

Can self selection create high-performing teams?

Roy Chen and Jie Gong*

This Version: Jan 2018

Abstract

Does the way that teams are formed affect their productivity? To address this question, we run an experiment comparing different methods of team formation: (1) random assignment; (2) self selection; and (3) algorithm assignment designed to maximize skill complementarity. We find that self selection creates high-performing teams. These teams perform better on a team task than randomly-assigned teams and as well as those assigned using the algorithm. Exploring the mechanism, we find evidence that, when given the choice, individuals self select into teams primarily based on their social networks and exert higher effort towards the team task.

JEL Classification: C93, D23

Keyword: group formation, teamwork, self selection, field experiment

*Chen: Department of Economics, Faculty of Arts and Social Sciences, National University of Singapore, AS2 #06-02, 1 Arts Link, Singapore 117570, ecscwr@nus.edu.sg. Gong: Department of Strategy and Policy, NUS Business School, National University of Singapore, 15 Kent Ridge Drive, Singapore, 119245, gong@nus.edu.sg. We would like to thank Yan Chen, Nancy Kotzian, Jin Li, Tracy Liu, Ivan Png, and seminar participants at the National University of Singapore, and the 2015 Economic Science Association World Meetings, Sydney for helpful discussions and comments, and Xuyuan Liu and Bo Shen for research assistance.

1 Introduction

Teamwork is prevalent in today’s workplaces. Over 80% of firms use teams as part of their organizational approach, and over 70% of firms use self-directed teams (Lazear and Shaw, 2007). Previous literature into teamwork studies how to make teams more productive, either assuming that the team formation process is exogenous, or focusing on a particular team formation method—e.g., self selection (Bandiera et al., 2013; Hamilton et al., 2003). An important but underinvestigated question is which group formation method firms should employ. In this study, we explicitly compare different team formation methods—i.e., self selection and principal assignment—in a field experimental setting.

Theoretically, it is not obvious which method is superior in creating productive teams. On the one hand, it is possible that workers have limited understanding of the technology of team production, or their incentives regarding team composition are not perfectly aligned with those of the firm. As a result, we may see a decrease in team productivity by self-chosen teams relative to those assigned by managers. On the other hand, if social connection plays a role in self selection, it may help mitigate free-riding behavior in teamwork, and therefore promote greater effort from members and lead to an increase in productivity. Socially-connected members may also know more information about each other, such as work style or personality, that is generally hidden to the firm but that affects team communication and coordination.

To investigate how individuals self select into teams and how these teams perform compared to teams formed through alternative processes, we randomly place subjects into one of three treatments in forming teams for a group project: 1) random assignment, 2) self selection, and 3) algorithm assignment designed to maximize skill complementarity. The random assignment serves as a baseline control, self selection allows the subjects to sort into teams freely, and the algorithm assignment is based on the premise that members with diverse skills may complement each other and therefore boost team productivity. We collect data on members’ pre-existing social networks, evaluate their project as the main outcome variable, and, after project completion, elicit information on how much time each member contributed (individual effort).

The results of our experiment show that the group formation process has a significant effect on group performance. When given the choice, subjects sort into groups based on social connections, at the expense of skill complementarity. Evaluating their group projects, we find that self-selected groups perform significantly better than those that are randomly assigned, and about as well as those that are assigned by the algorithm. These results provide evidence that self selection can create high-performing teams. Examining potential sources

of this superior performance, we find evidence that self-selected teams exert higher effort.

Our findings support self selection as an effective method for forming teams. The fact that self-chosen teams perform as well as algorithm-assigned teams provides evidence that social connections can compensate for a lack of skill complementarity. Furthermore, letting workers choose their own teams lowers the firms' expenses involved in collecting and processing information about individual workers, suggesting that decentralized group formation can be more efficient.

Our results have applicability to workplace team formation for several reasons. First, our subjects are representative of real Singaporean workers because they are business students at the National University of Singapore who will soon be joining the workforce. Second, our group formation treatments reflect possible firm approaches to team assignment. In our algorithm treatment, we obtain the same type of skill information as that obtained by firms during the interview process. In the self-chosen treatment, we mimic a decentralized group formation process that firms can undertake. Finally, the group project reflects tasks undertaken in real organizations.

This paper adds to previous work in the composition and performance of teams. Hamilton et al. (2003) find that heterogeneous teams in a garment factory in California are more productive than teams with members of homogeneous ability. Hoogendoorn et al. (2014) examine undergraduates in the Netherlands and find an initial increase in productivity for teams with greater ability dispersion, with an ultimate productivity decline as dispersion becomes too wide. In comparison, we examine skill complementarity in multiple dimensions. That is, in our algorithm treatment, we utilize four different skills instead of one. In this way, our study better mirrors the economic theory approach to team configuration highlighted by Lazear and Shaw (2007).

Bandiera et al. (2013), the closest study to the current paper, examine how strengthening team incentives affects the composition and performance of worker-chosen teams. They find in a field experiment with U.K. fruit pickers that with stronger incentives, workers are more likely to choose their groups based on ability rather than social connections. Our observation that self-chosen teams tend to sort based on social networks is consistent with their findings. However, our primary research question and design are different from theirs. While the teams in their study are all chosen by the workers themselves, we compare the outcomes of self-chosen groups and those exogenously assigned by a principal. Also, unlike their within-worker experimental design, our experiment is run between subjects in a randomized controlled trial.

