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 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, conditional on their gender, randomly to teams. We find that teams with an equal gender mix perform better than male-dominated and female-dominated teams in terms of sales, profits and earnings per share. We explore various mechanisms and find that mutual monitoring is more intense in mixed teams than in homogeneous teams. This partially explains the better results of mixed teams.

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

1 Introduction

This paper reports about a field experiment conducted to investigate the effect of team diversity on team performance. Many decisions in organizations are nowadays assigned to teams, not to individuals.[1] 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

*This version: April 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, Robert Dur and seminar participants in Amsterdam, Eindhoven, Lisbon, London, Maastricht and Uppsala 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.

[1] As Hamilton et al. (2003) put it: "During the past 30 years the use of teams has become a mainstay for the organization of work" (p. 465).

a team is its diversity. While there are theoretical results about the effect of team diversity on team productivity (Kremer, 1993; Prat, 2002), the empirical evidence is thin (Hamilton et al., 2003). Our study examines one particular dimension of team diversity, i.e. gender diversity.

The teams in our field experiment are business teams, and the performance of these teams is measured in terms of sales, profits and earnings per share. This links our study to the much discussed underrepresentation of women 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 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 a firm's corporate identity and brand image, better stakeholder relationships, higher employee motivation, higher customer satisfaction, more complementarities and better mutual learning (Desvaux et al., 2007).[2] Moreover, from a labor supply perspective, increases in the share of females in higher education and the labor force have strengthened the policy relevance of more gender diversity in management boards. Some countries (Norway, The Netherlands, Spain, France, Iceland) even enforce or are planning to enforce a higher representation of women by setting minimum quota of board seats for female directors.

Despite the pleas and implemented policies to reach 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 more equal representation of women does not have a positive impact on performance and may even be harmful.

Adams and Ferreira (2009) analyze data with characteristics of firms and their 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 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

[2] In a laboratory experiment, Niederle and Vesterlund (2007) find that men often choose to compete even if they perform poorly while 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.

(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.[3]

Finally, Apesteguia et al. (2011) 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.[4]

Our study contributes to this recent literature by providing fresh evidence on the impact of the share of women on the financial performance of business teams from a field experiment. In this experiment we manipulated the share of women in business teams and assigned participants - conditional on their gender - randomly to teams. Teams were started up at the same time and under equal circumstances. This method should address concerns regarding issues of self-selection, omitted variable bias and reversed causality.

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

[3]Ahern and Dittmar emphasize that the gender mix in a board has no impact on firm 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.

[4]Related is also Hansen et al. (2006) who measure 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 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).

company with a profit objective and shareholders for the duration of one year. Students face strong incentives, both individually and as a team, to perform a substantial and truly joint task of setting up and running a company with the objective of maximizing profit and shareholder value. The size of the team is realistic: the average European board is composed of 11.7 people (cf. Woods, 2010). These factors contribute to the external validity of our results. Based on this, one can argue that our study is informative about the effect of the gender composition of corporate boards on business performance. However, there are also noticeable differences between the business teams in our study and corporate boards. Some of these differences are interesting because they may become realistic in the future. The potentially realistic distinct features are the occurrence of teams with a majority of female members, the availability of male and female potential board members of the same caliber and the lack of a (predominantly) male selection committee. However, the external validity of our study is decreased due to the newness of the companies studied, the inevitable random composition of the team and, most importantly the age and experience profiles of the team members leading the student companies.

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 a team's 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 occurs more often in mixed gender teams than in more homogeneous teams and that more intense monitoring has a positive impact on company performance.

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 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,

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

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

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 professor, 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 is defined.[5] At the same time, positions are defined and allocated over 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 non-management positions. After half a (program) year, roles are switched and the management team positions are reallocated among the non-managing part of the team.[6] 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 compulsory, 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, financial management, trade management Asia and business languages. The experiment

[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.

[6]The relationship between the gender composition in the total team and the division of team roles across genders is discussed below.

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 did not know to which team they were assigned to at that stage).

Figure 1 shows the frequency distribution of the share of women per team at baseline. This shows that there is substantial gender variation across teams. The share of females varies from 0.1 to 1.0.

