Received: 20 April 2021 | Revised: 23 March 2022 | Accepted: 17 April 2022

DOI: 10.1111/ssqu.13164

ORIGINAL ARTICLE

Differences in team performance: Gender versus ability

Silvio Städter[1] | **Andreas Roider**[2] | **Michael Dowling**[3] | **Roland Helm**[3]

[1]Department of Economics, University of Regensburg, Regensburg, Germany

[2]Department of Economics, University of Regensburg, Germany, and CEPR, CESifo, and IZA, Germany

[3]Department of Business Administration, University of Regensburg, Germany

Correspondence
Silvio Städter, Department of Economics, University of Regensburg, Universitätsstr. 31, D-93040 Regensburg, Germany.
Email: silvio.staedter@ur.de

We thank Suzanne Weinberger and Ingrid Mittermeier for very valuable research assistance. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sector. Declarations of interest: none.

Abstract

We empirically study the relationship between the gender composition and the performance of teams. In our setting, teams of students are incentivized to develop business plans as part of a compulsory course in a business administration program. At first sight, and in line with earlier empirical literature, the gender composition seems to matter. Our article differs in that we have access to a measure of task-specific individual ability. When controlling for ability, the relation between gender and team performance vanishes, highlighting the importance of controlling for other characteristics in empirical and experimental studies of gender effects.

Objective: While various studies document a relationship between the gender composition and the performance of teams, we ask whether such findings may depend on a lack of controlling for other dimensions of heterogeneity among team members.

Methods: In our setting, teams of students are incentivized to develop business plans as part of a compulsory course in a business administration program. Our article differs from the literature in that we have access to a measure of task-specific individual ability. Using ordinary least squares regressions, controlling for year and team size fixed effects, and conducting various robustness checks, we study the association between team performance and team characteristics.

Results: At first sight, and in line with earlier empirical literature, the gender composition seems to matter for team performance. However, when controlling for the individual abilities of team members, the relation between the gender composition and team performance vanishes.

Conclusion: Our study highlights the importance of controlling for other characteristics in empirical and experimental studies of gender effects, which in our context leads them to disappear.

KEYWORDS
ability differences, gender effects, team composition, team performance

Over the course of the last decades, there has been a move toward more teamwork in many firms and organizations. A case in point is the gaining popularity of "agile" organizations where self-managed teams fluidly form to tackle clearly defined tasks with a strong focus on getting to results quickly. Initially applied to software development, such team-based organizational forms have recently spread across a broad range of industries and functions (see, e.g., Rigby, Sutherland, and Noble 2018; Rigby, Sutherland, and Takeuchi 2016).

Given the increasing relevance of teamwork, it seems to be important to gain a better understanding of performance differences across teams. In this respect, team composition (i.e., within-team heterogeneity in characteristics such as experience, ability, or socio-demographic characteristics) has been argued to play an important role (see, e.g., Hamilton, Nickerson, and Owan 2003).[1]

In particular, the effects of a team's gender composition have received growing attention. One reason for this is the emerging evidence on gender differences in individual preferences and behavior (for overviews, see, e.g., Eckel and Grossman 2008, or Croson and Gneezy 2009). Moreover, the underrepresentation of women in many decision-making bodies has triggered important policy discussions. For example, gender quotas in corporate boards have been mandated in countries such as Norway, France, Germany, Italy, and Great Britain, with varying degrees of sanctions for non-compliers across countries.

The empirical economics literature on gender and teams is still sparse (see, e.g., Kuhn and Villeval 2015). One focus of the literature has been to assess how female representation and gender quotas in the board room affect corporate performance. The results are, however, mixed. For example, Adams and Ferreira (2009) consider a sample of U.S. firms and find the average effect of gender diversity on performance (as measured by the return on assets and Tobin's Q) to be negative. More recently, Ferrari et al. (2021) studied the introduction of gender quotas in Italy but found no significant effect on corporate performance.[2]