Our results further lend support to previous findings that endogenous group formation or governance can foster performance. For example, Chen (2017) finds that coordination

in a minimum-effort game in a laboratory setting is improved when subjects are allowed to choose their own groups. Similarly, Blasco et al. (2013) find that coders allowed to choose their own groups in an online field experiment perform better on a coding task. Our results are consistent with the existing evidence that having the right to choose group features fosters cooperation.

This paper also makes useful contributions to the literature of social identity and group membership. Evidence from the social identity literature in economics suggests that people are less likely to shirk their group responsibilities when they feel more attached to a group (see, for example, Akerlof and Kranton, 2000; Eckel and Grossman, 2005; Chen and Li, 2009). In addition, Pan and Houser (2013) observe that the group formation process can affect not only ingroup trust, but also outgroup trust. By allowing people to choose their own groups, attachment and trust among group members seems to increase, improving group outcomes. As emphasized by Goette, Huffman and Meier (2012), separating labeling and the effects of social ties is important to understanding the micro-foundations of social group preferences. In addition, Goette, Huffman and Meier (2006) use random assignment of officers in the Swiss Army to platoons and find that group membership affects individuals' willingness to cooperate unselfishly, enforce norms, and punish outsiders vindictively. Our results add field experimental evidence to the value of social ties in within-group cooperation.

Lastly, our findings contribute to the extensive literature on decentralization in large organizations. Within this field, most studies examine decentralization from a mechanism design standpoint (see Mookherjee, 2006 for a comprehensive survey). These studies emphasize that decentralization provides benefits through improved communication, but may create costs arising from the principal-agent problem. However, there are fewer empirical examinations of the impact of decentralization on productivity. Ichniowski et al. (1997) and Ichniowski and Shaw (1999, 2003) document that U.S. businesses have increased the use of innovative human resource management practices, delegating production decisions to worker teams, and find that this has a significant impact on productivity. This study contributes to the literature by providing field experimental evidence on the benefits of decentralizing the group formation decision.

2 Experimental Design

In this section, we discuss the setting and design of our experiment, as well as the data. Our experiment addresses the question of how different group formation protocols impact group performance. Our subject pool consists of 685 students in a large undergraduate class at the National University of Singapore. The class meets in both a large lecture format, delivered

weekly for 13 weeks, and smaller discussion sections, which meet weekly starting in the third week of lecture. Each student attends one of 28 discussion sections for the duration of the class.¹ The discussion sections, led by teaching assistants, primarily review the material presented in lectures.

As part of their course requirements, students must present a group project in their discussion section worth 25% of their final grades in the course. This group project has been a course requirement since 2011, so the students experience a very similar classroom environment to previous cohorts.² Given the importance of university grades in Singapore for future career prospects, our subjects have a strong incentive to perform as well as possible on the project.³

Table 1: Experimental Design

Treatment	Number of sections	Number of 4-person groups	Age	Gender (Male=1)	Race (Chinese=1)	Birthplace (Singapore=1)
Random	8	46	19.90	0.45	0.92	0.75
Algorithm	10	56	19.88	0.43	0.93	0.78
Self Chosen	10	58	20.11	0.47	0.92	0.83

We run the experiment with three treatments, with all subjects in the same section also in the same treatment. This is done to reduce the possibility that subjects become aware of the other treatments and adjust their behavior. We randomly assign different sections into one of three treatments, stratifying the randomization based on section popularity and tutor. A summary of our experimental design is displayed in Table 1. The treatments appear to be balanced in terms of subject characteristics. Using Kruskal-Wallis tests, we cannot reject the hypothesis of an equal distribution of these demographics across the treatments.⁴ With four subjects per group and a total of six groups per section, each treatment has around 52-58 groups and 9-10 sections.⁵ In the “Random” treatment, students are placed into four-person

¹A university-wide balloting system assigns students to sections based on their submitted preference rankings. In our randomization procedure, we account for the possibility that some sections may be more popular than others.

²This creates an “ideal experiment” as described by Harrison and List (2004), in that the subjects do not perceive any of our treatments as being unnatural, or that an experiment is taking place.

³Pan et al. (2015) find that a 1 unit increase in cumulative average points at the National University of Singapore (an increase in grade average from B- to B+ or B+ to A) is associated with a 12.3% increase in monthly salary.

⁴We also run a multinomial logit model of treatment on all subject characteristics, and find that none of the individual coefficients are statistically significant and they are also jointly insignificant.

⁵Since not all sections have exactly 24 students, some groups may consist of sizes other than four students. We exclude all groups of size other than four from our analysis (45 students total).

groups randomly. In the “Self-chosen” treatment, students are asked to choose their own groups of four people from the students in their section.

In the “Algorithm” treatment, we assign students to four-person groups based on their responses to a demographic survey, administered to all students during the first lecture. In particular, one of the questions in this survey asks students to rank their own proficiency in four different skills that we identify as important for the group project, also highlighted in the rubric given to the students for the group project: presentation, research, quantitative analysis, and economic theory. Since our “Algorithm” treatment is based on the premise that skill complementarity benefits team performance, an ideal group under this treatment would consist of subjects with distinct highest-ranked skills.⁶ We consider the group formation process a linear assignment problem in the spirit of Munkres (1957). The solution to this assignment problem maximizes, at the section level, the sum of individuals’ rating for their assigned roles. Specifically, for a section of 24 subjects, we assign each subject to one of four roles (i.e., each section has 6 presentation “specialists”, 6 research “specialists”, etc.). We then randomly choose one subject from each role to create a group. This method is Pareto optimal in the sense that no subject can change groups to increase a group’s total ratings without decreasing another group’s total ratings.⁷⁸ While the algorithm assigns a role to each group member, we do not inform the groups of, or require members to perform, these assigned roles.⁹

Our experimental design allows us to test for several effects of group formation. First, by comparing the “Random” and “Self-chosen” treatments, we can observe how people self select into teams and whether they perform better than those who are randomly assigned to groups. Second, we can verify whether skill diversity affects group performance by comparing the “Random” and “Algorithm” treatments. Subjects in these two treatments experience identical experimental procedures (since these subjects are all given a group assignment with no further explanation), so any differences in performance can be attributed to the group assignment process.

