance and share of women for teams with less than 50% women, but poorly fits the relation between performance and the share of women for teams with at least 50% women. To address this concern, we also fitted spline functions allowing for different slopes below and above 50% of women.[10] The results are reported in panel B. All coefficients for the share of women when this is below 0.5, are positive and (with one exception) statistically significant. All coefficients for the share of women when this is at least equal to 0.5 are negative, and in two cases significantly so. Clearly, the number of teams here limits the precision of the estimates.

[10]We allow for a spline at 0.5 instead of for instance 0.55 (the optimum according to the quadratic specification) because the number of teams with at least 55% women is rather small.

Table 6. Effect of share of women on team performance

	Sales			Profits		Positive profits	Profits per share	
	Mean	Median	ln(sales)	Mean	Median	Mean	Mean	Median
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A: Polynomial								
Share women	4796.5***	4012.6**	4.983***	915.5	1277.2*	3.930**	30.927**	35.836***
	(1697.1)	(1719.1)	(1.770)	(660.2)	(755.8)	(1.587)	(14.799)	(13.218)
Share women squared	-4351.9***	-3703.0**	-4.794***	-751.3	-1121.2*	-3.236**	-27.151**	-32.001***
	(1521.3)	(1494.0)	(1.541)	(553.2)	(655.6)	(1.420)	(12.705)	(11.441)
Optimum	0.55	0.54	0.52	0.61	0.57	0.61	0.57	0.56
B: Spline								
Share women<0.5	2374.0**		2.462**	682.2		1.868**	18.750*	
	(1088.6)		(1.061)	(507.2)		(0.771)	(10.025)	
Share women$\geqslant 0.5$	-1800.9*		-2.261**	-359.6		-0.766	-11.900	
	(948.1)		(0.924)	(456.6)		(0.694)	(9.026)	

Note: Number of teams equal 43. The linear effect of the proportion of females on venture performance turned out insignificant in all specifications. We also tested higher order specifications and these turned out being insignificant too. Standard errors are given in parentheses. ***/**/* denotes significance at the 1%/5%/10%-level.

The results presented in this subsection all point in the same direction: the business' performance of teams first increases when the share of women in the team increases and then decreases in the share of women in a team. The precise share of women at which teams' performance peaks, varies a bit across performance measures, but in all specifications the optimum is around 0.55. Gender diverse teams perform better than male dominated or female dominated teams.

4.2 Robustness

The results presented in the previous subsection show an inverse u-shape effect of teams' gender composition on their performance with a peak around 55 percent female team members. This result turns out robust, at least upon variations in performance measures. The results are also not driven by outliers.

We performed some additional robustness checks by including additional controls or excluding observations from the regression analyses presented in Table 6. We limit the analyses to specifications with ln sales, the dummy indicator for positive profits and profits per share as performance measures. Thus, robustness checks are performed based on the specifications in columns (3), (6) and (7) of Table 6.

We repeated the analyses, now including all possible control variables discussed in Section 3, one-by-one, due to the small number of observations. None of the results previously obtained was affected significantly by any of the included variables, including indicators of the teams' distribution of individual personality, skills or knowledge levels. The main result was also maintained while controls were included for team effort, team results in terms of learning or appreciation and team processes such as conflicts or coordination. Table A2 in the Appendix only shows these results for the inclusion of specific controls (groupwise for the sake of presentation) to address the possibility that the main result is driven by specific (pretreatment) differences between males and females as were revealed in Table 5.

In particular, Table 5 showed some indications that teams with a higher share of females have higher average high school math grades, lower average levels of business and management knowledge and higher team-average scores on agreeableness. Moreover, the various fields of study are also associated with a teams' gender composition, especially 'Financial Management'. Table A2 in the Appendix shows that none of the results is changed significantly by controlling for (i) average math grade, business and management knowledge (ii) fields of study and (iii) the average scores on the big five personality characteristics including agreeableness. Table A3 shows that the results also remain similar when the two teams from 'Financial Management' are excluded. All in all, these results alleviate our concern that the main results are driven by any contaminating pretreatment

differences between teams of different gender compositions.[11]

