I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 13201

Two Field Experiments on Self-Selection, Collaboration Intensity, and Team Performance

Mira Fischer
Rainer Michael Rilke
B. Burcin Yurtoglu

APRIL 2020

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 13201

Two Field Experiments on Self-Selection, Collaboration Intensity, and Team Performance

Mira Fischer
WZB Berlin Social Science Center and IZA

Rainer Michael Rilke
WHU - Otto Beisheim School of Management

B. Burcin Yurtoglu
WHU - Otto Beisheim School of Management

APRIL 2020

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Two Field Experiments on Self-Selection, Collaboration Intensity, and Team Performance*

We analyze how the team formation process influences the ability composition and performance of teams, showing how self-selection and random assignment affect team performance for different tasks in two natural field experiments. We identify the collaboration intensity of the task as the key driver of the effect of self-selection on team performance. We find that when the task requires low collaborative efforts, the team performance of self-selected teams is significantly inferior to that of randomly assigned teams. When the task involves more collaborative efforts, self-selected teams tend to outperform randomly assigned teams. We observe assortative matching in self-selected teams, with subjects more likely to match with those of similar ability and the same gender.

JEL Classification: I21, M54, C93

Keywords: team performance, self-selection, field experiment, education

Corresponding author:
Rainer Michael Rilke
WHU – Otto Beisheim School of Management
Am Weidendamm 1a
D-10117 Berlin
Germany
E-mail: rainer.rilke@whu.edu

* This paper analyses two natural field experiments. The field experiments were pre-registered with the code AEARCTR-0002757 and AEARCTR-0003646 under the title "Peer selection and performance - A field experiment in higher education". We thank Steffen Loev, Marek Becker, and Andrija Denic for extremely helpful advice and assistance with the data. We also thank Bernard Black, Robert Dur, Ayse Karaevli, Simeon Schudy, Gari Walkowitz, participants of the Advances with Field Experiments Conference in Boston, seminar participants at Higher School of Economics in Moscow, Humboldt University of Berlin, University of Trier, University of Duisburg-Essen, University of Mannheim, Burgundy School of Business in Dijon, University of Amsterdam, and WHU - Otto Beisheim School of Management for helpful comments and suggestions on earlier versions of this paper.

1 Introduction

In today's highly complex economic environment, cooperation among individuals is crucial for organizational success. As business becomes increasingly global and cross-functional, the adoption of teamwork increases in all domains of work life (O'Neill and Salas, 2018; Cross et al., 2016). The nature and effectiveness of teamwork in a variety of productive activities matter for outcomes in diverse settings ranging from entrepreneurial ventures (Reagans and Zuckerman, 2019) to the mutual fund industry (Patel and Sarkissian, 2017) and from medical practice (Geraghty and Paterson-Brown, 2018) to achieving scientific breakthroughs (Wuchty et al., 2007).

Recognizing the importance of cooperation in teams, economists and management scholars are extensively studying the influence of various forms of team incentives (e.g., team bonuses or tournaments) on team performance. While team bonuses and team piece rates tend to have a positive influence on productivity (e.g., Englmaier et al., 2018; Friebel et al., 2017; Hamilton et al., 2003; Erev et al., 1993), the evidence on the influence of team tournament incentives on performance is ambiguous (e.g., Delfgaauw et al., 2019, 2018; Bandiera et al., 2013). Thus the precise channel through which team incentives lead to differences in team performance remains an open question.

Moreover, another performance-enhancing effect has gone largely unnoticed: the power of team incentives to influence the team formation process and, through it, team performance. When people are allowed to choose their teammates, they tend to match with people they like (e.g., Curranrini et al., 2009). However, they also appear to trade off both the pecuniary benefits from better cooperation and the non-pecuniary benefits of working in teams with friends against the pecuniary benefits stemming from working with higher-ability team members (Bandiera et al., 2013; Hamilton et al., 2003).[1]

[1]Other laboratory experiments examine the link between different group formation mechanisms in public goods or minimum-effort games. This literature shows that the level of cooperation in endogenously formed groups is similar to the level of contributions in groups with exogenous matching (e.g., Gächter and Thöni,

Only a small number of studies focus on the effects of team composition. Therefore, very little is known about the different effects on team performance of self-selection versus the random assignment of team members. Indeed, while studies in this literature find that self-selection can lead to higher team performance (Dahlander et al., 2019; Chen and Gong, 2018; Hamilton et al., 2003), none of them look at whether this effect varies across tasks with different levels of collaboration intensity or how self-selection influences the ability composition of teams.

This study both analyzes how the team formation process influences the ability composition of teams and team performance, and thus shows how self-selection and random assignment influence team performance for different tasks. Using two randomized natural field experiments, we identify the collaboration intensity of the team task as the key driver of the effect of self-selection on team performance.

We expect that endogenous team formation will lead to a higher degree of cooperation and a concentration of people with similar abilities and social preferences within teams. Consequently, self-selection should lead to a larger variance in team performance than in randomly assigned teams. However, whether these selection patterns lead to the overall superior performance of self-selected teams is prima facie unclear.

We argue that the impact of team formation on performance will hinge on the production function of a particular team task. The practical relevance of this question stems from team tasks taking different forms. For example, when a team of employees is assigned a task such as implementing a strategy or preparing a presentation of a product, the teammates have to work closely together to combine their efforts effectively. This process requires high-intensity collaboration, which is easier when the teammates already know one another, but individual abilities may not be as important. In contrast, a team of employees tasked to invent a new product or create a new research idea are involved in a process requiring a lower intensity of collaboration, so that the ability of each individual matters more for the team's outcome.

2005; Guido et al., 2019; Chen, 2017). A related and higher-level question of why people choose to join teams at all has recently also been investigated in laboratory experiments (Cooper et al., 2019).

Thus, when the underlying task is collaboration-intensive, we hypothesize that self-selection is beneficial for average team performance because it increases collaborative efforts. However, when the underlying task is ability-intensive, self-selection can be detrimental to average team performance because it leads to a concentration of skills in some of the teams. To test this hypothesis, we conduct a randomized field experiment that incorporates tasks requiring different intensities of collaborative efforts.

We embed the experiments in a mandatory course for first-year undergraduate students at a major German business school. We conducted the experiments during the winter quarters 2017/18 and 2018/19 with two cohorts of students randomly assigned to two separate study groups. In both experiments each study group received the same course content from the same instructor but in separate classes. The course required the students to work on two distinct tasks in teams of two. In one class, the instructor told the students on the first day to form a team with a fellow student of their choice. The students also had to give the instructor their teammate choice in written form before the next class (treatment *Self*). In the other class, students were randomly assigned to a team of two before the second class meeting (treatment *Random*).

All teams had to work on two types of tasks that were similar in the skill set they required but different in terms of collaboration intensity: a written task and a presentation task involving the participation of both teammates. In both tasks, teams had to solve microeconomics problems, and their grades depended solely on the accuracy of their solutions, presented in written form and presentation form, respectively. For the written task, we expected that if one student alone came up with the correct solution, whether the other team member could also produce it would be of less relevance, so that collaboration was necessary only for agreeing on the solution to be submitted. In contrast, for the presentation task, we expected collaboration to be essential for ensuring that each member could present their parts of the correct solution and for ensuring coherence of the different parts.