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 which required to stick to the imposed team assignment. 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.

During the year, 104 students (19%) dropped out.[7] 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 a team's 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 the dropout decision of students is not affected by the gender composition of

[7] High dropout rates from the first year of tertiary education are common in the Netherlands where the admission of students based on grades or previous achievements is not allowed.

their team.[8]

Student teams elected a management team, including the CEO and CFO, twice, once for the duration of the first half of the program, and once for the second half. In theory, two gender homogeneous management teams could be elected from a mixed team. This could then contaminate our design. This turns out not to be the case. Regressing the share of women in the management team on the share of women in the team returns a coefficient of 0.983 (s.e. 0.069) for the first half of the program and of 0.997 (s.e. 0.038) for the second half of the program. Similarly, we relate the gender diversity index of the management team to the gender diversity index of the total business team, where the gender diversity index is defined as the share of women times the share of men. The regression coefficients are 1.181 (s.e. 0.204) for the first half of the program and 1.075 (s.e. 0.152) for the second half. 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. This implies that our estimates of the effect on performance of the gender composition of the team can be interpreted as the effect of the gender composition of the managing part of the team. The division of team roles by the teams themselves does not contaminate our design.

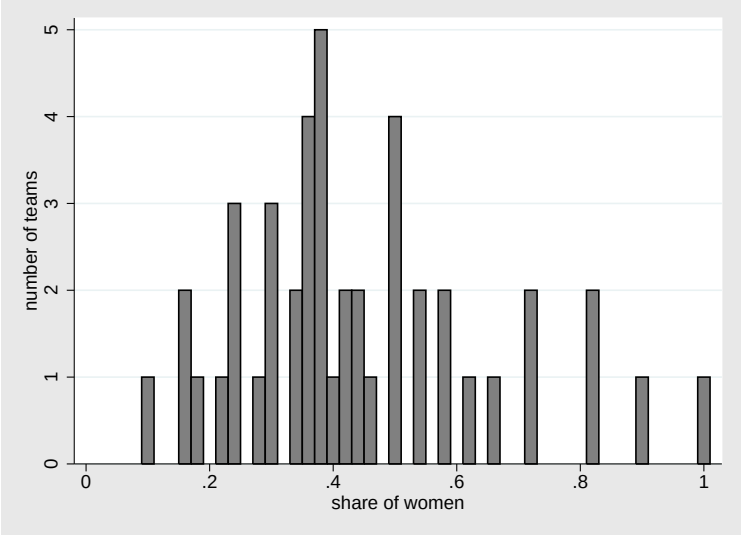

Figure 1. Frequency of share of women per team

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 'fired' by team members. The decision

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

to fire someone requires that two thirds of the team agrees, together with the consent of the professor. In the case of being fired, 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 fired 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. Firing team members is not uncommon and thus a credible threat: 50% of the teams has experienced at least one fire and the average number is 0.74. The occurrence and number of fires are not related to the team's business 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 GPA (20%). Both individual and team performance determine the grade. The individual component has no predetermined weight in the total grade. 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 of 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.

Specific team incentives are further provided by the formal competition among teams. At the end of the year, 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 bowl, 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 (s.d. 3.8) on the program. This is a high number given that the program counts for twenty percent of the students' curriculum; the average actual number of hours students in Dutch vocational schools spend on their education is 32 hours per week (Allen et al., 2009). The positive correlation between grades and sales/profit, 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 towards the business outcomes we measure.

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. Response rates are slightly higher for women than for men. The surveys provide background information about individuals and teams. This information is required to assess whether team assignment was random, 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, 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. Twenty percent are non-Dutch (exchange) students and nearly half of the Dutch students has at least one parent not born in the Netherlands.[9]

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, 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 (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).[10]

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

[10] These measures are taken using the so-called Escan (Driessen, 2005), a validated self-assessment test

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's 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.

Finally, all three surveys include self-assessments of the 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 be able to explain possible differences in performance between teams based on their 'modus operandi' and the team processes they developed and employed, 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 with the professor and the result, 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 these scores are related to teams' performance on the one hand and to the teams' gender composition on

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.

the other.