4.3 Mechanisms

The previous subsections establish that there is a causal link between teams' gender diversity and their business outcomes. This subsection inquires possible mechanisms underlying this finding. In particular, we explore three possible mechanisms that we derive from the literature and were discussed in Section 1. These are: (i) the possible benefits of a more diverse pool of knowledge and skills due to complementarities (cf. Hamilton et al., 2003); (ii) (mutual) learning (cf. Hamilton et al., 2003) and an egalitarian distribution of (learning) benefits (cf. Dufwenberg and Muren, 2006); and (iii) (mutual) monitoring practices.[12] These exploratory efforts mainly aim at providing directions for future research in explaining differences between gender diverse and homogeneous teams.

Complementarities Men and women in mixed teams may complement each others' skills and knowledge. The resulting broader base of skills and knowledge may have a positive impact on business outcomes. To test this hypothesis derived from Hamilton et al. (2003), we first derive the correlations between within team skill and knowledge diversity and the team's gender composition. The team skill/knowledge diversity is measured by the standard deviation of the individual baseline scores of each observed relevant skill and knowledge area within teams.[13] Second, for those skills and knowledge areas whose standard deviation is associated with the gender composition of the team, we calculate the correlation between this standard deviation and the teams' performance measures sales, profit and profit/share. None of the results in the first stage turn out significant; more gender diverse teams are not associated with skills or knowledge levels at baseline with a wider spread. Thus, we find no evidence that skill and knowledge complementarities may underlie the higher effectiveness of more gender diverse teams.

Learning and egalitarianism When teams learn, mean skill and knowledge levels increase. Learning may be related to the gender composition of the team. This may be due to differential initial distributions of skills and knowledge levels – for which we find no evidence – or due to differential team processes that may be unobserved. Testing whether the positive effect of gender diversity on business performance is related to more learning, proceeds in two steps. First, the changes in teams' mean skills/knowledge levels

[11] Nevertheless, some of the controls themselves may be associated significantly with team performance.

[12] Along with monitoring practices, we also address the role of other team processes, such as conflicts and team atmosphere that the management literature has pointed out to be associated with team diversity and team performance (Pelled, 1996; Pelled et al., 1999).

[13] We control for team size in all of these regressions that establish this conditional correlation coefficient to account for the fact that this standard deviation may be related to size.

in the various areas are regressed on the share of women and its square, where we control for the baseline levels of the skill/knowledge area and team size. Second, for those skills/knowledge levels whose development is associated significantly with gender composition, we measure the correlation of the change in the skill/knowledge level with the various business performance measures in a regression framework with the same controls. The team average increases in skill/knowledge levels turn out to be unrelated to teams' gender composition. There is thus, on average, not more or less learning in gender diverse teams than in other teams.

The average learning gain in a team is not informative about the distribution of the learning gain among team members. This distribution of learning gains may be different in gender diverse teams than in male dominated or female dominated teams. For four skills this is the case. We obtain this result in regressions of learning gains on gender composition (proportion females and its square) where we control for team size and the base level of the distribution of skills and knowledge. Table 7 reports the impact of the share of women and its square on the within-team standard deviation of learning in these four different areas. This shows significant u-shaped relationships between the share of women and the within-team standard deviations in the gains in creativity, external orientation, organizing and pro-activity. More gender diversity thus reduce the within-team variances in learning in these areas. Apparently, gender diverse teams distribute the benefits from learning more egalitarian.

In regressions relating these changes in the distribution of skill levels to team performance (again controlling for the base levels and team size) we find significant negative associations with sales for organizing skills and pro-activity (underlined in the table) and with profit and/or profit/share for creativity, external orientation and organizing skills (bold in the table).[14] This suggests that higher variances in learning are detrimental for teams' performance. It thus seems that one of the underlying mechanisms of the positive effect of gender diversity on business performance may be a more equal spread of learning benefits in more gender diverse teams.[15]

Monitoring Adams and Ferreira (2009) report that in a sample of US firms, female directors have better attendance records than male directors and that male directors have fewer attendance problems in more gender diverse boards. Women are also more likely to

[14]The regression coefficients (and standard errors) for the regression of the standard deviation in a team of the developments of its members in creativity, external orientation, organizing skills and pro-activity and log of sales are -0.325 (0.225), -0.157 (0.173), -0.433 (0.253) and -0.536 (0.212) respectively. For profits the numbers are -0.748 (0.463), -0.709 (0.348), -1.686 (0.606 and -0.369 (0.514).