For the ability-intensive written task, we first find that the self-selected teams performed

significantly worse than the randomly assigned teams, whereas, for the collaboration-intensive presentation task, self-selected teams performed better overall. We then studied two channels through which the team formation process may affect outcomes: ability composition and the quality of within-team cooperation. Consistent with prior empirical evidence, we find that self-selection results in teams of students with more similar abilities. In contrast, random assignment increases the spread of high-ability students across teams, thereby increasing the number of teams with at least one high-ability member. We also find that self-selection enables more students to work with a teammate whom they like and with whom they cooperate more intensively, a factor that seems to drive the superior performance in the more cooperative task.

Our study adds to the thus far small literature examining the consequences of team composition mechanisms in real-world settings. Chen and Gong (2018) find that university students who may self-select their teammates perform better on a presentation task than students who are randomly assigned to teams. Likewise, Dahlander et al. (2019) find that students – in an entrepreneurial task – who can freely choose with whom they want to work with perform better when they were given an entrepreneurial task as to when the group is free to choose their task. While Chen and Gong (2018) show that self-selection leads to team formation based on social connections rather than member skills, neither they nor Dahlander et al. (2019) examine how the assignment mechanism affects actual skill composition within groups. Our setting, however, allows for a nuanced assessment of the compositional effects of different mechanisms of team formation.

Our study makes three contributions to the literature. First, we formalize how the ways teams form may interact with task features to affect performance. Second, using two randomized natural field experiments, we test how self-selection of teammates affects ability composition as well as team performance in different tasks. Finally, our study combines theses insights to offer an explanation of how the effect of self-selection on team composition may explain why self-selected teams are superior in tasks that require a high level of collaborative efforts but not in other team tasks.

Specifically, our study shows that these differential effects are likely due to the potential of the self-selection mechanism affecting both the social and ability composition of teams. In a task that, beyond cognitive skills, requires a high level of collaborative efforts, self-selection is favorable because it allows friends to work together. However, if the task problem can be solved primarily by cognitive skills, then self-selection—by raising the proportion of teams in which both members have low ability—leads to lower overall performance.

2 Theoretical considerations

We introduce a theoretical framework to demonstrate why the ability composition of the team and the intensity of cooperation can affect team performance, depending on the type of task. We consider two types of team tasks: an *ability-intensive* task, in which teams have to submit a handwritten solution to a problem set, and a *collaboration-intensive* task, in which teams have to submit a short video presentation of their solution to a problem set (section 3.3 gives a detailed description of the two tasks).

Two students, denoted as i and j, have to form a team. Each teammate has a unidimensional cognitive ability level a_i for solving the task. Providing the right solution to problem sets is determined by the ability composition of the team and the quantity of collaborative effort $(q_{i,j})$ between both teammates.

We assume that the quality of the team's solution, s, depends on the ability level of the ablest student and the level of collaborative effort between the two teammates. The higher one teammate's ability and the higher the collaborative effort between the two teammates, the better the solution will be. A task may differ by the relative amount of cognitive skills and collaboration necessary for solving. We introduce the weights p and $1-p$ for representing the relative importance of cognitive skills and collaboration, respectively. Thus the quality of the team's solution, s, is given by:

$$s = p \cdot max(a_i, a_j) + (1 - p) \cdot q_{i,j}.$$

Although highly stylized, this relationship captures the intuition that, in a task that is highly dependent on the one dimension of cognitive ability, only the ability of the ablest teammate matters for team performance. If one person can find a suitable solution to a problem or spot the mistakes in the other teammate's solutions, the ability of the other teammate is not important, and such a task does not require any collaborative effort $q_{i,j} = 0$. Thus, in the *ability-intensive* team task, the team grade is determined by (s_a):

$$s_a = max(a_i, a_j).$$

In contrast, in the *collaboration-intensive* team task, in addition to finding the right solution, both students have to actively participate in presenting their team solution in a short video clip. We assume that, in addition to cognitive ability, this task requires a certain degree of collaboration. We expect that the level of collaborative effort is higher in self-selected teams due to better communication, a higher willingness to help the other member, and higher intrinsic motivation. In this case, the team's grade (s_c) is determined by the weighted sum of the maximum ability and the quality of collaboration:

$$s_c = p \cdot max(a_i, a_j) + (1 - p) \cdot q_{i,j}.$$

The literature suggests that when subjects are allowed to choose their teammates, they tend to choose a teammate who has similar abilities and with whom they are acquainted. Thus the maximum ability in self-selected teams is—on average—lower than in randomly assigned

teams. Yet the expected quality of collaboration is higher in self-selected teams. Therefore, we expect that randomly assigned teams perform—on average better—on the written task than self-selected teams. Furthermore, we expect the advantage of the randomly assigned teams over the self-selected teams to be smaller in the presentation task, because self-selected teams lacking a high-ability member can usually make up for this disadvantage by superior cooperation. If p is sufficiently small (i.e., the importance of collaborative effort is higher and the importance of the maximum ability within the team is lower), then self-selected teams may even outperform randomly assigned teams.

3 Study

3.1 Context and background

The field experiment was conducted with students of the BSc program at a German business school between October 2017 and April 2019. The business school offers university education in business administration, with degrees at the BSc, MSc, MBA, and PhD levels, as well as executive education programs. The school has around 2000 students. At the BSc level, the school offers the International Business Administration program, to which the school admits roughly 230 students in each year. In academic year 2017/2018, a total of 672 students were enrolled in the program, 26% of whom were female.

Studying the impact of team formation mechanisms on team performance requires an environment in which participants can choose teammates, in which the selection mechanism can be exogenously varied, and in which team performance can be objectively measured. The environment of the business school class we study fulfills all of these criteria while allowing us to maintain a high degree of control. Furthermore, to observe self-selection not only on demographic characteristics but also on ability we need a sample of participants that have prior acquaintance with each other. This is given for our student subjects as the class takes

place at a point in time when students already completed courses together and had ample opportunities to get to know each other at extra-curricular activities (e.g., student societies, sports teams, and opportunities for involvement in music, drama, political campaigning, or community work) organized at the business school, located in a small town.

3.2 Experimental timeline and treatments

The field experiments took place in the Microeconomics I course, with two cohorts of first-year students in the BSc program in International Business Administration. In each cohort, students were randomly assigned to two separate classes, both taught by the same instructor (one in the morning and one in the afternoon of the same day). During the first week, students learned that fulfilling the course requirements included completing two tasks in teams of two and passing an exam at the end of the quarter. As the instructor did not announce any task-specific details on the team tasks in the first week, the students knew only that these tasks were take-home assignments that they had to complete during study hours. They also knew that they would have to solve both tasks with the same team member, because re-matching was not permitted.

For each cohort, in one class—the *Self* treatment—the instructor told the students on the first day to form a team with a fellow student of their choice. Students had to submit their team composition to the instructor in writing before the second meeting. In the other class—the *Random* treatment—students were randomly assigned to a team of two, and team composition was announced by email before the second meeting.