The outcome variables in our analyses are measures of teams' business performance. Information about business performance was retrieved from the annual reports that we managed to obtain from 43 out of 45 teams. 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 the profit and loss statements 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 than the exact level of profit reported. We therefore add the binary indicator for positive profits to the 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 of 20 euro initially, profits vary between -15 and +15 euros.[11] The final column in the table shows that the three profit measures are positively and significantly correlated with sales.

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.

3.2 Randomization

In this subsection we first examine whether students are randomly assigned to teams of different gender composition, conditional on their gender. 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

[11] The mean number of shares issued is 52, the standard deviation is 21.5. The minimum and maximum numbers of shares sold amount to 15 and 100. On average, half of the shares are bought by the team members themselves. We have no exact information on the identity of the companies' shareholders.

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 team's gender diversity index (the share of women times the share of men). No systematic differences are observed between the characteristics of males and females in 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.

The share of women and gender diversity of teams possibly correlate with other team characteristics. Table 5 reports these correlations. The first two columns show that the share of women correlates 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. Moreover, in line with expectations, the average degree of risk aversion is significantly higher in teams with more female members (e.g. Dohmen et al., 2011). 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.[12]

The last two columns of Table 5 indicate that there are no significant correlations between a team's gender diversity index and other team characteristics measured at baseline. 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 if and to what extent these gender composition related differences between teams at baseline confound the results.

[12]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).

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

		Share of women				Diversity index			
		Males		Females		Males		Females	
Personal	Age	0.006	(0.005)	0.002	(0.005)	−0.001	(0.002)	−0.002	(0.002)
characteristics	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)
Traits	Need for power	0.005	(0.006)	−0.009	(0.009)	0.000	(0.002)	0.003	(0.003)
	Perseverance	0.003	(0.005)	−0.001	(0.007)	0.001	(0.002)	−0.003*	(0.002)
	Need for achievement	−0.001	(0.005)	0.006	(0.007)	0.002	(0.001)	−0.002	(0.002)
	Self-efficacy	0.007	(0.006)	−0.010	(0.010)	0.001	(0.002)	0.002	(0.003)
	Risk aversion	0.003	(0.006)	0.016	(0.010)	−0.000	(0.002)	−0.004	(0.003)
	Social orientation	0.006	(0.006)	0.000	(0.006)	0.002	(0.002)	0.000	(0.002)

Note: Each coefficient 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)
Entrepreneurial traits (average)				
Need for power	−0.040	(0.074)	0.024	(0.024)
Perseverance	0.107*	(0.058)	−0.022	(0.024)
Need for achievement	0.036	(0.060)	−0.007	(0.016)
Self-efficacy	−0.050	(0.075)	0.025	(0.023)
Risk aversion	0.162**	(0.071)	−0.042*	(0.024)
Social orientation	0.066	(0.046)	0.003	(0.016)
Team size	−0.021	(0.014)	0.011	(0.009)

Note: Based on information from 45 teams. Standard errors in parentheses. ***/**/* denotes significance at the 1%/5%/10%-level.

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 0.55. Beyond that percentage of female team members, further increases in the share of women tend to reduce team 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 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