[15]However, including the changes in these spreads in the specifications of which Table 6 shows the results remain unchanged: The positive effect of gender diversity on business performance remains similar and significant.

Table 7. Gender diversity, learning and performance

		Mean	SD
△ Creativity	Share female	0.479	**-4.228*****
		(1.666)	(1.530)
	Share female squared	−1.007	**3.607****
		(1.718)	(1.527)
△ External orientation	Share female	1.391	**-3.854****
		(2.011)	(1.812)
	Share female squared	−1.334	**3.244****
		(2.061)	(1.579)
△ Organizing	Share female	0.043	**-3.250****
		(1.365)	(1.252)
	Share female squared	−0.457	**3.226*****
		(1.128)	(1.061)
△ Pro-activity	Share female	0.790	-3.984***
		(1.244)	(1.052)
	Share female squared	−0.592	3.989***
		(1.157)	(0.962)

Note: Based on information from 43 teams. Regressions in column 1 and 2 control for teamsize and the level of the (mean or standard deviation of the) skill at baseline. When the coefficient is underlined (bold) the effect of the skill on sales (profit) is significantly positive. Standard errors are given in parentheses. ***/**/* denotes significance at the 1%/5%/10%-level.

join monitoring committees. Accordingly, they find that CEO turnover is more sensitive to firm performance and directors' pay is more strongly tied to performance in firms with more gender diverse boards. Together these results suggest that gender diverse boards allocate more effort to monitoring. We collected information to measure the level of mutual monitoring (along with other team processes) in both follow-up surveys.[16] Table 8 reports results from least square regressions of monitoring in a team on the team's share of women and its square. In the first column the dependent variable is the level of monitoring measured in the first follow-up (in January 2009), in the second column the dependent variable is the level of monitoring measured in the second follow-up (in May 2009), and in the third column the dependent variable is the change in monitoring between the first and second follow-ups.

In the first follow-up – four months after the start of the program – we see no significant impact of gender composition on the level of monitoring. However, in the second follow-up – administered at the end of the program – there appears to be a significant relation between gender composition and the level of monitoring. Again the relationship

[16]The measure of monitoring is based on four items: (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 (see Langfred, 2004). Cronbach's alpha for monitoring equals 0.88.

Table 8. Effect of share of women on monitoring

	Monitoring January	Monitoring May	Change May - Jan
Share female	-3.453	5.789**	10.412**
	(3.048)	(2.773)	(4.534)
Share female squared	3.253	-5.588**	-10.425**
	(2.584)	(2.780)	(4.292)

Note: Based on information from 43 teams. Monitoring is measured as based on factor analysis of four statements that were valued by individual students, see the text. The Cronbach alpha of 0.88 indicates the validity of the factor. Regressions in columns 1 and 2 control for teamsize, whereas column 3 controls for both teamsize and the level of monitoring intensity at first follow-up. When the coefficient is underlined (bold, italic) the effect of monitoring intensity on sales (profit, profit/share) is significantly positive. Standard errors are given in parentheses. ***/**/* denotes significance at the 1%/5%/10%-level.

is inverse u-shaped. Members of more gender diverse teams monitor each other more intensively: The intensity is an increasing function of the share of women until this share is 0.52 and decreases afterward. A similar pattern is found for the change in monitoring between the first and second follow-ups, now the relationship peaks at a share of women equal to 0.50. We thus find that during the second half of the program monitoring in teams is more intense in gender diverse teams than in more homogeneous teams. This is consistent with the findings reported by Adams and Ferreira (2009), although we should note that in their analysis more gender diversity is equivalent to a larger share of women because there are no companies in their data set for which the share of women in the board of directors exceeds 0.50.