Table 2 summarizes our experiment timeline as well as our data. In the first lecture of the semester, all students in the course complete a demographic survey (see Appendix A.1),

⁶Note that this ranking is in terms of their own skills and not as compared to other students. One may be concerned that these are measured “within person” and therefore cannot capture the absolute level of skills for each member. While direct measures of skills are unavailable, the algorithm in fact achieves a scenario where each member performs his or her comparative advantage.

⁷See the proof in Appendix A.5.

⁸For more details on the implementation of this algorithm, see Appendix A.6.

⁹In other words, we create teams by maximizing “potential” complementarity between skills. Members may not play the roles that our algorithm assigns, which might hinder teams in their ability to achieve their maximum productivity and attenuate the treatment effects. If we inform them of the roles, however, the treatment would be a combination of assigned groups and assigned roles, thus complicating the interpretation.

Table 2: Experiment Timeline and Data Collected

Week	Task(s)	Data collected
1	In-class survey	Skill rankings Demographics
3	Group assignment Topic 1 announcement	Network information
5	Topic 2 announcement	
8	Topic 1 presentations	Topic 1 performance
10	Topic 2 presentations	Topic 2 performance
16	Final exam	Effort

which includes a question regarding their skills. At the first section meeting (Week 3 of the lecture), an experimenter assigns the subjects to four-person groups (or asks them to form their own groups).¹⁰ After all groups are formed, subjects fill out a short “network survey” indicating if they know any of their teammates outside of class (see Appendix A.2). Then each group is randomly assigned a project topic. With six groups per section, three groups present Topic 1 in Week 8 of the discussion section and three groups present Topic 2 in Week 10. The presentations are given in two different weeks due to time constraints.¹¹ Topic 1 is announced at the end of Week 3 and Topic 2 is announced at the end of Week 5. All groups therefore have 5 weeks to complete their project.

The groups’ project performance constitutes our main outcome variable. We hire two research assistants (RA) to assess the performance of every group in every section, so that the evaluation is consistent. During Weeks 8 and 10, the two RAs attend each section to view the group presentations. They score each presentation on 10 different questions in 4 categories, corresponding to the 4 skills. The RAs’ rubric is included in Appendix A.3. These assistants are not told the details of the experiment. In particular, they are unaware of each section’s treatment assignment, so their scores cannot be affected by anticipated treatment differences.

Finally, to assess subject effort, we ask each individual to report the number of hours each group member spent working on the assignment. In order to incentivize subjects to tell the truth, we set up a game in the style of a Keynesian beauty contest. In this contest, subjects are given points based on how well their reported hours match the average reported hours of

¹⁰To ensure consistency in the delivery of this information, the same experimenter attends all section meetings that week. To minimize possible experimenter demand effects, the experimenter reads from a script when speaking to the subjects.

¹¹The two topics concern car taxes and the declining birth rate, respectively. In our analysis, we control for topic fixed effects so that we can effectively compare the performance of groups presenting on the same topic.

the other members of their group. In other words, subjects are incentivized to report what they believe others will report. To prevent collusion, we ask this question to subjects during the final exam for the class. Although this game has many Nash equilibria (where everyone’s reports match exactly), the instructor, using a similar example in a lecture, emphasized that the case where everyone tells the truth is a focal point. We also tell subjects on the final exam that “it is a Nash equilibrium for everyone to report the actual number of hours each person worked on the project” (Appendix A.4).

3 Results

In this section, we first present our main results for the effects of group formation on team composition and performance. We then examine a possible channel by analyzing how members’ effort varies across the different treatments.

Our analyses are based on our data for the four-person groups, which include 640 subjects across 160 groups.¹² In addition, we cluster standard errors at the section level to control for any possible dependency of decisions and performance across groups within a section. All non-parametric tests are two-sided Mann-Whitney U tests, and the unit of observation is the group.

3.1 Group Composition

In the “Self-chosen” treatment, we examine whether subjects sort into groups based on network connections, skill complementarity, or demographic characteristics. Figure 1 presents the level of network connections and skill complementarity within a group by treatment. A subject’s number of connections is the number of people she reports she knows outside the class. In order to calculate skill complementarity, we apply the Hungarian algorithm to find each member’s optimal role and sum the rank the subject assigns to that role. Therefore, a smaller number indicates better skill complementarity.

Figure 1 shows that subjects tend to sort themselves into groups based on social connections rather than skill complementarity. Specifically, we see that self-chosen groups have more connected members than either randomly-assigned or algorithm-determined groups (Figure 1a). This comparison (network in self-chosen groups $>$ network in control) is signif-

¹²If a section does not have exactly 24 subjects due to either original enrollment or subsequent dropping of the course by a student, then that section will have some non-four-person groups. It is possible that different treatments have different dropout rates, but we test the likelihood of having fewer than four subjects in a group and find no differences across treatments. Due to unforeseen circumstances, we do not have the grades for one section. We exclude this section from the regression sample.