The first team task, handed out in mid-November, had to be submitted in early December. The second team task, handed out in early December, had to be submitted by the end of January. The final exam took place in March. During the course students received no feedback on their performance in the team tasks. After the final exam, feedback consisted only of students' overall course grade. Upon request students could also receive detailed information

on both their team's performance in the different tasks and their individual performance in the exam. Figure 1 displays the timeline of the experiment.

In the winter quarter of 2017/18 (Experiment I, n=190, 31% female) students completed two written team tasks, whereas in the winter quarter of 2018/19 (Experiment II, n=192, 29% female), the first was a written task and the second was a presentation task. Across both experiments the first task was identical and the solutions were supposed to be submitted in written form. The second task, although very similar in content, differed between both experiments; while in Experiment I solutions were supposed to be submitted in written form, in Experiment II students were required to give a presentation (see subsection 3.3). This design allows us to identify interaction effects of the team formation mechanism with task characteristics as well as possible heterogeneous trends in collaboration across treatments.

3.3 Tasks

Given that we expected the effect of the team assignment mechanism on team performance to hinge on the degree of collaboration intensity in a given task, we chose two tasks that differed in collaboration intensity: a written task and a presentation task.[2] Although the written task required students to reach an agreement on whose solution is best, one student may produce the correct solution by himself or herself. In contrast, the presentation task requires students to jointly prepare a presentation in which both teammates' contributions had to be visible. This process required a higher degree of cooperation and coordination than simply agreeing on the best solution. In both the written and presentation tasks, students' solutions were graded solely on the correctness of their answers.

The written tasks consisted of microeconomic problems for which students had to submit written solutions. These problems called for applications of the theoretical knowledge that students had acquired during class lectures, such as analyzing demand patterns, calculating

[2]The exercise sets appear in the online appendix.

market outcomes, or designing pricing strategies. A solution usually involved an explanation of the theoretical background, a correct approach to the solution, and a series of calculations that possibly included one or two graphs.

In addition, the instructions for the written tasks specified that the students had to present their written answers clearly. Answers could be either typed or handwritten and they had to be legible. Some questions required outlining a correct methodological approach and the corresponding calculation. Other questions required conceptual explanations in complete sentences, or also in bullet points if they were clear and significant.

The presentation task consisted of questions for which students had to submit their solutions in a five-minute presentation (recorded as a video). The questions required a level of microeconomics skills similar to that in the written task. The instructions required both team members—and their individual contributions—be visible in the video.

In addition, the instructions specified that the video should be comprehensible, i.e., it should be possible to understand the presenters' speech. Teams were allowed to use graphs, illustrations, and slides to make their presentations more effective. The instructions further stated that teams could use their smartphones to produce the video and that the technical quality of the video itself would not be graded.

Although we conducted our experiments in an educational setting, the characteristics of the two tasks share commonalities with a variety of tasks in many organizational settings. The written tasks resembles problem-solving in product and strategy development, which call for high levels of technical skills. In contrast, the presentation task—while keeping constant the level of technical skills required—requires students to collaborate effectively. The presentation component thus resembles tasks in production and service jobs that do not require specific technical skills but are collaboration-intensive, requiring good communication and coordination among teammates.

3.4 Data

Data for the study was gathered from three distinct sources (see figure 1). Our baseline data contains students' high school performance (GPA), their performance on the admissions tests, and student-level demographic information. Both the GPA and the results of the admissions tests are clean measures of pre-experiment academic ability as they are not affected by peers at the business school. Moreover, our data contains information on students' performance in the two team tasks and on the final course exam at the end of the quarter. The endline data includes information collected through a post-experiment survey, in which we elicited students' perceptions of cooperative behavior in their team and their prior relationship with their team member. Furthermore, it contained an incentivized measure of pro-sociality. We provide a more detailed description of the different data sources in the following subsections.

3.4.1 Pre-experiment ability measures

Our pre-experiment ability measures and demographic variables come from the business school's student registry, in particular its admissions data. The business school's program, which is known as highly competitive, uses a selective admissions procedure. In the first step of the admissions process, applicants to the BSc program provide basic background information and their high school grade point average (GPA).[3] The admissions office ranks applicants by their GPA and invites the top 10% to an admissions day, where applicants have to give an oral presentation, participate in a group discussion, and take part in an interview. Two evaluators rate the applicants' performance on these tasks on a scale from 1 to 10. Towards the end of the admissions day, applicants take an analytical test, on which their performance is also rated on a scale from 1 to 10.

[3]The German GPA (Abiturnote) ranges from 4.0 (the worst) to 1.0 (the best) grade and is the most important criterion for university admission in Germany (e.g., Fischer and Kampkötter, 2017). Our entire sample has an average GPA of 1.79 (SD=.504). For our analysis, we invert the GPA so that higher values indicate better grades.

The first applicant task is to give a 15-minute presentation on a self-chosen topic (Presentation). A round of questions by two evaluators then follows, for up to five minutes. Applicants are evaluated on the presentation content and structure (e.g., is it logically structured), the presentation format and delivery (e.g., is the diction articulate), and on the quality of dialogue and their ability to respond to questions.

The second applicant task is to participate in a group discussion (Discussion) on a topic given to them on the spot. The discussion takes 50 minutes. Evaluators rank applicants according to group-related behavior (e.g., does the applicant contribute to a group solution), problem-related behavior (e.g., is the applicant able to argue well and convince others), and management-related behavior (e.g., does the applicant motivate other group members or search for compromises between opposing views).

The third applicant task is an individual interview (Interview). Evaluators ask questions about the student's personal background to assess the student as a future business administration student. Moreover, the evaluators rate the candidates' likelihood to persevere with their career goals.

The fourth applicant task is an analytical test (Analytical test), during which applicants solve quantitative problems, including diagramming techniques from the business world. Applicants have 60 minutes to complete the test. This written task, which constitutes the final round of the admissions testing, helps the school assess the analytical reasoning skills of the candidates.

The subjective ratings of both evaluators on the first three tasks are averaged and then added to the score on the analytical test. The result represents the applicant's final score. The admissions office ranks students according to this score and offers the best candidates a place in the program.

3.4.2 Measures of team performance and exam outcomes

Students' performance on both team tasks and the individual exam determined their final grade. Each team received a common grade for their performance per task and each task had a weight of 15% towards the individual final grade. The exam was written in the end of the course and contributed 70% to the final grade. Student assistants blinded to the experiment graded the performance on the team tasks and exam.

3.4.3 Post-experiment survey

On the day following the final exam, we invited the students to take part in an online post-experiment survey. This survey elicited students' perceptions of collaboration in their team, their prior relationship to and beliefs about their teammate's performance, and perceptions of the quality of teaching. To incentivize participation, we used a raffle.[4] Among all participants of the survey, a random draw was to pick one student, who would receive 200 EUR as a reward. For an incentivized measure of students' pro-sociality, we asked students which fraction of this amount they would like to donate to UNICEF if they won.

4 Results

We organize the presentation of our results around our main research question: How does self-selection influence team performance? We begin our analysis by establishing the internal validity of our experimental approach. We show that the student sample does not differ between the treatments on any observable variable elicited before the experiments. Then we analyze the influence of the two assignment mechanisms on performance. Our results show that team performance by randomly assigned teams is higher than that by self-selected

[4] An overview of the survey questions appears in Table 5, which we discuss in Section 4.3.2.

teams for a task not requiring close collaboration. However, for a task that requires teammate cooperation, the randomly assigned teams tend to perform worse than the self-selected ones. We then explore the mechanisms through which team formation affects performance.