Studies with field data face, however, various challenges. For example, often only relatively noisy measures of team performance (such as corporate profits) are available. Moreover, in many real-world settings, teams differ in their size as well as in the tasks, decisions, and circumstances they face, which makes them difficult to compare. These issues are ameliorated in articles on the performance of teams in the relatively more structured context of business-simulation games. Overall, these studies find that the gender composition matters for teams' performance, but again the details of the results differ. For example, Hoogendoorn, Oosterbeek, and Van Praag (2013) conduct a field experiment with teams of 12 students each, who are enrolled in a business program. As part of their curriculum, each team forms a start-up, which is liquidated after 1 year. Hoogendoorn, Oosterbeek, and Van Praag (2013) compare teams' profits and sales, and they find that performance is weakly increasing in the share of female team members.[3] Lamiraud and Vranceanu (2018) conduct a field experiment where teams of five students simulate corporate boards. They find that teams with a majority of female members perform best.[4] Finally, Apesteguia, Azmat, and Iriberri (2012) study non-experimental field data from an online business simulation game. There, each observation corresponds to three students that select into a team and play the role of general manager of a

[1] These questions have also received attention in the psychology literature (see, e.g., Amason, Shrader, and Tompson 2006; Edwards et al. 2006).

[2] Ellison and Mullin (2014) study the effects of gender composition at the office level.

[3] They also investigate potential channels, but do not find evidence for complementarities, learning, monitoring, or the reduction of conflicts.

[4] For laboratory experiments on the effects of the gender composition on team performance, see, for example, Ivanova-Stenzel and Kübler (2011).

company that competes against four other simulated firms. They find no systematic relationship between gender composition and team performance.

One potential reason for the mixed empirical results on the relationship between gender composition and team performance could be a lack of information about other characteristics of team members, such as heterogeneity with respect to ability for the task at hand. For example, Apesteguia, Azmat, and Iriberri (2012) can only control for the university where a given team member studies. Arguably, this is, however, a rather coarse measure of individual task-specific ability. Hence, it seems that in their study one cannot rule out that gender effects are masked by heterogeneity in other characteristics. Arguably, a similar issue arises in experimental studies of gender effects if experimental subjects of various genders also systematically differ in other characteristics that correlate with gender. In this case, randomly assigning subjects of different genders to various experimental groups would not ensure that the various groups only differ in their gender composition. They would most likely also differ in those other characteristics. Thus, in such a situation, a documented gender effect might be driven by differences in these other characteristics. For example, while Hoogendoorn, Oosterbeek, and Van Praag (2013) find team performance to be weakly increasing in the share of female team members, they only have access to participants' overall grade-point averages. Hence, they cannot control for task-specific ability, which, in principle, might drive their result.

Our article tries to highlight this issue. Similar to Apesteguia, Azmat, and Iriberri (2012), we study team performance in a business simulation task, but have access to individual measures of the task-specific (ex-ante) ability of team members. At first glance (i.e., without controlling for team members' abilities), team performance is increasing in female representation, which would be in line with Hoogendoorn, Oosterbeek, and Van Praag (2013). However, when taking ability of team members into account, this association vanishes, while ability exhibits a strong, significant correlation with team performance.

CONTEXT AND DATA

Our empirical analysis of team performance relies on data from a compulsory one-semester course on "Management and Entrepreneurship" in a bachelor program on business administration at a university in Germany.[5] The course is offered on a yearly basis, and we have access to data on students' performance in the years 2011–2015, that is, for five cohorts of a total of 1670 students (see Table 1, which provides summary statistics of our data). Students typically take this course in their fourth semester, and the final grades they obtain count toward their individual grade point averages (with a weight of 6 out of a total of 180 credit points). Hence, students have an incentive to perform well.

The course consists of two parts: weekly lectures (where a semester typically lasts 14 weeks) and the preparation of business plans by groups of students. A student's final grade is an equally weighted average of (i) the respective team's performance in the business plan project, which takes place over the course of the semester, and (ii) the individual performance in a final exam, which takes place in the weeks immediately following the end of term.[6] Students form teams right at the beginning of the semester before the course commences, and team composition remains fixed. Note that when teams are formed students had not been exposed to business plan tasks at earlier stages of their studies, and any prior individual grades are only privately known by the respective student. In general, teams consist of seven members, and on average 57 percent of team members are female.[7]

[5] Enrolling in a bachelor program in business administration is quite common in Germany, and such programs are offered by all major universities. According to the German Federal Statistical Office, in the winter term 2018/2019, a total of 240,572 students were enrolled in a program in business administration, which is by far the biggest group among all fields of studies (see https://de.statista.com/themen/56/studenten/).