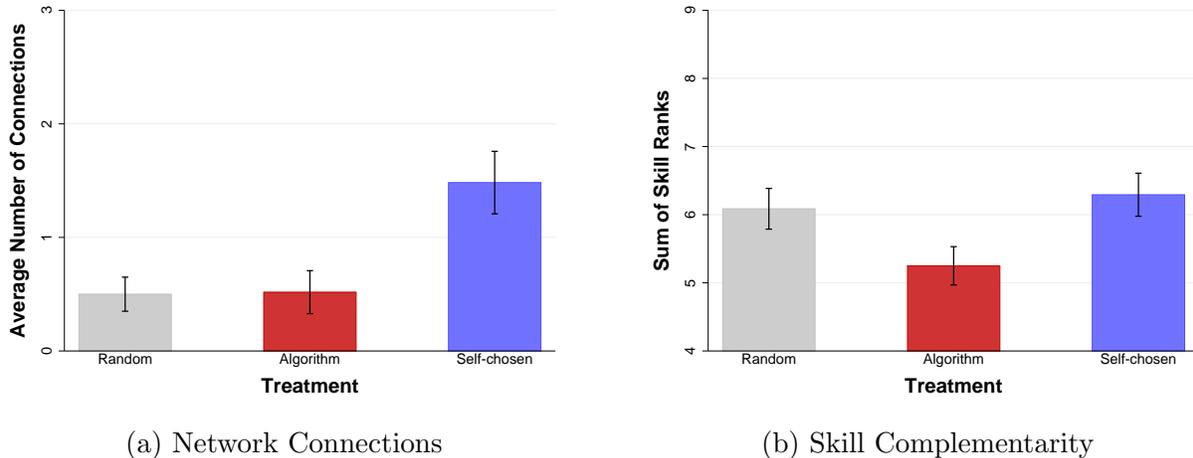


Figure 1: Group Composition by Treatment

icant ($p < 0.01$). However, the comparison between the randomly-assigned and algorithm-determined groups is not significant ($p = 0.46$). On average, a subject in the self-chosen treatment knows one more person in her group than a student in either the random or algorithm treatment. Figure 1b shows that the algorithm-determined groups have greater skill complementarity than either the self-chosen or randomly-assigned groups ($p < 0.01$ for both comparisons). However, there is no significant difference in skill complementarity between the randomly-assigned and self-chosen groups ($p = 0.45$). From this analysis, we can see that subjects who choose their groups rely on friendship networks while ignoring the skills of the other members.

We next run a series of OLS regressions across groups, accounting for any TA fixed effects in all analyses. The estimated treatment effects on group composition are presented in Table 3. Our dependent variables include the following group characteristics: (1) the average number of connections a member has within her group; (2) the sum of skill ranks; (3) the average student age; (4) the proportion of men in the group; (5) the proportion of business majors in the group; (6) the proportion of Chinese in the group, and (7) the proportion of individuals born in Singapore in the group. Our independent variables include the self-chosen and algorithm treatment dummies, with randomly-assigned groups as the reference.

The results of our regressions confirm our observations from Figure 1. First, we find that self-chosen groups have more network connections (Table 3, column 1) but about the same level of skill complementarity (column 2) as randomly-assigned groups. As for demographic characteristics, our results show that subjects do not sort into groups based on age, gender, discipline of study, ethnicity, or birthplace. We further find that our algorithm-determined groups exhibit greater skill complementarity—a result of our experimental design—but sim-

ilar demographic patterns to our randomly-assigned groups. This finding reinforces our observation that skills are not directly correlated with any of our observed demographic characteristics.

Table 3: OLS: Group Composition by Experimental Treatment

	(1) Network	(2) Skill	Demographics				
	connections	complementarity	(3) Age	(4) Male	(5) Business	(6) Chinese	(7) Native
Self-chosen	1.009*** (0.153)	0.213 (0.155)	0.175 (0.143)	0.017 (0.051)	-0.029 (0.041)	0.004 (0.039)	0.077 (0.057)
Algorithm	0.044 (0.139)	-0.829*** (0.243)	-0.059 (0.131)	-0.025 (0.057)	0.011 (0.029)	0.020 (0.026)	0.036 (0.058)
Constant	0.481*** (0.138)	6.158*** (0.242)	19.966*** (0.104)	0.459*** (0.039)	0.896*** (0.030)	0.896*** (0.037)	0.798*** (0.043)
Observations	160	160	160	160	160	160	160
R^2	0.268	0.158	0.049	0.022	0.044	0.023	0.048

Notes: 1. We control for TA fixed effects.
2. Robust standard errors clustered by section are in parentheses.
3. Significant at * 10%; **5%; *** 1%.

3.2 Treatment Effects on Group Performance

We next examine whether group performance differs across treatment. To measure group performance, we have our research assistants score each group on a 100 point scale.¹³ These RAs are unaware of the experimental manipulation, or the treatment assigned to each section. We take the RAs’ average scores as the measure of each group’s performance. Each group’s total score is assessed across four categories: research, theory, statistics, and presentation, with each category accounting for 25% of the group’s total score. We present the rubric used by the RAs in Appendix A.3.

Figure 2 presents the total and category scores for the groups in our experiment. From Figure 2a, we see that our self-chosen groups perform better than our randomly-assigned groups and approximately as well as our algorithm-determined groups. We also find that both our self-chosen and algorithm-determined groups perform better than our randomly-assigned groups ($p = 0.052$ and $p = 0.045$). The self-chosen and algorithm-determined groups’ scores are not significantly different ($p = 0.94$). The results in Figure 2b show that this superior performance is observed across all four categories, with the largest difference observed in the statistics category ($p < 0.01$ comparing either the self-chosen or algorithm-determined group to the randomly-assigned group).