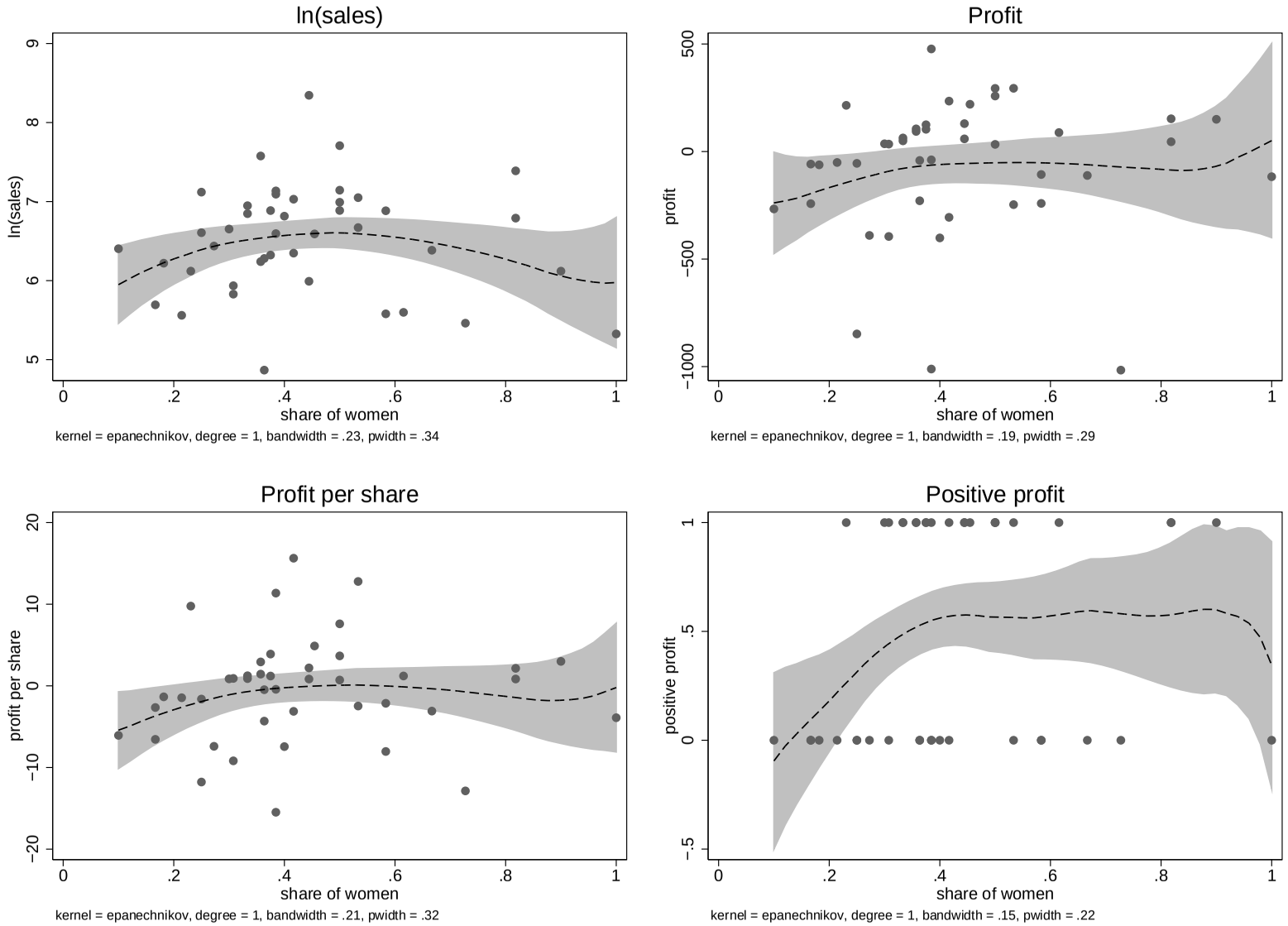

Figure 2. Effect of share of women on team performance

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: N= 43 teams. 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.

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.[13] 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.

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 a team's 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.[14]

4.2 Robustness

The results presented in the previous subsection show an inverse u-shape effect of the gender composition of teams 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

[13] 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.

[14] In Subsection 2.3 we mentioned that six teams were selected to present their results to a jury of entrepreneurs. Regressing an indicator for belonging to this top-6 on the share of women and the share of women squared returns coefficients of 1.03 (s.e. 0.83) and -0.77 (s.e. 0.77). While these estimates lack precision, the pattern of these coefficients confirms our other findings. The six selected teams have shares of women of 0.25, 0.42, 0.44, 0.55, 0.62 and 0.82. The winning team consists of 42 percent women.