4.1 Randomization checks

Table 1 provides an overview of the properties of our sample in the treatments and the experiments. We show separate summary statistics for Experiment I, Experiment II, and pooled for both experiments. The table shows that the randomization was successful in producing highly similar groups based on observable characteristics such as high school performance (GPA) and performance on the admissions test. The only characteristic that differs significantly between treatments in Experiment II is the percentage of female students ($p = .038$, χ^2-test, one-sided).[5] We therefore provide results from two regression specifications, both with and without controlling for gender (and other observables).

4.2 Team performance

Our objective outcome measure of performance is the score that teams receive for their work on two separate tasks during the quarter. We summarize our results in Figure 2, which plots the standardized average team score for each task and the individual performance on the final exam. The left panel shows the outcomes for Experiment I; the right panel shows the outcomes for Experiment II.

For Experiment I, where the first and the second team tasks had to be submitted in written form, the figure indicates that—on average—teams in *Random* perform better than teams in *Self*. A non-parametric comparison of average team scores yields a significantly lower score for teams in *Self* than teams in *Random* ($p = .007$, Mann-Whitney U test) (hereafter MWU

[5]Unless otherwise stated, all p-values are based on two-sided tests.

test). A non-parametric test for the equality of variances between the treatments underlines this pattern and shows that the variance of team performance is significantly larger in the *Self* treatment ($p = .002$, Levene's test).[6] We also observe no change in performance over time. A comparison of average performance between the first and second team tasks reveals no significant differences (*Self*: $p = .885$, *Random*: $p = .9291$, MWU test). Neither average student performance nor variance of student performance on the final exam differs significantly across treatments ($p = .455$, MWU test; $p = .995$, Levene's test).

First, for Experiment II, the figure indicates that teams in *Self* perform worse on the written task than those in *Random*, while the effect appears reversed when the teams work on the video task. Consequently, in the first team task of Experiment II, we replicate the observed pattern of Experiment I. Again, teams in *Self* perform significantly worse than those in *Random* when the task is written, but this time the variances are not significantly different ($p = .064$, MWU test; $p = .194$, Levene's test). A look at Figure 2 reveals that the average team performance is higher in *Self* than in *Random* in the second team task (presentation task). However, non-parametric tests comparing the mean and the variance of average team performance between *Self* and *Random* cannot reject the null hypothesis that performance in both treatments is equal ($p = .156$, MWU test; $p = .381$, Levene's test).

Second, we run regressions controlling for pre-experiment observables (i.e., GPA, admissions test score, gender) to verify these outcomes. To do so, we analyze the teams' performance for the first (written) and second (written or presentation) team tasks across both experiments. The first team task in both experiments was identical, with students submitting their work in writing. To test the influence of task characteristics on team performance, we vary the second team task. In Experiment I, teams had to submit their solutions in written form, while in Experiment II, teams had to submit video presentations (as described earlier). We use OLS regressions with standard errors clustered at the team level.

[6] A separate analysis of the first and second team task yields similar significant differences in averages (1^{st} team task: $p = .011$, 2^{nd} team task: $p = .068$, MWU test), and (marginally) significant differences in variances (1^{st} team task: $p = .104$, 2^{nd} team task: $p = .001$, Levene's test). A detailed pairwise comparison appears in Table A.1 in the Appendix.

Table 2 shows the results of the regression analysis, where the dependent variable is the team performance (z-standardized) for the both team tasks, separately. In Models (1)-(3), we predict the team performance on the first team task. Model (1) includes only a dummy variable for the treatment *Self* ("1 if *Self*"). Self-selected teams perform on average .42 standard deviations worse at the first task than randomly-assigned teams. Model (2) includes a dummy variable for the experiment ("1 if Experiment II") and an interaction term of the *Self* treatment and the experiment ("1 if *Self* x 1 if Experiment II") to control for potential interactions. While both of these control variables remain insignificant, the coefficient on the treatment dummy *Self* remains significant and almost unchanged at -.47, indicating that, for the first task, the treatment effect is not significantly different across experiments. In Model (3) we include additional controls, finding that the treatment effect is not affected by their inclusion. We find that GPA, but not the admissions test score, predicts team performance.

Next, we study the second team task. The regression results appear in Models (4)-(6). In Model (4), we pool observations from both experiments (ignoring the type of task) and include only a treatment dummy. Consistent with the results of the non-parametric analysis, we find no significant effect of self-selection, suggesting that a meaningful investigation of the effects of team assignment on performance should take into account the task characteristics. After we control for the experiment and interact with the treatment, we find that teams in the *Self* treatment in Experiment I performed .50 standard deviations worse in the second task than teams in the *Random* treatment (model 5). We thus find very similar treatment effects for the first and the second tasks in Experiment I, suggesting that there is no heterogeneous learning across treatments and that the ordering of the tasks does not matter.

In addition, adding up the first and the third coefficients in Model (5), we find that in Experiment II, in the second task (a presentation), teams in *Self* tended to perform .28 standard deviations better than teams in *Random.* In line with the non-parametric analysis, a joint F-test shows that this difference is not significant ($p = .1774$).

4.3 Mechanisms

The critical question that now arises is why self-selection has a detrimental effect on team performance in ability-intensive but not in collaboration-intensive tasks. We shed light on two mechanisms that help answer this question. Specifically, we originally hypothesized that our treatment variation would influence the composition of teams and the quality of within-team cooperation. We expected that when individuals can self-select into teams, they would tend to prefer teammates with similar characteristics and abilities, and thus exhibit higher quality cooperation.

4.3.1 Team composition

We begin by looking into how students (in treatment *Self*) form teams. To do so, we use pre-experiment registry data on students' ability (measured as their performance on the various tasks in the admissions test and their GPA), gender, and an incentivized measure of pro-sociality from the post-experiment survey. For each team and measure m, we calculate the absolute difference between both teammates

$$m_{ij} = |x_i - x_j|$$

where i and j are teammates.

Thus lower absolute differences indicate higher similarity of teammates, and higher values indicate higher dissimilarity. If students in *Self* were to match on certain measures, we would observe a higher similarity, i.e., a lower average absolute difference. Moreover, as a reference point, we calculate the average absolute difference after simulating the matching of each student with all potential teammates from the respective treatment. This simulation provides us with information on how a within-sample random team composition would hypothetically look.

The results appear in Table 3. The first column shows the absolute difference for all measures in *Self*, while the second column shows the absolute difference for all measures in the simulated *Random* "treatment." A comparison of the values in the first and the second columns suggests that students sort themselves into teams with students of similar levels of ability and pro-sociality, and the same gender. More specifically, we observe that self-selected teams display a higher degree of similarity in terms of their GPA, their score on the analytical test, and gender. These differences are significant (GPA: $p = .003$; Analytical test: $p = .012$; Female: $p = .0001$, Wilcoxon signed rank test, WSR test). Interestingly, we do not find significant differences in either the *Self* treatment or the simulated *Random* "treatment" for interview, discussion, and presentation skills.[7] Furthermore, students tend to have more similar levels of pro-sociality in the *Self* treatment, however, this difference not statistically significant (Pro-sociality: $p = .224$).