¹³We find a Pearson’s correlation of $r = 0.794$ between our two RAs, which indicates good correlation between their scoring of the groups.

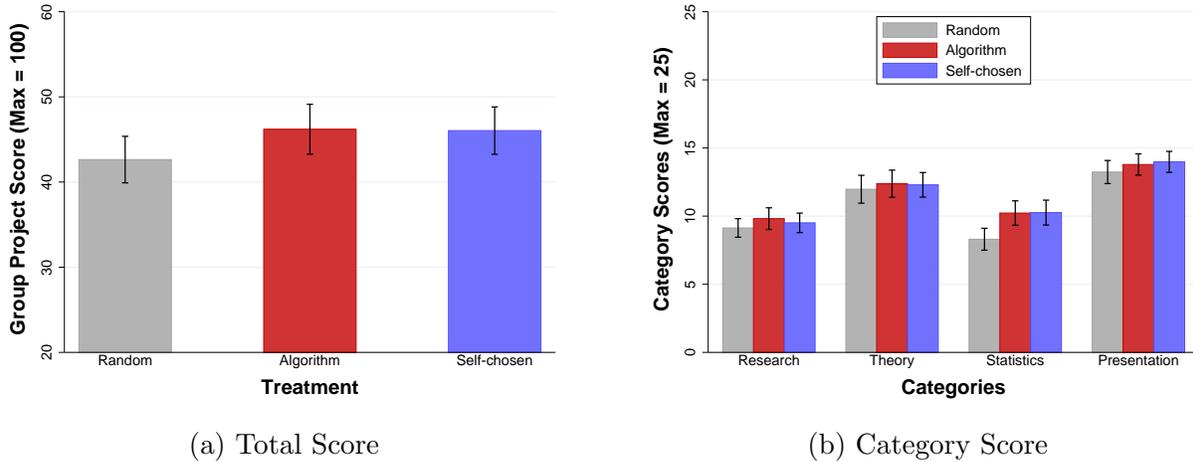


Figure 2: Group Composition by Experimental Treatment

Table 4: OLS: Treatment Effects on Group Performance

	(1) Total	(2) Research	(3) Theory	(4) Statistics	(5) Presentation
Self-chosen	3.285** (1.484)	0.392 (0.347)	0.293 (0.621)	1.832*** (0.595)	0.769* (0.386)
Algorithm	3.250* (1.788)	0.682 (0.420)	0.328 (0.704)	1.760** (0.690)	0.480 (0.474)
Constant	51.224*** (3.302)	9.694*** (1.042)	15.47*** (0.964)	10.48*** (1.319)	15.58*** (0.813)
Observations	160	160	160	160	160
R^2	0.226	0.196	0.212	0.187	0.192

Notes: 1. We control for TA, presentation order, topic and time fixed effects.
2. Robust standard errors clustered by section are in parentheses.
3. Significant at * 10%; **5%; *** 1%.

Table 4 reports the OLS estimates of our treatment effects on group performance. Column 1 presents the results of treatment on the total performance score while columns 2 to 5 break down the score by category. All regressions control for TA, topic (topic 1 or 2), section meeting day (Monday to Friday) and time (morning or afternoon) fixed effects, and presentation order within a section. Across our three groups, we find that both the self-chosen and algorithm-determined groups outperform the randomly-assigned groups, with the average treatment effect about 3.285 points higher for the total score, significant at the 5% level. Evaluating at the mean score of random groups (42.636), we find that this effect is equivalent to about a 7.7% increase in performance. The results in columns 2 to 5 show positive treatment coefficients for all four categories, with the effect on “statistics” performance estimated very precisely. These results suggest that all four categories contribute to the superior performance of our treatment groups, and that “statistics” performance is the main contributor to overall group performance.

Our results regarding statistics performance may reflect student perceptions of the difficulty of the statistical element of the project. In our student skill survey, we find that statistics is most often ranked as the skill in which students feel least proficient (57.8% of students), and least often ranked as the skill in which they feel most proficient (9.5% of students).¹⁴ These relatively low self-rankings of the statistics skill lead to a dearth of group members who can specialize in the statistics role. For our algorithm treatment, we are able to spread out the group members who list statistics as one of their top skills into different groups. In fact, 73% of our algorithm-determined groups have at least one member who ranks statistics as her first or second best skill, compared to only 54% (64%) of our randomly-assigned (self-chosen) groups.¹⁵

Lastly, we conduct a Wald test comparing the coefficients of our “Self-chosen” and “Algorithm” treatments. The results of this analysis indicate that we cannot reject the null hypothesis that the two coefficients are the same ($p > 0.5$ in columns 1 to 5). This reinforces our conclusion that our self-chosen groups perform as well as our algorithm-determined groups.

Table 5: OLS: Treatment Effects on Group Members' Effort

	(1)	(2)	(3)	(4)
	log(Man-hours)	Coef. of variance	Minimum hours	Maximum hours
Self-chosen	0.124* (0.067)	-0.015 (0.017)	0.185 (1.027)	0.247 (1.022)
Algorithm	0.087 (0.079)	-0.009 (0.015)	0.301 (1.104)	0.553 (1.224)
Constant	4.002*** (0.777)	0.116 (0.108)	5.987 (6.654)	6.312 (7.115)
Observations	615	615	615	615
R^2	0.063	0.056	0.044	0.053

- Notes:*
1. We control for student demographics.
 2. We control for tutor, presentation order, topic and time fixed effects.
 3. Robust standard errors clustered by section are in parentheses.
 4. Significant at * 10%; **5%; *** 1%.