The higher level of monitoring in gender diverse teams can of course only explain the better performance of these teams if monitoring is positively related to performance. This is the case. The correlations between monitoring (measured as the change between the first and second follow-up) and the various business outcomes are: 0.189 (s.e. 0.082) for (ln) sales, 0.007 (s.e. 0.163) for profits and 0.605 (s.e. 0.287) for profits per share. The bold and underlined coefficients in Table 8 indicate that the regression coefficient of the monitoring variable is significantly positive in the sales and profit per share equations. Nevertheless, upon adding monitoring as a control in the equations in Table 6 the main effects remain similar and significant. Hence, more intense monitoring in gender diverse teams only partially explains why these teams perform better.[17]

[17] In the same manner as for monitoring, we have also checked whether any of the other team processes is related to gender diversity (and if so to team performance). This was not the case. Moreover, 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 (see Appendix Table A1) in many different ways based on the information in the annual reports. No systematic differences were observed in the market orientation of teams of various levels of diversity.

5 Discussion and conclusion

The key finding of this study is that of a causal inverse u-shaped impact of the share of women in a team on the business outcomes of the team. Performance peaks when the share of women is around 0.55. Further investigation suggests that this positive impact of gender diversity is caused by more intense mutual monitoring and more equal learning in gender diverse teams.

In the introduction we motivated our field experiment by reference to the public discussion about gender diversity in boards of directors in companies, to policy measures in several countries and to recent papers on this topic. We argued that while some recent studies attempt to address endogeneity issues, results may still be biased. Our field experiment is likely to provide clean unbiased evidence, because we were in the position to randomly assign participants to business teams with different shares of women. This comes, however, at a cost. While previous studies (Adams and Ferreira, 2009; Ahern and Dittmar, 2010; Matsa and Miller, 2010) use data from people who really made it into boards of directors of large companies, our study uses data from students in international business studies who run a student company for the duration of one year. We believe, however, that several features of the context in which our field experiment was conducted, contribute to the generalizability of our findings. First of all, the tasks that students perform in their companies are similar to the tasks performed by boards of directors in their companies; there is a CEO and there are managers responsible for different departments of the company. Even the size of teams closely resembles the size of boards in companies. Second, students attribute a substantial amount of their time to their companies. It is not a full-time job, but their input of time is certainly non-trivial. Third, students face strong incentives. Underperformance can result in study delay, which in turn can have serious financial consequences for the students. Moreover, many teams are motivated by the prospect to be the best performing business team.

An important feature of the context of the college where we conducted our study is that men and women are of similar quality and that the numbers of men and women are roughly equal. This is currently not the situation among the people that have the knowledge and experience that qualifies them for a position in a board of directors. The average quality of men in that group is probably equal to the average quality of women in that group, but the men outnumber the women in that group. To get to equal numbers, experienced and qualified men have to leave and inexperienced and perhaps not yet qualified women have to enter the group. This is exactly what happened in Norway and what explains the reduction in firms' performance. Our study shows that if there are enough women that are equally qualified as men, it is in firms' best interest to increase the share of women in their boards. It would therefore not be surprising if results like those re-

ported in this study have in the end a larger impact on the position of women in boards of directors than policies based on good intentions but which ignore supply constraints, such as the quotas set in Norway and other countries.

References

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

Ahern, K. and Dittmar, A. (2010). The changing of the boards: The value effect of a massive exogenous shock. Unpublished working paper.

Apesteguia, J., Azmat, G., and Iriberri, N. (2010). The impact of gender composition on team performance and decision-making: Evidence from the field. Unpublished working paper.

Catalyst (2010). Catalyst Census of Women Board Directors of the Fortune 1000. Technical report, Catalyst, New York.

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

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

Driessen, M. (2005). E-scan ondernemerstest. Technical report, 's Graveland. Entrepreneur Consultancy BV, The Netherlands.

Driessen, M. and Zwart, P. (1999). The role of the entrepreneur in small business success: the Entrepreneurship Scan. Working paper, University of Groningen.

Dufwenberg, M. and Muren, A. (2006). Gender composition in teams. *Journal of Economic Behavior and Organization*, 61(1):50–54.

European Commission (2006). Entrepreneurship education in Europe: Fostering entrepreneurial mindsets through education and learning. In *Final Proceedings Conference Entrepreneurship Education*, Oslo.