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) 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 and risk aversion. 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 (iii) the average scores on the big five personality characteristics including agreeableness and (iv) average scores on a set of entrepreneurial traits, including risk aversion. 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.[15]

4.3 Mechanisms

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 prac-

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

tices. In support of the findings by Adams and Ferreira (2009), Gul et al. (2011) find that gender diverse boards improve the quality of public disclosure through better monitoring. 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.[16]

These studies suggest that mechanisms such as complementarities, learning, monitoring and conflicts/friendships are possible explanations for our findings. In addition we considered: effort, risk aversion (e.g. Dohmen et al., 2011), delegation of decision making and the type of product or service. For each of these factors we examined whether it is significantly related to team gender diversity and to their business performance. Both conditions should hold for a factor to possibly explain the relation between a team's gender composition and its performance. We find that this is only the case for monitoring. None of the other factors finds any support in our data.[17]

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. Cronbach's alpha for monitoring equals 0.88.

Table 7 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 we see no significant impact of gender composition on the level of monitoring. However, in the second follow-up there appears to be a significant relation between gender composition and the level of monitoring. The relationship 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

[16]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.

[17]In Appendix B we describe how each of the other factors has been operationalized.

Table 7. Monitoring and share of women

	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's 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. Standard errors are given in parentheses. ***/**/* denotes significance at the 1%/5%/10%-level.

in teams is more intense in gender diverse teams than in more homogeneous teams. This is in line with the results from the recent studies by Adams and Ferreira (2009) and Gul et al. (2011). A more equal representation of females on the board adds to promoting better board attendance, whereas a mixed team demands greater accountability from managers for poor performance (Adams and Ferreira, 2009).

Monitoring is positively related to performance. The correlations between monitoring (measured as the change between the first and second follow-up) and the various business outcomes are: 0.117 (s.e. 0.069) for (ln) sales, 0.002 (s.e. 0.058) for positive profits and 0.552 (s.e. 0.586) for profits per share. The positive effect of mutual monitoring on team performance is in line with the effectiveness of monitoring to mitigate free-riding to increase productivity (Hamilton et al., 2003). Mutual monitoring may be particularly effective in the case of large teams – of people who don't know each other – (Knez and Simester, 2001; Hamilton et al., 2003) or for firms with weak corporate governance, thereby suggesting that gender-diverse boards could act as a substitute mechanism for corporate governance (Adams and Ferreira, 2009; Gul et al., 2011). It is conceivable that the teams in our experiment have not put sophisticated corporate governance mechanisms into place. 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.

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.

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 being fired and excluded from the program and lead to study delay, which in turn can have serious financial consequences for the students. Performance in the program determines 20 percent of students' GPA. Moreover, teams are motivated by the ownership of shares and 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 case in the group of people qualified for a position in a board of directors: 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. This kind of evidence may provide incentives to firms to diversify their management teams. It would therefore not be surprising if results like those reported in this study have in the end a larger impact on the position of women in boards of directors than enforcing policies that ignore supply constraints, such as the quota 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.

Allen, J., van Breugel, G., Coenen, J., Fouarge, D., Meng, C., Ramaekers, G., van Dijk, J., and Venhorst, V. (2009). Afgestudeerden van het hbo tijdens een crisis: geen verloren generatie. *Den Haag, HBO-raad*.

Apesteguia, J., Azmat, G., and Iriberri, N. (2011). The Impact of Gender Composition on Team Performance and Decision-Making: Evidence from the Field. *Management Science*, forthcoming.

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.

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., and Wagner, G. G. (2011). Individual risk attitudes: Measurement, determinants and behavioral consequences. *Journal of the European Economic Association*, forthcoming.

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.

Gul, F., Srinidhi, B., and Ng, A. (2011). Does board gender diversity improve the informativeness of stock prices? *Journal of Accounting and Economics*, forthcoming.

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.

Knez, M. and Simester, D. (2001). Firm-wide incentives and mutual monitoring at Continental Airlines. *Journal of Labor Economics*, 19:743–772.

Kremer, M. (1993). The O-ring theory of economic development. *Quarterly Journal of Economics*, 108:551–575.

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.

Niederle, M. and Vesterlund, L. (2007). Do women shy away from competition? Do men compete too much? *Quarterly Journal of Economics*, 122(3):1067–1101.

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.

Prat, A. (2002). Should a team be homogeneous? *European Economic Review*, 46(7):1187–1207.

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 percent. *HR Magazine, United Kingdom*.

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.