3.3 Effort

Finally, we examine whether differences in group performance can be explained by differences in effort across the treatment groups. Using the reported hours worked from the effort survey, we calculate group effort by adding the hours reported by each student for each group member, i.e., her perceived man-hours worked by the group. Students have an incentive to report the actual hours worked, but they may not perfectly observe each other's effort if they spend some time working alone on the project. However, any measurement error due to not observing hours would tend to bias our estimates downward, i.e., work against finding any differences in effort across groups.

Table 5, column (1) presents the estimates of a set of OLS regressions. The dependent variable is the logarithm of total man-hours.¹⁴ Note that we include the same group characteristics as in Table 4 as well as individual demographic characteristics such as age, gender, discipline of study, ethnicity, and birthplace. We find that self-chosen groups invest 12.4%

¹⁴We compare these self-assessments to those for the other three skills: 1) presentation (28.1% rank it first, 20.2% rank it last), 2) research (26.7% rank it first, 13.4% rank it last), and 3) economic theory (35.6% rank it first, 8.6% rank it last).

¹⁵This lack of group members who consider themselves to be proficient in a specific skill only substantially affects the statistics skill. For the other three skills, at least 90% of groups in each treatment have a member who ranks the respective skill as either first or second.

¹⁶We use the log transformation because the reported hours are lognormally distributed. Meanwhile, it automatically drops observations with zero hours. These are unlikely to be the actual effort exerted, given that every group, at minimum, prepares presentation slides.

more hours on the project than do randomly-assigned groups. This effect is significant at the 10% level.¹⁷ Evaluating at the mean of the control (37.83 man-hours), the difference is about 4.69 hours per group. We find no significant difference in total effort between our algorithm-determined and randomly-assigned groups.

To gain further insight into effort as a mechanism for the effect of group formation on performance, we examine the dispersion of effort among group members. We measure dispersion as the coefficient of variation (the ratio of the standard deviation to the mean), and minimum and maximum reported hours across the four members of a group. As shown in Table 5, we find no significant difference between our treatment and control groups in the distribution of effort within a group.

The treatment effect on effort shows that self-chosen groups exert higher effort than those in the control. Does the higher effort explain their superior performance? We conduct a causal mediation analysis and find that around 16% of the treatment effect on group performance works through the effort channel.¹⁸ Again, this ratio might be attenuated due to measurement error in the students' reported effort. It also suggests that there are other sources of the better performance by self-chosen groups. For instance, it is possible that socially-connected members work more efficiently together, coordinate more smoothly, and have lower communication costs. Unfortunately, we do not have relevant data on these factors. Future work can explore how self selection may affect members' efficiency to uncover the unexplained part of their superior performance.

4 Conclusion

In this study, we experimentally manipulate how groups are formed in order to examine the effect of group formation on group composition, effort, and performance. Our results show that people tend to select groups based on social connections rather than member skills. We further find that these self-chosen groups perform better than do randomly-assigned groups, and about as well as algorithm-determined groups with optimal skill complementarity. We also find some evidence that self-chosen groups exert a higher amount of effort, partly explaining their superior performance. These findings are consistent with the prediction that

¹⁷If we exclude one outlier who reports 100 hours per member, the estimate increases to 13.8% and is significant at the 5% level.

¹⁸The exercise consists of three steps: First, as shown in Table 4, column 1, the total effect of self selection on team performance is 3.285. Second, we include team effort in the performance regression and find a coefficient of 4.2, i.e., the impact of effort on team performance. Finally, we take the product of the coefficient and the change in team effort induced by self-selection, 0.124 (Table 5, column 1). This produces the effort-induced change in the team performance. In other words, about $4.2 \cdot 0.124 / 3.285$, or 16% of the treatment effect works through its impact on effort.

social connections can mitigate free riding behavior in teamwork.

Regarding the decentralization of group formation in the workplace, our findings suggest that allowing employees to form their own workgroups can lead to a similar level of productivity as manager-determined groups. Delegating group formation may take advantage of hidden information, such as social connections, which are valued in group work. Furthermore, taking into account the costs associated with collecting worker information and designing formation rules, self-chosen groups provide cost savings for a firm.

In theory, a centralized mechanism should be able to mimic the outcome of any decentralized system, a direct implication of the Revelation Principle. However, the validity of this theory relies on the absence of any communication or information processing costs. In contrast, in actual organizational contexts, groups are formed based on rich information about participants; this information is collected at a cost. The choice between centralization and decentralization therefore reflects a trade-off between the communication costs involved in obtaining information when forming groups and the potential incentive problems if workers act in their own self-interest when forming their own groups. When decision-making is delegated to the workers, our findings suggest that they sort by friendship, which in turn may overcome the incentive problems in teams, making decentralization the superior organizational arrangement.

There are several extensions to this work that are worthy of exploration. First, while we find that self-chosen groups seem to prioritize social connections, it is not clear if these increased connections are the only contributor to the increased performance in these groups. It is possible that the freedom to choose one's own group also motivates the students to perform better. Future work can further explore the benefit of autonomy per se and the effects of working with friends. Second, while our subjects only work together on a single task, a possible extension is to add repeated interaction to examine whether workers sort and perform differently. Finally, it should be noted that the effectiveness of self-chosen teams may depend on other policies adopted by the firm. For example, Bandiera et al. (2013) find that self-chosen teams may select differently depending on the strength of their incentives. Together with our results, these findings suggest that a firm that adopts a decentralized group formation policy must also carefully choose the incentive scheme that it employs.