[6] The final exam consists of 30 multiple-choice questions on various theories of management and entrepreneurship. The grade on the final exam depends on the respective student's individual performance only, that is, there is no teamwork component. In the empirical analysis, we do not use the individual grades on the final exam because it takes place after the business plan project has been concluded. Hence, participating in the business plan project might affect students' enthusiasm to study for the final exam. As a result, team performance might influence individual performance in the final exam.

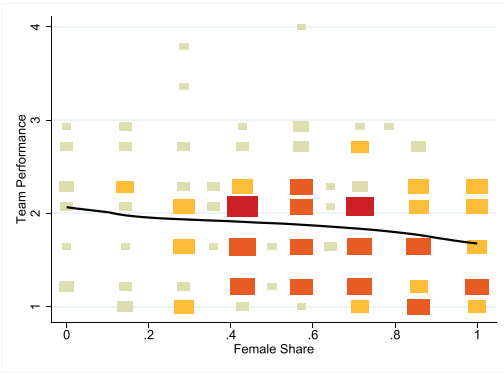
[7] For non-systematic reasons, a small number of teams consists of six or eight students (see Table 1 for details).

TABLE 1 Summary statistics

Year	Participants (Number)	Teams (Number)	Female Share (Mean)	Female Share (SD)	Team Per-formance (Mean)	Team Per-formance (SD)	Ability (Mean)	Ability (SD)
2011	298	44	0.58	0.27	1.80	0.53	3.60	0.93
2012	319	46	0.53	0.28	1.82	0.48	3.41	1.18
2013	378	54	0.54	0.30	1.84	0.64	3.11	1.06
2014	337	48	0.61	0.28	1.99	0.57	3.14	1.01
2015	338	48	0.58	0.26	1.89	0.59	3.07	1.07
Overall	1670	240	0.57	0.28	1.87	0.57	3.25	1.07

Note: Teams generally consist of seven students. In each of 2011, 2014, and 2015, there are, however, two teams consisting of eight students. In 2011, 2012, and 2014, there are 12, 3, and 1 team(s) consisting of six students, respectively. Standard deviations (SD) are calculated across teams. Note that both team performance and ability are measured on scales that run from 1.0 (excellent) to 5.0 (failed), that is, lower values indicate better results.

(a) Team Performance and Female Share



(b) Team Ability (Mean) and Female Share

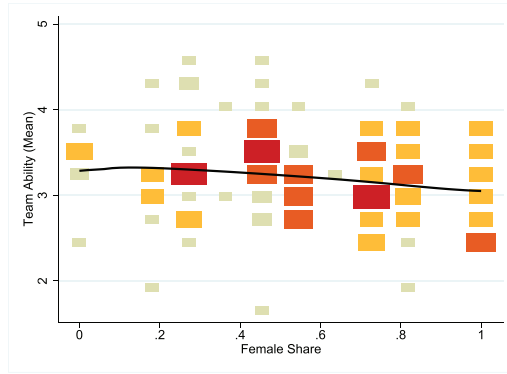


FIGURE 1 Raw data

Note: Higher values of team performance indicate worse performance, where this variable varies between 1.0 (best) and 5.0 (worst). Observations for all 240 teams in panel (a), and observations for all 152 teams without missing individual ability information in panel (b). Each rectangle corresponds to teams with a given combination of female share and team performance (a) and team ability (mean) (b), respectively. The color of a given rectangle indicates the respective number of observations in panel (a) (in panel (b)): red = 10–12 observations (7–8 observations), orange = 7–9 observations (5–6 observations), yellow = 4–6 observations (3–4 observations), green = 1–3 observations (1–2 observations). The size of a given rectangle expresses differences with respect to the number of observations within these ranges. The black lines are based on kernel-weighted local polynomial smoothing (kernel = epanechnikov, degree = 1)