Goldberg, L. R. (1990). An alternative description of personality: The big five factor structure. *Journal of Personality and Social Psychology*, 59(6):1216–1229.

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

Hansen, Z., Owan, H., and Pan, J. (2006). The impact of group diversity on performance and knowledge spillover-An experiment in a college classroom. NBER Working Paper.

Hoxby, C. (2000). Peer effects in the classroom: Learning from gender and race variation. NBER Working Paper.

Karlan, D. and Valdivia, M. (2011). Teaching entrepreneurship: Impact of business training on microfinance clients and institutions. *Review of Economics and Statistics*, forthcoming.

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

Lavy, V. and Schlosser, A. (2010). Mechanisms and impacts of gender peer effects at school. *American Economic Journal: Applied Economics*, forthcoming.

Lewis, K. (2003). Measuring transactive memory systems in the field: Scale development and validation. *Journal of Applied Psychology*, 88:587–604.

Matsa, D. A. and Miller, A. R. (2010). A female style in corporate leadership? Evidence from quotas. Unpublished working paper.

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

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

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

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

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

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

Pelled, L., Eisenhardt, K., and Xin, K. (1999). Exploring the black box: an analysis of work group diversity, conflict, and performance. *Administrative Science Quarterly*, 44(1):1–3.

Shane, S. (2010). Entrepreneurship and the big five personality traits: A behavioral genetics perspective. Unpublished working paper.

Woods, D. (2010). Female representation on UK boards has risen to 13.6 *HR Magazine, United Kingdom*, 05 October:xx–xx.

Zhao, H. and Seibert, S. (2006). The Big Five personality dimensions and entrepreneurial status: A meta-analytical review. *Journal of Applied Psychology*, 91:259–271.

Table A1: Team characteristics

#	Name	Female (share)	Team size	Field*	Sales (euro)	Profits (euro)	Profit/ share	P/S*
1	A-Card	0.25	14	BM	1236.15	-848.05	-11.78	S
2	A'dam Gadgets	0.36	11	BM	534.12	-41.40	-0.47	P
3	Appie	0.90	9	BL	454.75	149.86	3.00	P
4	Aqua de Coctail	0.42	11	FM	1130.47	-305.94	-3.12	P
5	ArtEco Bags	0.40	9	M	912.00	-401.69	-7.44	P
6	BubbleMania	0.18	10	BM	503.00	-61.79	-1.34	P
7	D'Wine	0.25	8	M	740.00	-55.00	-1.62	P
8	Eastern Green	0.36	13	BM	513.00	105.51	2.93	P
9	Escapade Inc	0.67	9	BL	592.55	-111.30	-3.09	P
10	eyeBMA	0.38	14	BB	557.50	124.66	3.90	P
11	Firefly	0.50	11	BB	2225.65	293.62	3.67	P
12	Fl!pthat	0.23	11	BM	455.00	214.88	9.77	S
13	Ginger	0.58	11	TMA	976.50	-106.81	-2.14	P
14	Himitsu	0.30	10	TMA	775.00	36.00	0.86	n/a
15	I-Care	0.38	14	BM	1204.45	477.15	11.36	P
16	iJoy	0.36	12	TMA	1952.85	93.56	1.44	P
17	I-Juice	0.38	13	BM	1255.38	-38.54	-0.42	P
18	IMSC	0.27	10	M	625.00	-390.00	-7.41	n/a
19	iShield	0.44	10	BM	4209.49	129.76	2.20	P
20	KISBag	1.00	9	BL	205.48	-117.02	-3.90	P
21	Laservibes	0.36	11	BM	130.00	-228.90	-4.32	S
22	Mengelmoes	0.33	10	TMA	941.50	63.14	1.24	P
23	My-Buddy	0.17	10	TMA	297.00	-58.33	-2.65	P
24	Nine2Five	0.73	12	BL	235.45	-1016.36	-12.87	P
25	Picture Perfect	0.21	14	BM	260.09	-50.87	-1.45	P
26	Pietje Plu	0.73	11	BL	n/a	n/a	n/a	P
27	Pocket Memory	0.38	15	FM	978.94	103.46	1.20	P
28	Pro'Lux	0.31	13	BM	378.25	-394.90	-9.18	P
29	Qwinlok	0.31	12	TMA	340.00	34.61	0.91	P
30	Reflection	0.82	11	BL	889.51	45.43	0.84	P
31	SAME	0.82	9	BL	1618.35	152.37	2.15	P
32	Sappho	0.50	7	BL	980.00	n/a	n/a	n/a
33	Sharity	0.58	11	BL	265.00	-241.12	-8.04	P
34	ShoeTattoo	0.62	13	BL	270.00	88.32	1.21	S
35	Student Promotion	0.42	13	BM	571.32	234.54	15.64	S
36	StuPill	0.38	13	M	731.33	-1011.33	-15.48	P
37	Test-a-Holic	0.45	11	BM	728.45	219.77	4.88	P
38	We-Do Solutions	0.10	9	TMA	604.00	-266.82	-6.06	P
39	We 'R U	0.33	13	BM	1041.11	49.77	0.89	P
40	XNG	0.50	11	TMA	1087.50	258.31	7.60	P
41	YEN Empowered	0.50	11	TMA	1266.67	33.33	0.71	n/a
42	YET's Wear	0.53	13	BM	789.08	-246.81	-2.47	P
43	YOU	0.17	12	M	0.00	-242.41	-6.55	P
44	Young Legends	0.44	9	BL	400.00	59.00	0.84	n/a
45	YUVA	0.53	15	BM	1153.00	294.11	12.79	P