We observe that the teams in *Self* are significantly more homogeneous in terms of ability (GPA, analytical test) and gender, and tend to be more similar in terms of their pro-sociality. This observation suggests that high levels of ability and pro-sociality (and same gender) are desirable characteristics in a potential teammate. We further investigate whether students in treatment *Self* trade off these two desirable characteristics against each other. To do so, we examine whether (a) the student's ability positively predicts the teammate's ability, (b) the student's pro-sociality positively predicts the teammate's pro-sociality (as suggested by the previous analysis of similarity) and (c) whether the student's pro-sociality also predicts the teammate's ability and vice versa.

In Table 4 (model 1) we regress GPA on the teammate's GPA. The positive and significant coefficient ($p = .019$) indicates that more (less) able students match with more (less) able teammates. This association is robust and stays significant at conventional levels after the inclusion of various additional control variables such as the student's and the teammate's scores on the analytical test (model 2), and the individual's and the teammate's level of

[7]These measures also do not correlate with performance on the different team tasks (see table A.2).

pro-sociality (model 3). This finding corroborates the results in Table 3 that a student's score on the analytical test does not predict his or her teammate's GPA. Model (3) shows that the student's pro-sociality positively predicts the teammate's ability.

In Models (4)-(6) we regress the GPA, the analytical test score and pro-sociality on the teammate's score in the analytical test but find no significant associations. Model (7) shows that the student's pro-sociality does not significantly predict the teammate's pro-sociality. Moreover, the student's GPA positively predicts the teammate's pro-sociality (model 8), also when we control for the student's pro-sociality (model 9).

Overall, the findings in Tables 3 and 4 show (a) that students tend to choose a teammate of similar abilities and the same gender, (b) that students of higher ability (independent of their own pro-sociality) match with more pro-social teammates, and (c) that students who are more pro-social (independent of their own ability) match with higher-ability teammates. These findings suggest that a high level of ability, a high level of pro-sociality, and the same gender are desirable traits in a teammate because students expect them to positively contribute to team performance, to make working together more pleasant, or both. The next subsection investigates whether this student expectation is fulfilled.[8]

4.3.2 Collaborative behavior within teams

A second mechanism that we hypothesized to be affected by the treatment variation and to influence team performance is the quality of cooperation. In our post-experiment survey, we asked students to evaluate their collaboration experience in their team during the course (see table 5 for an overview of all questions).

We asked students to agree or disagree (on a 7-point Likert scale) with several statements aimed at capturing various aspects of team collaboration and organization. More specifically,

[8]We run the very same analysis with the data from *Random*. The results appear in Table A.3 in the Appendix. We observe no systematic patterns here, although some coefficients are weakly significant.

we also asked questions about the perceived quality of cooperation and the pleasure of working together.[9]

Table 5 reports the results from the post-experiment survey, pooled for Experiments I and II and for each experiment separately. For their experience during the task, students in *Self* reported having communicated more ("We communicated a lot"; $p < .0001$, MWU test) and to have cooperated better ("We helped each other a lot"; $p = .0188$) than students in *Random*. Moreover, they indicated that the teammates' contributions were more equally distributed ("Both team members contributed equally"; $p = .021$) and that both teammates exerted effort ("Both team members exerted effort"; $p = .002$). These comparisons clearly show that teams in *Random* solve the problem sets with a different approach than teams in *Self*, likely by assigning the task to the more able teammate but also cooperating less.[10]

Furthermore, the student's mood ("The mood in our team was good"; $p = .189$), levels of stress ("Our team was very stressed."; $p = .134$), and motivation ("Our team was very motivated"; $p = .151$) in teams in *Self* was not different in their teams compared to teams in *Random*.

Although student in *Self* were more likely to report being friends ("My team member was a friend"; $p < .0001$) or having been acquainted with their teammate before the course ("I knew the team member very well before the course"; $p < .0001$), it was not the overall pleasure of working together but rather the higher level of cooperation that was different in team between the treatments.

These findings highlight a potentially important channel through which random assignment may increase performance in the written task while tending to decrease performance in the

[9] We also ask a battery of questions about the perceived teaching quality, which might influence performance. However, we found no significant differences between the treatments and experiments, indicating that the lecturer's teaching was of the same quality in both classes and experiments.

[10] As Table 5 shows, these differences between the treatments are mostly driven by the reports of students from Experiment I, where students were not explicitly required to collaborate. As the video presentation in Experiment II required each teammate to cooperate equally and to appear in the video to present the results, they might have tried to fulfill this expectation. Therefore, a desirability bias might explain why we do not find as strong a difference in self-reported cooperation in Experiment II as in Experiment I. That the average ratings of cooperation also tend to be higher in Self in Experiment II than in Experiment I points in the same direction.

presentation task. For the written task, which requires fewer collaborative efforts, letting the ablest student perform the task is most efficient; for the presentation task, which requires collaborative efforts, both communication and coordination work better in self-selected teams.[11]

5 Conclusion

This paper provides evidence from a two natural field experiments that study how team formation processes influence team performance. We use data on the individual characteristics and behavior of students at a business school to understand the effect on team performance of varying both the team formation process and the collaboration-intensity of tasks. The results of our randomized field experiments add a new dimension to the debate on the effects of team formation on team performance. Previous experiments do not use objective ability measures to capture selection patterns, nor do they offer an explanation for observed effects of the team formation process on performance at different tasks. In these natural field experiments, we use registry data on student ability generated prior to the experiments to study how the team formation process affects the teams' ability and social composition, which in turn affect team performance in two distinct team tasks with different collaboration intensities.

We find that the team formation mechanism chosen for assigning subjects to teams is a useful tool for strategically inducing performance. Importantly, this relationship hinges on the specific requirements of the underlying task. When subjects are allowed to choose their teammate, the team assignment mechanism substantially influences performance on the team tasks through assortative selection patterns. These selection patterns prove to be performance-enhancing when the underlying task requires a high degree of collaborative efforts. In contrast, random assignment of teammates leads to better team performance when the task requires little

[11] 70% (Experiment I: 74%, Experiment II: 67%) of students responded to our request to participate in the survey. We test and find no significant difference in the fraction of participating students between *Random* and *Self* (Experiment I: $p = .282$, Experiment II: $p = .261$, χ^2 test). Furthermore, participation in the survey was balanced for GPA ($p = .146$, $p = .466$, MWU test), the analytical test ($p = .091$, $p = .334$, MWU test), and gender ($p = .730$, $p = .822$, χ^2 test).

collaborative efforts but a high level of (unidimensional) technical skills. After the team task, we measure individual performance of subjects and find no differences between the team formation mechanisms, indicating that the effect that we observe at the team-level does not translate to individual performance differences.

The study offers valuable insights for managers and team leaders, i.e., those who decide how teams are put together in firms and other organizations. If managers want to maximize team performance, they first need to consider the type of task involved before deciding whether employees can self-select their teammates. Given that randomly assigned teams can produce superior outcomes for tasks that are characterized by a low level of collaboration intensity, our findings also reveal a weakness in trends towards more "agile work practices" (e.g., Mamoli and Mole, 2015), which give employees the freedom to choose their working groups regardless of circumstances.