All students in a given team receive the same grade on the business plan project, and this joint grade serves as our measure of team performance. Importantly, note that in the German grading system, grades run from 1.0 (excellent) to 5.0 (failed). That is, a lower value of the variable team performance indicates better performance. Over the course of 10 weeks, each team has to prepare a distinct 10-page business plan for some real or fictitious company, which is freely chosen by the respective team. Business plans have to be structured into four parts: an executive summary, a description of the product or service, a discussion of its potential as well as of its likely competition, and plans for marketing and sales (plus, potentially, a two-page appendix providing further details). The business plan is supposed to contain only a rough sketch of financials. However, it should convince a potential investor of the plan's appeal, and it should lay out a preliminary marketing strategy. Hence, teams with a better grip of marketing and sales will presumably perform better.

As a measure of each student's individual *ex ante* ability, we use the respective student's individual grade from another closely related compulsory course in their program, which is called "Introduction to Marketing." Like in the case of team performance, this grade varies between 1.0 (excellent) and 5.0 (failed), that is, a lower grade indicates better performance. Students typically take the course "Introduction to Marketing" one semester prior to the business plan project in the "Management and Entrepreneurship" course. Given the business plan project's focus on developing a marketing strategy (as discussed above), the "Introduction to Marketing" grade might arguably be viewed as a signal about an important part of students' *ex ante* ability for the subsequent team task. For data privacy reasons, we do not have access to other grades or other socio-demographic characteristics.

RESULTS

In order to study the relationship between gender composition and team performance, we use observations at the team level. Panel (a) of Figure 1 displays the raw data on the relationship between team performance on the vertical axis and female share (i.e., the fraction of female team members) on the horizontal axis. In Panel (a), each rectangle corresponds to teams with a given combination of female share and team performance, where the color and size of the rectangle indicate the number of teams that exhibit

TABLE 2 Regression analysis of team performance

	(1)	(2)
Female share	−0.320**	−0.141
	(0.015)	(0.341)
Team ability (Mean)		0.339***
		(0.001)
Top share		1.082*
		(0.059)
Constant	1.937***	0.442
	(0.000)	(0.278)
Observations	240	152
Adjusted R^2	0.021	0.118
Team size FE	YES	YES
Year FE	YES	YES

Note: The table reports the results from OLS regressions with team performance as the dependent variable. Team performance and team ability are measured on grids that run from 1.0 (best) to 5.0 (failed). All regressions include year dummies as well as team-size dummies. *p*-Values are reported in parentheses, where *, **, and *** indicate statistical significance at the 1 percent, 5 percent, and 10 percent level, respectively.

the given combination (as explained in more detail in the note below Figure 1). The black line in Panel (a) is based on kernel-weighted local polynomial smoothing and suggests a slightly positive relationship between team performance and the share of female members on the team. That is, a higher female share is associated with a lower (i.e., better) grade on the business plan project.

In Table 2, we more systematically study the relationship between gender and team performance through regression analysis. Note that all regressions include year dummies in order to control for cohort effects as well as team-size dummies. Column (1) of Table 2 suggests that teams with a larger share of female members perform better (i.e., obtain a lower grade).

In a next step, we incorporate team members' ability into the analysis. To this end, we define the variable team ability (mean), which is the team-specific average of the individual ability measures. Note that some students postpone taking "Introduction to Marketing" to a later date, in which case no individual grade is available to us. This applies to 142 out of 1670 students (and in 88 out of 240 teams at least one member's grade is missing). In Column (2) of Table 2 and in panel (b) of Figure 1, we drop these teams from the analysis.

In particular, Figure 1b displays the joint distribution of female share and team ability (mean), where, analogous to panel (a), each rectangle corresponds to teams with a given combination of female share and team ability (mean). The two variables seem to be positively related, that is, the larger the share of females, the better, that is, lower, the value of team ability (mean).[8]

One could also hypothesize that, beyond the mean of within-team ability, team performance is also related to the share of high-ability members on the team. To this end, for each group, the variable top share measures the fraction of students who have one of the top-two grades, that is, ability measures of either 1.0 or 1.3.