Appendix A

Table A1: Team characteristics

#	Name	Female (share)	Team size	Sales (euro)	Profits (euro)	Profit/ share	Description of product/service
1	A-Card	0.25	14	1236.15	-848.05	-11.78	Discount card for Amsterdam nightlife
2	A'dam Gadgets	0.36	11	534.12	-41.40	-0.47	USB hot plate for coffee, tea, etc.
3	Appie	0.90	9	454.75	149.86	3.00	Apple-shaped box to preserve apples
4	Aqua de Coctail	0.42	11	1130.47	-305.94	-3.12	Comprehensive cocktail shaker set
5	ArtEco Bags	0.40	9	912.00	-401.69	-7.44	Durable give-away bag for clothes stores
6	BubbleMania	0.18	10	503.00	-61.79	-1.34	Multifunctional protective key chain
7	D'Wine	0.25	8	740.00	-55.00	-1.62	Bottles of wine
8	Eastern Green	0.36	13	513.00	105.51	2.93	Engravable text bean that grows a plant
9	Escapade Inc	0.67	9	592.55	-111.30	-3.09	Tube clip for sealing food, toiletry, etc.
10	eyeBMA	0.38	14	557.50	124.66	3.90	Package with easy-to-use eye shadow
11	Firefly	0.50	11	2225.65	293.62	3.67	Ascending fire lantern for celebrations
12	Fl!pthat	0.23	11	455.00	214.88	9.77	Redecorating already existing websites
13	Ginger	0.58	11	976.50	-106.81	-2.14	Multifunctional solar energy charger
14	Himitsu	0.30	10	775.00	36.00	0.86	n/a
15	I-Care	0.38	14	1204.45	477.15	11.36	Beauty products with Dead Sea minerals
16	iJoy	0.36	12	1952.85	93.56	1.44	Wristband with USB storage capacity
17	I-Juice	0.38	13	1255.38	-38.54	-0.42	Pocket-size lightweight mobile charger
18	IMSC	0.27	10	625.00	-390.00	-7.41	n/a
19	iShield	0.44	10	4209.49	129.76	2.20	Invisible protective shield for iPhones
20	KISBag	1.00	9	205.48	-117.02	-3.90	Tiny foldable bag to replace plastic bags
21	Laservibes	0.36	11	130.00	-228.90	-4.32	Organizing lasershows for companies
22	Mengelmoes	0.33	10	941.50	63.14	1.24	Easy-to-wear telephone charger device
23	My-Buddy	0.17	10	297.00	-58.33	-2.65	USB doll for kids that reflects emoticons
24	Nine2Five	0.73	12	235.45	-1016.36	-12.87	USB hot plate for coffee, tea, etc.
25	Picture Perfect	0.21	14	260.09	-50.87	-1.45	Customized t-shirts for men and women
26	Pietje Plu	0.73	11	n/a	n/a	n/a	Trendy umbrellas
27	Pocket Memory	0.38	15	978.94	103.46	1.20	Business cards with USB storage capacity
28	Pro'Lux	0.31	13	378.25	-394.90	-9.18	Promotional gifts with USB storage capacity
29	Qwinlok	0.31	12	340.00	34.61	0.91	Boxer shorts for female adolescents
30	Reflection	0.82	11	889.51	45.43	0.84	Cosmetics mirror including mascara clip
31	SAME	0.82	9	1618.35	152.37	2.15	Comfortable unisex earwarmer
32	Sappho	0.50	7	980.00	n/a	n/a	n/a
33	Sharity	0.58	11	265.00	-241.12	-8.04	Necklace with peace sign for teenagers
34	ShoeTattoo	0.62	13	270.00	88.32	1.21	Shoe customization by graphic artists
35	Student Promotion	0.42	13	571.32	234.54	15.64	Promotional activities for companies
36	StuPill	0.38	13	731.33	-1011.33	-15.48	Comfortable Indonesian anti-RSI pillow
37	Test-a-Holic	0.45	11	728.45	219.77	4.88	Alcohol breath tester for nightlife
38	We-Do Solutions	0.10	9	604.00	-266.82	-6.06	Multifunctional trendy key chain
39	We 'R U	0.33	13	1041.11	49.77	0.89	Compact wallet in several colors
40	XNG	0.50	11	1087.50	258.31	7.60	T-shirts of "Chicks on Kicks" community
41	YEN Empowered	0.50	11	1266.67	33.33	0.71	n/a
42	YET's Wear	0.53	13	789.08	-246.81	-2.47	Customized t-shirts of own YET-brand
43	YOU	0.17	12	0.00	-242.41	-6.55	Hotel door hanger to store keys, money, etc.
44	Young Legends	0.44	9	400.00	59.00	0.84	n/a
45	YUVA	0.53	15	1153.00	294.11	12.79	Engravable grain of rice in a glass covering