A Appendices

A.1 Skills Survey

BSP 1005 Survey

Please answer the following questions.

1. Name: _____

2. Matriculation number: _____

3. Home Faculty: _____

4. What is your age? _____

5. What is your gender? Male Female

6. Which of the following best describes your racial or ethnic background?

Chinese Malay Indian Other

7. Were you born in Singapore? Yes No

8. What was the most recent school you attended before joining NUS?

(Examples: Victoria JC, Nanyang Polytechnic, etc.)

9. Have you taken any economics classes before? If so, at what level did you take your most advanced class? (Examples: JC, Polytechnic, etc.)

10. At what level did you take your most advanced mathematics class?

(Examples: JC, Polytechnic, etc.)

(See next page →)

11. A major grading component of BSP 1005 is a group project. The following four skills are essential for completing the project. **Please rank these skills by how competent you feel in each.** This information may be shared with your potential group members.

- (1) Presentation skills: ability to effectively communicate and present information
- (2) Research skills: ability to find relevant background information and data
- (3) Quantitative skills: ability to use statistical tools (such as Excel) to analyze data
- (4) Economic theory skills: ability to apply what you have learned in the module to the project

Example: If you feel you are best at economic theory, next best at quantitative methods, next best at presenting, and worst at research, then you would say that your own ranking is:

4 > 3 > 1 > 2

Your own ranking:

___ > ___ > ___ > ___

A.2 Group Network

Group 1

Do any members of your group know each other outside of this module?

<hr/>	<hr/>
●	●
●	●
<hr/>	<hr/>

A.3 Presentation Rubric

Category	Item	Description	Score
Research	Background	policy, institutional, and cultural details	
	Evidence	how relevant and authoritative are their arguments?	
	Data Collected	amount, scope and usefulness of data collected	
Theory	Framework	unifying framework?	
	Economic Analysis	how reasonable is their analysis?	
Statistics	Methods	how detailed and appropriate is their econometrics?	
	Statistical Analysis	how good is their answer and explanation?	
Presentation	Flow	how well did their presentation flow?	
	Professionalism	were they and their slides professional?	
	Engagement	was it interesting?	
Factual Questions	How many people spoke during the presentation?		
	How did they divide the presentation?		By topic / section / time
	How much time did they take?		<15 / =15 / >15

A.4 Effort Elicitation

Recall the flat tire example we discussed in lecture. This question is a similar task regarding your group project.

- Please write down the number of hours each person in your group spent on the group project using the table below. As long as you fill out this table, you will receive 2 marks.
- You can receive up to 4 more marks based on how well your reported hours match the *average* hours your group members report. The better you match the average, the more marks you receive.

Just as in the flat tire example, it is a **Nash equilibrium** for everyone to report the actual number of hours each person worked on the project. **Your responses to this question will not affect anyone's grade on the group project.** (For the exact method we will use to calculate your marks, see the next page).

Name of Group Member	Number of Hours
You	_____ hours
_____	_____ hours
_____	_____ hours
_____	_____ hours

The number of marks you receive is determined by the following procedure:

1. For each member of your group, we will compare two numbers: a) the number reported by you for that member and b) the average number reported by the others in your group for that member.
2. We calculate the absolute difference between these two numbers.
3. We now have this difference for each member of your group, and we take the maximum of these differences.
4. The number of marks you receive is based on this maximum difference:
 You receive 4 marks if this difference is at most 1 hour.
 You receive 3 marks if this difference is between 1 hour and 2 hours.
 You receive 2 marks if this difference is between 2 hours and 4 hours.
 You receive 1 mark if this difference is between 4 hours and 6 hours.
 You receive 0 marks if this difference is more than 6 hours.

Example:

Group Member	Your Report	Average of Others' Reports	Difference
You	\underline{v} hours	a hours	$ v-a $
Member 2	\underline{w} hours	b hours	$ w-b $
Member 3	\underline{x} hours	c hours	$ x-c $
Member 4	\underline{y} hours	d hours	$ y-d $

Call the maximum of these differences $D = \max\{|v-a|, |w-b|, |x-c|, |y-d|\}$. You will receive marks P based on the value of D :

$$P = \begin{cases} 4, & \text{if } D \leq 1 \\ 3, & \text{if } 1 < D \leq 2 \\ 2, & \text{if } 2 < D \leq 4 \\ 1, & \text{if } 4 < D \leq 6 \\ 0, & \text{if } D > 6 \end{cases}$$

A.5 Pareto Efficiency of the Algorithm Treatment

Here we prove that our method of group formation for the algorithm treatment is Pareto optimal, in the sense that no students can change groups to increase a group's total ratings without decreasing another group's total ratings.

Proof. Suppose that there exists a change in the formation of groups that makes one group better off without making any other groups worse off. Let us start with a simple case in which the change involves two students: student i from group m and student j from group n . Without loss of generality, assume that after swapping the two students, group m has strictly higher skill ratings and group n has weakly higher ratings. The other four groups in the section remain unchanged.

After swapping the two students, group m and group n each experience a change related to the roles within each group. If every student performs the same role in the new group as initially assigned, the skill ratings for each member remain the same and so do the total ratings for each group.

If the two students change their roles upon entering their new groups, this leads to strictly higher ratings for group m and weakly higher ratings for group n . This will increase the students' ratings for their roles at the section level, contradicting the optimal assignment solved by the Munkres algorithm, which already maximizes the sum of all students' ratings at the section level.