Columns (2) of Table 2 reveals that team ability (mean) has a highly significant positive coefficient (implying that higher average ability of team members is associated with better team performance). However, the coefficient on the share of women in the team drops substantially and becomes insignificant.

[8] This is in line with earlier literature that documents that females often perform better than males in academic environments. There, it is argued that factors such as females' higher self-discipline and differing expectations with respect to future career choices might play an important role in this respect (see, e.g., Fortin, Oreopoulos, and Phipps 2015; Duckworth and Seligman 2006). In this sense, our individual "ability" measure (i.e., the grade on the "Introduction to Marketing" course) might not only capture innate ability for the topic at hand, but it might also reflect other (personality) attributes.

Perhaps surprisingly, the share of top students has a (weakly) significant, positive coefficient (i.e., a larger share of top students is associated with a worse grade on the team project, i.e, with lower team performance).[9]

Our main result of a lacking relationship between gender composition and team performance (as well as a highly significant relationship with team members' ability) is robust along a variety of dimensions (for details, see Table A1 in the Appendix): It holds when additionally controlling for the within-team standard deviation of ability (Column (1)), when controlling for the standard deviation but not the top share (Column (2)), and when not at all controlling for the within-teams dispersion of ability (Column (3)).[10] Perhaps interestingly, when controlling for the within-team standard deviation of ability in these robustness regressions, the coefficient on team ability (SD) is negative but only weakly significant which might suggest that more heterogeneity (i.e., both weak and strong students on the team) is associated with better team performance (i.e., a lower, and hence better, grade on the team task).[11]

CONCLUSION

Teamwork is steadily gaining importance in firms and organizations. In practice, teams vary substantially in their performance, and it has been argued that team composition might play an important role. In particular, various studies suggest that a team's gender composition impacts performance. In contrast to other articles in the literature, we have access to an arguably task-specific measure of team members' individual *ex ante* abilities. Without controlling for ability, there is a gender composition effect also in our context. However, when controlling for ability, the gender composition effect vanishes, while average task-specific ability correlates positively with team performance.

Of course, in the present setting, members are not randomly allocated to teams, which prevents a causal interpretation of our results. However, note that this is also the case in many empirical studies, for example, on the effects of legally mandated gender quotas in the board room. While the nature of our data potentially limits generalizability, our results suggest that controlling for personal characteristics beyond gender might be of particular importance in such studies.

REFERENCES

Adams, Renée, and Daniel Ferreira. 2009. "Women in the Boardroom and Their Impact on Governance and Performance." *Journal of Financial Economics* 94(2):291–309.

Amason, Allen, Rodney Shrader, and George Tompson. 2006. "Newness and Novelty: Relating Top Management Team Composition to New Venture Performance." *Journal of Business Venturing* 21(1):125–48.

Apesteguia, Jose, Ghazala Azmat, and Nagore Iriberri. 2012. "The Impact of Gender Composition on Team Performance and Decision Making: Evidence from the Field." *Management Science* 58(1):78–93.

Croson, Rachel, and Uri Gneezy. 2009. "Gender Differences in Preferences." *Journal of Economic Literature* 47(2):448–74.

Duckworth, Angela, and Martin Seligman. 2006. "Self-Discipline Gives Girls the Edge: Gender in Self-Discipline, Grades, and Achievement Test Scores." *Journal of Educational Psychology* 98(1):198.

Eckel, Catherine C., and Philip J. Grossman. 2008. Differences in the Economic Decisions of Men and Women: Experimental Evidence. In *Handbook of Experimental Economics Results*, edited by Charles R. Plott and Vernon L. Smith, Vol. 1, 509–19. New York: Elsevier.

Edwards, Bryan, Eric Day, Winfred Arthur Jr, and Suzanne Bell. 2006. "Relationships Among Team Ability Composition, Team Mental Models, and Team Performance." *Journal of Applied Psychology* 91(3):727–36.

[9] One potential explanation could be free-riding by other team members. The average value of top share is 0.05, and there is only one team where this share is larger than 0.5. Hence, excellent students are usually in the minority. Consequently, other team members might reduce their effort and try to free-ride on high-ability students, which might outweigh any direct positive impact of these "star" members on their team's success.