Moreover, when managers want to create a more inclusive work environment by forming more diverse teams or teams with a similar average skill level, random team assignment might prove more beneficial. Our field experiment shows that students are more likely to match with teammates of the same gender when they are allowed to self-select. This finding suggests that self-selection might create not only inequalities in abilities across teams but also less gender-diverse teams.

References

Bandiera, O., I. Barankay, and I. Rasul (2013). Team incentives: Evidence from a firm level experiment. *Journal of the European Economic Association 11*(5), 1079–1114.

Chen, R. (2017). Coordination with endogenous groups. *Journal of Economic Behavior & Organization 141*(5), 177–187.

Chen, R. and J. Gong (2018). Can self selection create high-performing teams? *Journal of Economic Behavior & Organization 148*, 20–33.

Cooper, D. J., K. Saral, and M. C. Villeval (2019). Why join a team? *IZA Discussion Paper* (12587).

Cross, R., R. Rebele, and A. Grant (2016). Collaborative overload. *Harvard Business Review*.

Curranrini, S., M. O. Jackson, and P. Pin (2009). An economic model of friendship: Homophily, minorities, and segregation. *Econometrica 77*(4), 1003–1045.

Dahlander, L., V. Boss, C. Ihl, and R. Jayaraman (2019). The effect of choosing teams and ideas on entrepreneurial performance: Evidence from a field experiment. *Mimeo*.

Delfgaauw, J., R. Dur, O. A. Onemu, and J. Sol (2019). Team incentives, social cohesion, and performance: A natural field experiment. *Tinbergen Institute Discussion Paper*.

Delfgaauw, J., R. Dur, and M. Souverijn (2018). Team incentives, task assignment, and performance: A field experiment. *The Leadership Quarterly*.

Englmaier, F., S. Grimm, D. Schindler, and S. Schudy (2018). The effect of incentives in non-routine analytical team tasks - Evidence from a field experiment. *CESifo Working Paper Series* (6903).

Erev, I., G. Bornstein, and R. Galili (1993). Constructive intergroup competition as a solution to the free rider problem: A field experiment. *Journal of Experimental Social Psychology 29*(6), 463–478.

Fischer, M. and P. Kampkötter (2017). Effects of German universities' excellence initiative on ability sorting of students and perceptions of educational quality. *Journal of Institutional and Theoretical Economics 173*(4), 662.

Friebel, G., M. Heinz, M. Krüger, and N. Zubanov (2017). Team incentives and performance: Evidence from a retail chain. *American Economic Review 107*(8), 2168–2203.

Gächter, S. and C. Thöni (2005). Social learning and voluntary cooperation among like-minded people. *Journal of the European Economic Association 3*(2), 303–314.

Geraghty, A. and S. Paterson-Brown (2018). Leadership and working in teams. *Surgery (Oxford) 36*(9), 503–508.

Guido, A., A. Robbett, and R. Romaniuc (2019). Group formation and cooperation in social dilemmas: A survey and meta-analytic evidence. *Journal of Economic Behavior & Organization 159*, 192 – 209.

Hamilton, B. H., J. A. Nickerson, and H. 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), 465–497.

Mamoli, S. and D. Mole (2015). *Creating Great Teams: How Self-selection Lets People Excel*. Pragmatic Bookshelf.

O'Neill, T. A. and E. Salas (2018). Creating high performance teamwork in organizations. *Human Resource Management Review 28*(4), 325–331.

Patel, S. and S. Sarkissian (2017). To group or not to group? Evidence from mutual fund databases. *Journal of Financial and Quantitative Analysis 52*(5), 1989–2021.

Reagans, R. and E. W. Zuckerman (2019). Networks, diversity, and productivity: The social capital of corporate R&D teams. *Organization Science 12*(4), 502–517.

Wuchty, S., B. F. Jones, and B. Uzzi (2007). The increasing dominance of teams in production of knowledge. *Science 316*(5827), 1036–1039.

6 Figures

Figure 1: Sequence of events and data sources

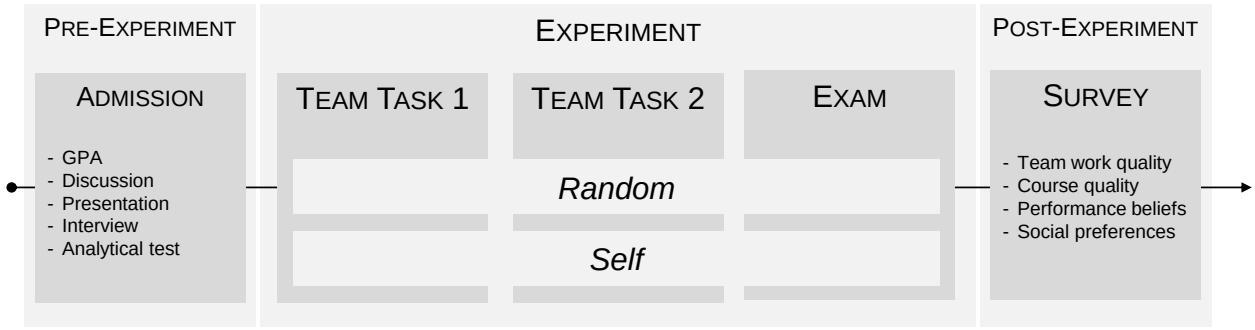

Figure displays variables and the sequence of events in the experiments. The sequence of events is the same for both *Experiment I* and *Experiment II*.