Note: Team size represents average team size during the program; Field refers to field of study: BM is business management; BL is business languages; FM is financial management; M is management and TMA is trade management Asia. P/S indicates product or service.

Table A2: Robustness to inclusion of control variables

	Log Sales	Positive profits	Profits/share	Log Sales	Positive profits	Profits/share	Log Sales	Positive profits	Profits/share
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Share female	5.075**	3.403**	21.308	5.517***	4.045**	25.940*	5.295**	4.500**	30.226
	(1.999)	(1.567)	(15.178)	(1.804)	(1.745)	(14.521)	(1.998)	(1.884)	(18.855)
Share female squared	-4.851***	-3.159**	-23.633*	-3.916**	-3.199**	-14.741	-5.254***	-3.751**	-30.596*
	(1.744)	(1.416)	(12.826)	(1.784)	(1.553)	(12.737)	(1.778)	(1.663)	(16.965)
Mathematics grade	-0.021	0.233	4.425**						
	(0.169)	(0.191)	(2.176)						
Business knowledge	0.178	-0.349	-2.096						
	(0.428)	(0.323)	(3.593)						
Management knowledge	-0.113	-0.119	-1.833						
	(0.316)	(0.270)	(3.619)						
Business Management				-0.168	0.229*	7.142**			
				(0.186)	(1.680)	(2.803)			
Trade Management Asia				0.012	0.180**	5.782**			
				(0.177)	(2.070)	(2.362)			
Business Languages				-0.972***	0.365	-0.389			
				(0.324)	(0.740)	(4.154)			
Agreeableness							0.075	0.082	1.944
							(0.214)	(0.139)	(1.795)
Conscientiousness							0.038	0.033	-0.122
							(0.191)	(0.129)	(1.576)
Extroversion							0.057	0.137	-0.626
							(0.131)	(0.157)	(1.167)
Neuroticism							0.159	0.218	1.540
							(0.201)	(0.161)	(2.163)
Openness to experience							-0.292	-0.213	-3.234
							(0.231)	(0.201)	(1.982)

Note: Number of teams equal 43. Standard errors are given in parentheses. ***/**/* denotes significance at the 1%/5%/10%-level.

Table A3: Effect of share of women on team performance (excluding Financial Management)

	Sales (ln)	Positive profits (mean)	Profits/share (mean)	Profits/share (median)
	(1)	(2)	(3)	(4)
Share female	4.829**	4.074**	31.873**	38.382***
	(1.808)	(1.617)	(15.164)	(13.264)
Share female squared	-4.633***	-3.362**	-28.048**	-34.173***
	(1.589)	(1.445)	(13.109)	(11.458)

Note: Number of teams equal 41. Standard errors are given in parentheses. ***/**/* denotes significance at the 1%/5%/10%-level.