For changes involving more than two students, the proof is similar: if the roles within groups remain the same, there is no improvement in groups' total ratings. If roles change, then there is a new assignment of students to roles such that, at the section level, total ratings are improved. This result again contradicts the optimal assignment solved by the Munkres algorithm.

A.6 Implementation of the Algorithm Treatment

For our algorithm, we treat group formation as a linear assignment problem, which is a classic problem in combinatorial optimization, as described by Munkres (1957):

The personnel-assignment problem is the problem of choosing an optimal assignment of n men to n jobs, assuming that numerical ratings are given for each man's performance on each job. An optimal assignment is one which makes the sum of the men's ratings for their assigned jobs a maximum.

For each section, we have 24 students, whom we must assign to 24 jobs. We treat these jobs as 6 copies of 4 skill-based roles (i.e., there are 6 presentation roles, 6 research roles,

6 quantitative roles, and 6 economic theory roles). The output is a role assignment for all 24 students, with 6 students assigned to each role. This calculation is performed by the Munkres/Hungarian algorithm, where the input is the cost for each student-job combination and the algorithm must minimize overall costs.

The cost of each student-job combination is determined such that the roles corresponding to the student's best skill have the lowest cost, the roles corresponding to the student's second best skill have the next lowest cost, and so on. In particular, we give the rank of 1 a cost of 0, the rank of 2 a cost of 1, the rank of 3 a cost of 25, and the rank of 4 a cost of 601. This ensures that, when given the option of placing 1 student in her third-ranked role (with all others in their best role), or placing all students in their second-ranked role, the algorithm will choose the latter option.

Once the algorithm gives us each student's job assignment, we form groups by randomly placing one student from each role into a group, so that each group has 1 student assigned to each different role. This process should, in principle, create groups that are balanced with a diverse set of skills.

References

- Akerlof, George A. and Rachel E. Kranton**, “Economics and Identity,” *The Quarterly Journal of Economics*, August 2000, *115* (3), 715–753.
- Bandiera, Oriana, Iwan Barankay, and Imran Rasul**, “Team Incentives: Evidence From a Firm Level Experiment,” *Journal of the European Economic Association*, October 2013, *11* (5), 1079–1114.
- Blasco, Andrea, Kevin J. Boudreau, Karim R. Lakhani, Michael Menietti, and Christoph Riedl**, “Do Crowds have the Wisdom to Self-Organize?,” 2013. Working Paper.
- Chen, Roy**, “Coordination with Endogenous Groups,” *Journal of Economic Behavior and Organization*, 2017, *141* (1), 177–187.
- Chen, Yan and Sherry Xin Li**, “Group Identity and Social Preferences,” *American Economic Review*, March 2009, *99* (1), 431–457.
- Eckel, Catherine C. and Philip J. Grossman**, “Managing Diversity by Creating Team Identity,” *Journal of Economic Behavior & Organization*, November 2005, *58* (3), 371–392.
- Goette, Lorenz, David Huffman, and Stephan Meier**, “The Impact of Group Membership on Cooperation and Norm Enforcement: Evidence Using Random Assignment to Real Social Groups,” *American Economic Review*, May 2006, *96* (2), 212–216.
- , – , and – , “The Impact of Social Ties on Group Interactions: Evidence from Minimal Groups and Randomly Assigned Real Groups,” *American Economic Journal: Microeconomics*, 2012, *4* (1), 101–115.
- Hamilton, Barton H., Jack A. Nickerson, and Hideo Owan**, “Team Incentives and Worker Heterogeneity: An Empirical Analysis of the Impact of Teams on Productivity and Participation,” *Journal of Political Economy*, June 2003, *111* (3), 465–497.
- Harrison, Glenn W and John A List**, “Field experiments,” *Journal of Economic literature*, 2004, *42* (4), 1009–1055.
- Hoogendoorn, Sander, Simon C. Parker, and Mirjam van Praag**, “Ability Dispersion and Team Performance,” 2014. Tinbergen Institute Discussion Paper 14-053/VII.

- Ichniowski, Casey and Kathryn Shaw**, “The Effects of Human Resource Management Systems on Economic Performance: An International Comparison of U.S. and Japanese Plants,” *Management Science*, May 1999, *45* (5), 704–721.
- **and** – , “Beyond Incentive Pay: Insiders’ Estimates of the Value of Complementary Human Resource Management Practices,” *Journal of Economic Perspectives*, Winter 2003, *17* (1), 155–180.
- , – , **and Giovanna Prennushi**, “The Effects of Human Resource Management Practices on Productivity: A Study of Steel Finishing Lines,” *American Economic Review*, June 1997, *87* (3), 291–313.
- Lazear, Edward P. and Kathryn L. Shaw**, “Personnel Economics: The Economist’s View of Human Resources,” *The Journal of Economic Perspectives*, Fall 2007, *21* (4), 91–114.
- Mookherjee, Dilip**, “Decentralization, Hierarchies, and Incentives: A Mechanism Design Perspective,” *Journal of Economic Literature*, June 2006, *44* (2), 367–390.
- Munkres, James**, “Algorithms for the Assignment and Transportation Problems,” *Journal of the Society for Industrial and Applied Mathematics*, March 1957, *5* (1), 32–38.
- Pan, Jessica, Dominic Soon, and Wei-Kang Wong**, “Are Employers Susceptible to Coarse Thinking? Implications for Wage Determination and Returns to Signaling,” 2015. Working paper.
- Pan, Xiaofei Sophia and Daniel Houser**, “Cooperation during cultural group formation promotes trust towards members of out-groups,” *Proceedings of the Royal Society B*, 2013, *280* (1762).