Figure 2: Team assignment, performance, and task characteristics

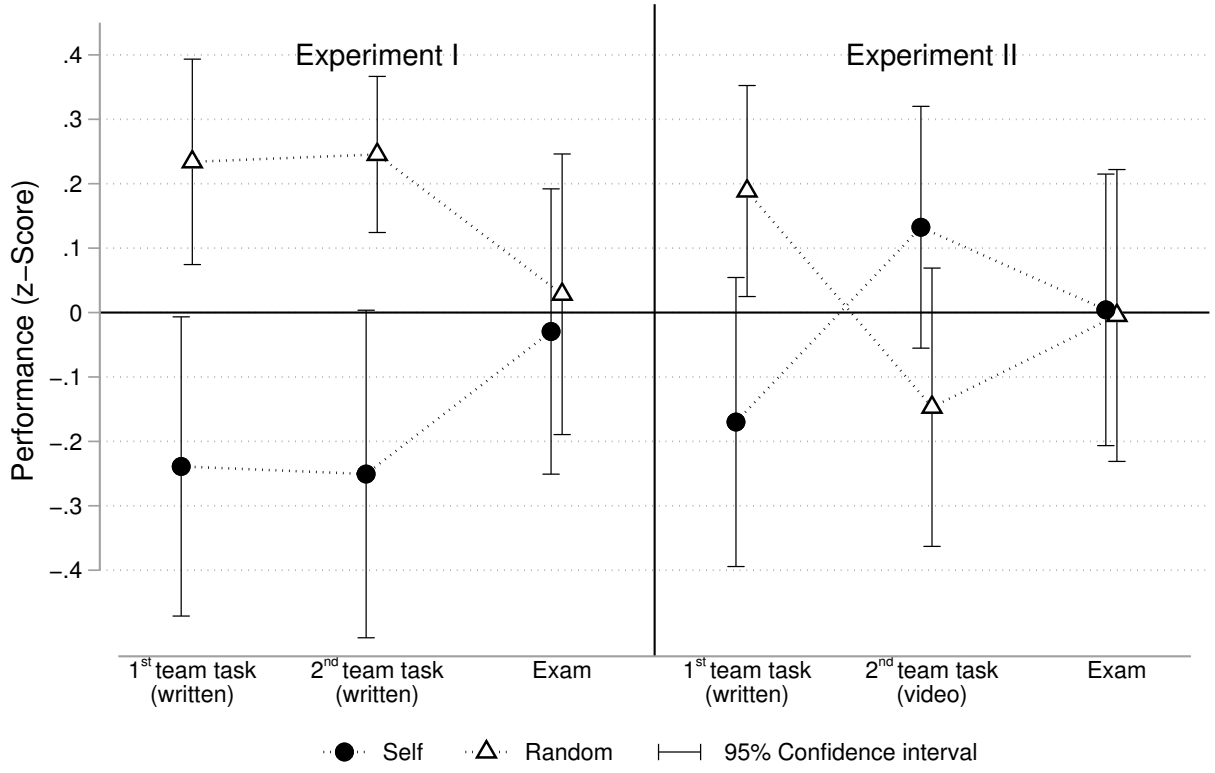

Figure shows the average team performance (z-standardized) for the tasks in our experiments. The left panel shows the results from Experiment I; the right panel, from Experiment II.

7 Tables

Table 1: Randomization checks

	Experiment I			Experiment II			Experiment I + II		
	Self	*Random*	*p-value*	*Self*	*Random*	*p-value*	*Self*	*Random*	*p-value*
Pre-experiment data									
GPA	.057	-.053	.195	.074	-.084	.523	.066	-.068	.134
% female	.287	.323	.593	.356	.220	.038	.323	.273	.282
Admissions test									
Admissions test score	-.024	.026	.533	.007	-.008	.911	-.008	.010	.753
Analytical	.022	-.010	.889	-.056	.064	.428	-.019	.026	.666
Presentation	.052	-.047	.549	.107	-.122	.125	.080	.083	.141
Interview	-.086	.086	.128	-.002	.002	.879	-.042	.045	.344
Discussion	-.046	.033	.688	-.017	.019	.932	-.031	.027	.744
Post-experiment data									
Pro-sociality	-.029	.009	.693	.008	-.010	.899	.001	-.011	.906

Descriptive statistics of pre-experiment data, admissions test scores and pro-sociality. GPA is inverted and z-standardized, with a higher GPA indicating better school performance. Analytical Test, Presentation, Interview, and Discussion are z-standardized. *p*-Values are from a Mann-Whitney U test (two-sided) comparing differences in mean ranks between both treatments. *p*-values for the comparison of % female are from a χ^2-test (one-sided). The correlation matrix appears in Table A.2 in the Appendix.

Table 2: Predicting team performance

	Dependent variable:					
	Performance on 1st team task (Exp. I and II: written)			Performance on 2nd team task (Exp. I: written, II: video)		
Independent variables	(1)	(2)	(3)	(4)	(5)	(6)
1 if *Self*	-0.42***	-0.47**	-0.48**	-0.11	-0.50**	-0.52***
	(0.14)	(0.20)	(0.20)	(0.14)	(0.20)	(0.20)
1 if Experiment II		-0.05	0.03		-0.39**	-0.32*
		(0.16)	(0.15)		(0.18)	(0.16)
1 if *Self* x 1 if Experiment II		0.11	0.03		0.78***	0.72**
		(0.28)	(0.28)		(0.29)	(0.29)
Controls						
GPA			0.12**			0.13**
			(0.06)			(0.06)
1 if female			-0.03			-0.28*
			(0.12)			(0.15)
Admissions test score			-0.02			-0.07
			(0.05)			(0.06)
Constant	0.21***	0.23**	0.24**	0.05	0.25***	0.34***
	(0.08)	(0.11)	(0.12)	(0.09)	(0.09)	(0.11)
Observations	382	382	377	382	382	377
R^2	0.04	0.04	0.07	0.00	0.04	0.07

Columns (1) - (3) show OLS regressions of z-standardized team performance on the first task. In both experiments students had to submit a written task. Columns (4) - (6) show OLS regressions of z-standardized team performance on the second task. In Experiment I, students have to submit a written task; in Experiment II students have to submit a video presentation. Control variables are GPA, admissions test score, and gender. GPA and admissions score have been z-standardized. Standard errors clustered on teams are in parentheses. Significance indicators: *** $p<.01$, ** $p<.05$, * $p<.1$.

Table 3: Self selection and composition of teams

Pre-experiment data	Observed *Self*	Simulation *Random*
GPA	.978 ***	1.117
% female	.204 ***	.4198
Admissions test		
Admissions test score	1.174	1.124
Analytical	1.012 **	1.135
Presentation	1.112	1.136
Interview	1.195	1.129
Discussion	1.154	1.129
Post-experiment data		
Pro-sociality	1.049	1.126

The table displays the average absolute difference between teammates on pre-experiment observables. Simulation Random denotes the average absolute difference for the respective variable from a simulation in which we pairwise match all students within a treatment within an experiment. Stars indicate the two-sided significance level of a Wilcoxon Signed Rank test comparing the observed score against the simulated value from Simulation Random. Significance indicators: *** $p < .01$, ** $p < .05$, * $p < .1$.

Table 4: Predicting self-selection (Treatment *Self*)

	Dependent variable:								
	Teammate's GPA			Teammate's Analytical test			Teammate's Pro-sociality		
Ind. variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Individual characteristics									
GPA	0.25**	0.23**	0.30**		-0.00	0.02		0.21**	0.18*
	(0.11)	(0.11)	(0.15)		(0.07)	(0.09)		(0.09)	(0.10)
Analytical test		-0.01	0.02	0.06	0.03	-0.01			0.15
		(0.08)	(0.10)	(0.07)	(0.07)	(0.12)			(0.12)
Pro-sociality			0.17*			0.12	0.12	0.15	0.17*
			(0.09)			(0.10)	(0.09)	(0.10)	(0.09)
Teammate characteristics									
GPA					0.30***	0.26***		-0.11	-0.09
					(0.06)	(0.08)		(0.10)	(0.10)
Analytical test		0.35***	0.30***						-0.10
		(0.07)	(0.09)						(0.12)
Pro-sociality			-0.08			-0.08			
			(0.09)			(0.10)			
Constant	0.04	0.04	-0.01	0.01	-0.01	0.13	0.02	0.02	0.03
	(0.11)	(0.11)	(0.14)	(0.10)	(0.10)	(0.12)	(0.13)	(0.13)	(0.13)
Observations	184	182	104	182	182	104	106	106	104
R^2	0.06	0.16	0.22	0.00	0.11	0.13	0.02	0.06	0.09

The table displays the coefficient estimates from specifications using data only from treatment *Self*. All variables are z-standardized. Robust standard errors are in parentheses. Significance indicators: *** $p<.01$, ** $p<.05$, * $p<.1$.