Table A2: Robustness to inclusion of control variables

	(ln) sales	Positive profits	Profits/share	(ln) sales	Positive profits	Profits/share
	(1)	(2)	(3)	(4)	(5)	(6)
Share female	5.075**	3.403**	21.308	5.517***	4.045**	25.940*
	(1.999)	(1.567)	(15.178)	(1.804)	(1.745)	(14.521)
Share female squared	-4.851***	-3.159**	-23.633*	-3.916**	-3.199**	-14.741
	(1.744)	(1.416)	(12.826)	(1.784)	(1.553)	(12.737)
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)

Note: Based on information from 43 teams. Standard errors are given in parentheses. ***/**/* denotes significance at the 1%/5%/10%-level.

Table A2: Robustness to inclusion of control variables (continued)

	(ln) sales	Positive profits	Profits/share	(ln) sales	Positive profits	Profits/share
	(7)	(8)	(9)	(10)	(11)	(12)
Share female	5.295**	4.500**	30.226	5.471**	5.385**	32.238**
	(1.998)	(1.884)	(18.855)	(2.074)	(2.255)	(13.436)
Share female squared	-5.254***	-3.751**	-30.596*	-5.304***	-4.615**	-28.413**
	(1.778)	(1.663)	(16.965)	(1.779)	(1.913)	(12.091)
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)			
Need for power				-0.431	-0.301	-3.896*
				(0.256)	(0.272)	(2.012)
Perseverance				-0.253	0.268	-1.064
				(0.317)	(0.250)	(2.216)
Need for achievement				0.232	0.528**	6.010***
				(0.259)	(0.243)	(2.041)
Self-efficacy				0.268	-0.560**	-3.579
				(0.226)	(0.235)	(2.555)
Risk aversion				0.309	-0.159	-1.258
				(0.218)	(0.264)	(2.185)
Social orientation				0.122	-0.164	-0.216
				(0.187)	(0.196)	(1.770)

Note: Based on information from 43 teams. Standard errors are given in parentheses. ***/**/* denotes significance at the 1%/5%/10%-level.

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

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

Note: Based on information from 41 teams. Standard errors are given in parentheses. ***/**/* denotes significance at the 1%/5%/10%-level.

Appendix B: Mechanisms

Complementarities

Men and women in mixed teams may complement each others' skills and knowledge. To assess this, we first standardized the various skill and knowledge dimensions. We then computed for each skill and knowledge dimension the maximum in a team. Subsequently, we computed 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. Neither do these constructed minimums have a significant impact on business performance.

Learning

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 (see Table 5) – or due to differential team processes that may be unobserved. The team average increases in skill/knowledge levels turn out to be unrelated to teams' gender composition. There is thus not more or less learning in gender diverse teams than in other teams.

Conflicts, friendships and team 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. Teams' average conflict score is also unrelated to business outcomes.

The second follow-up survey 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 and business performance.

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. This variable is unrelated to gender diversity and business performance.

Effort

We asked students in the third survey how many hours per week they devoted to the program on average. We also asked students in the third survey to rate how much each of the students within their team contributed to the student company. Teams' averages of these two effort measures are not significantly related to teams' gender diversity or business performance.

Risk aversion

The first survey asked respondents to rate their risk attitudes on a scale from 1 to 5. Neither teams' average risk attitude nor the standard deviation or the maximum or minimum are signficantly related to teams' gender diversity or to business performance in regressions that also include team size.

Decision making

In the final 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 or business performance.

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. No systematic differences were observed in the market orientation of teams and their gender mix.