[10] This result also upholds when controlling for the team-specific ability median rather than the mean.

[11] This would be reminiscent of Hamilton, Nickerson, and Owan (2003) who find similar results for teamwork within firms. In a recent field experiment on the effects of cognitive ability on team performance, Hoogendoorn, Parker, and Van Praag (2017) study teams of undergraduate students who have to startup and manage a company as part of an international business program. In their study, average cognitive ability has no effect on team performance, while ability dispersion is beneficial (but only up to a point). In our context, in a regression that does not control for team ability (mean), the coefficient on team ability (SD) is −0.287 ($p = 0.103$).

Ellison, Sara, and Wallace Mullin. 2014. "Diversity, Social Goods Provision, and Performance in the Firm." *Journal of Economics & Management Strategy* 23(2):465–81.
Ferrari, Giulia, Valeria Ferraro, Paola Profeta, and Chiara Pronzato. Forthcoming. "Do Board Gender Quotas Matter? Selection, Performance, and Stock Market Effects." *Management Science*.
Fortin, Nicole, Philip Oreopoulos, and Shelley Phipps. 2015. "Leaving Boys Behind: Gender Disparities in High Academic Achievement." *Journal of Human Resources* 50(3):549–79.
Hamilton, Barton H., Jack A. Nickerson, and Hideo Owan. 2003. "Team Incentives and Worker Heterogeneity: An Empirical Analysis of the Impact of Teams on Productivity and Participation." *Journal of Political Economy* 111(3):465–97.
Hoogendoorn, Sander, Hessel Oosterbeek, and Mirjam Van Praag. 2013. "The Impact of Gender Diversity on the Performance of Business Teams: Evidence from a Field Experiment." *Management Science* 59(7):1514–28.
Hoogendoorn, Sander, Simon Parker, and Mirjam Van Praag. 2017. "Smart or Diverse Start-Up Teams? Evidence from a Field Experiment." *Organization Science* 28(6):1010–28.
Ivanova-Stenzel, Radosveta, and Dorothea Kübler. 2011. "Gender Differences in Team Work and Team Competition." *Journal of Economic Psychology* 32(5):797–808.
Kuhn, Peter, and Marie Claire Villeval. 2015. "Are Women More Attracted to Co-operation Than Men?" *Economic Journal* 125(582):115–40.
Lamiraud, Karine, and Radu Vranceanu. 2018. "Group Gender Composition and Economic Decision-Making: Evidence from the Kallystée Business Game." *Journal of Economic Behavior & Organization* 145:294–305.
Rigby, Darrell, Jeff Sutherland, and Andy Noble. 2018. "Agile at Scale." *Harvard Business Review* 96(3):88–96.
Rigby, Darrell, Jeff Sutherland, and Hirotaka Takeuchi. 2016. "Embracing Agile." *Harvard Business Review* 94(5):40–50.

How to cite this article: Städter, Silvio, Andreas Roider, Michael Dowling, and Roland Helm. 2022. "Differences in team performance: Gender versus ability." *Social Science Quarterly*. 103:926–933. https://doi.org/10.1111/ssqu.13164

APPENDIX

TABLE A1 Regression analysis of team performance–robustness

	(1)	(2)	(3)
Female share	−0.132	−0.105	−0.117
	(0.362)	(0.478)	(0.430)
Ability (Mean)	0.384***	0.241***	0.240***
	(0.000)	(0.007)	(0.007)
Ability (SD)	−0.441**	−0.290*	
	(0.014)	(0.093)	
Top share	1.556***		
	(0.009)		
Constant	0.685*	1.066***	0.803**
	(0.098)	(0.007)	(0.028)
Observations	152	152	152
Adjusted R^2	0.149	0.113	0.102
Team size FE	YES	YES	YES
Year FE	YES	YES	YES

Note: The table reports the results from OLS regressions with team performance as the dependent variable. Team performance and team ability are measured on grids that run from 1.0 (best) to 5.0 (failed). All regressions include year dummies as well as team-size dummies. *p*-Values are reported in parentheses, where *, **, and *** indicate statistical significance at the 1 percent, 5 percent, and 10 percent level, respectively.