Table 5: Overview of survey items and survey results

Survey item	Experiment I *Random*		*Self*	Experiment II *Random*		*Self*	Experiment I + II *Random*		*Self*
Perceived quality of cooperation (1=Not agree, 7=Completely agree)									
We communicated a lot.	5.16	$<^{*}$	6.08	5.58	$<^{**}$	6.14	5.35	$<^{***}$	6.11
We helped each other a lot.	5.41	$<^{**}$	5.95	5.82	$<$	6.07	5.60	$<^{**}$	6.01
Both team members exerted effort.	5.46	$<^{**}$	6.07	5.93	$<^{*}$	6.34	5.67	$<^{***}$	6.20
Both team members contributed equally.	5.12	$<^{*}$	5.59	5.44	$<$	5.87	5.26	$<^{**}$	5.73
Our individual skills complemented very well.	4.99	$<^{**}$	5.56	5.26	$<$	5.52	5.11	$<^{**}$	5.54
Our team was very stressed.	2.87	$<^{*}$	3.30	2.53	$<$	2.69	2.71	$<$	3.00
Our team was very motivated.	5.53	$<$	5.78	5.70	$<$	5.85	5.61	$<$	5.81
The mood in our team was good.	5.79	$<^{*}$	6.10	6.19	$<$	6.27	5.98	$<$	6.18
The coordination of our team was very good.	5.03	$<^{*}$	5.38	5.56	$<$	5.59	5.27	$<$	5.49
I was dominant in leading the team.	4.49	$>$	4.22	4.30	$>$	4.21	4.40	$>$	4.22
One person was dominant in leading the team.	4.10	$>$	3.84	4.14	$>$	3.85	4.12	$>$	3.84
Attitude towards the other (1=Not agree, 7=Completely agree)									
My team member is a friend.	3.82	$<^{***}$	6.25	4.33	$<^{***}$	6.06	4.06	$<^{***}$	6.15
I knew the team member very well before the course.	2.60	$<^{***}$	6.19	2.93	$<^{***}$	5.66	2.75	$<^{***}$	5.93
Observations	68		73	57		71	125		144

Table reports descriptive statistics of student responses in the post-experimental survey. P-values stem from a two-sided Mann-Whitney U test for a comparison of averages between *Self* and *Random*. Significance indicators: *** $p<.01$, ** $p<.05$, * $p<.1$.

A Appendix

Table A.1: Average and variance of performance (z-Standardized)

	Experiment I					
	Self		*Random*		*p-Value*	
Performance	Avg	SD	Avg	SD	MWU	Leven's
Tot. team task	-.303***	1.214***	.297	.621	.007	.002
1^{st} team task (written)	-.239**	1.140	.234	.791	.011	.104
2^{nd} team task (written)	-.251*	1.248***	.245	.599	.068	.001
Exam	-.022	1.004	.028	1.002	.455	.995
	Experiment II					
	Self		*Random*		*p-Value*	
Performance	Avg	SD	Avg	SD	MWU	Leven's
Tot. team task	.049	.967	-.089	1.046	.451	.534
1^{st} team task (written)	-.196*	1.150	.183	.791	.064	.193
2^{nd} team task (video)	.114	.956	-.153	1.047	.156	.381
Exam	.004	.989	-.005	1.018	.984	.603

Descriptive statistics (z-Scores) of students performance in the experiment. Avg (SD) shows the average (standard deviation) of team performance for the team tasks at the team level and for the exam at the individual student level. MWU p-values stem from a two-sided Mann-Whitney U test for a comparison of averages between *Self* and *Random*. Leven's p-values are the results of a comparison of variances between both treatments. p-Values denote results from a Mann-Whitney U test (two-sided) comparing differences in distribution between both treatments. p-Value for the comparison of % female are the result of a χ^2-test. Significance indicators: *** $p<.01$, ** $p<.05$, * $p<.1$.

Table A.2: Pairwise correlations of pre-study data, admissions test score, and performance on team tasks and exam

	GPA	Female	Admissions	Analytical	Presentation	Interview	Discussion	Pro-sociality	Tot. team task	1^{st} Team task	2^{nd} Team task	Exam score
GPA	1.000											
Female	.135**	1.000										
Admissions total	.262***	-.123*	1.000									
Analytical	.205***	-.310***	.385***	1.000								
Presentation	.199***	.023	.522***	-.020	1.000							
Interview	.068	-.001	.660***	-.120*	.102*	1.000						
Discussion	.089	.036	.662***	-.115*	.092	.538***	1.000					
Pro-sociality	-.032	.101	.038	-.106	.091	.008	.107	1.000				
Tot. team task	.126*	-.056	.007	.056	-.020	-.000	-.035	.117	1.000			
1^{st} Team task	.101*	-.000	.016	-.011	.085	.013	-.062	.064	.566***	1.000		
2^{nd} Team task	.092	-.073	-.017	.073	-.070	-.029	-.024	.100	.919***	.220***	1.000	
Exam score	.320***	-.090	.132*	.291***	.170**	-.092	-.121*	.020	.122*	.127*	.078	1.000

Table displays correlation coefficients of pairwise correlations. Admissions total covers the aggregated score a student received on the admissions day. Analytical, Presentation, Interview, and Discussion are the score the student received for each respective task during the admissions day. Tot. team task, 1^{st} Task, 2^{nd}, and Exam are the score for the respective performance during the course. All scores are z-standardized. Table includes data from all treatments and experiments. Significance indicators: *** $p<.01$, ** $p<.05$, * $p<.1$.

Table A.3: Predicting random assignment (Treatment *Random*)

	Dependent variable:								
	Teammate's GPA			Teammate's Analytical test			Teammate's Pro-sociality		
Ind. variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Individual characteristics									
GPA	0.03	0.05	0.19		-0.08	-0.32**		0.14	0.15
	(0.08)	(0.08)	(0.11)		(0.08)	(0.15)		(0.10)	(0.10)
Analytical test		-0.08	-0.21**	0.01	0.03	0.16			-0.00
		(0.07)	(0.10)	(0.08)	(0.08)	(0.12)			(0.09)
Pro-sociality			0.15			-0.00	-0.13	-0.11	-0.11
			(0.10)			(0.13)	(0.11)	(0.12)	(0.12)
Teammate characteristics									
GPA					0.11	0.25*		-0.08	-0.10
					(0.08)	(0.14)		(0.10)	(0.11)
Analytical test		0.10	0.17*						0.05
		(0.07)	(0.10)						(0.10)
Pro-sociality			-0.09			0.08			
			(0.11)			(0.15)			
Constant	-0.05	-0.05	0.02	-0.02	-0.02	0.04	0.05	0.05	0.04
	(0.09)	(0.09)	(0.12)	(0.11)	(0.11)	(0.16)	(0.13)	(0.13)	(0.13)
Observations	180	178	82	178	178	82	82	82	82
R^2	0.00	0.02	0.13	0.00	0.02	0.10	0.02	0.05	0.05

The table displays the results of a OLS regression analysis (robust standard in parentheses). Specifications include data from treatment *Random* only. All variables have been z-standardized. Significance indicators: *** $p<.01$, ** $p<.05$, * $p<.1$.