TEAM SIZE AND PERFORMANCE:
EXPERIMENTAL EVIDENCE FROM INDIA[*]

Giorgia Barboni[†]
University of Warwick

Ayah Bohsali[‡]
UPF

Elisa Giannone[§]
CREI and CEPR

April 2025

Abstract

What are the gains of team size for performance? Does new technology diffuse faster in larger groups? To answer these questions and solve the endogenous sorting bias, we run the first randomized controlled trial where 1,582 Indian women were randomly assigned to 206 groups of different size and encouraged to engage in a production process. We find smaller teams to perform better than larger teams and that technology diffuses non-linearly with team size. By measuring bilateral communication within teams and conducting network analysis across team members, we find that smaller teams are more centralized than medium and larger. Our findings can be explained by smaller teams organizing the production process more efficiently, for example, by assigning tasks to group members and being more likely to reach a consensus on production decisions. To estimate the welfare effects of the intervention, we propose to build a network model with a heterogeneous group size. Size affects coordination costs as well as the impact of new technologies.

1 Introduction

Teamwork is a fundamental feature of the modern economy (Deloitte, 2016). Understanding how team production works is first-order for effective technology diffusion and overall economic growth and development. Despite the benefits associated to a larger size, 90% of firms worldwide remain small, especially in developing countries (OECD, 2000). The theoretical literature has advanced and built numerous theories in the attempt to rationalize the costs of teams. First, coordination and monitoring are harder to achieve when more people interact (e.g. Becker and Murphy 1992, Garicano and Rossi-Hansberg 2006). Second, organizational constraints, which are more frequent in larger firms (Williamson, 1967; Hsieh and Klenow, 2009), may offset the gains from of human capital spillovers and specialization. We put forward a complementary hypothesis: The larger the team, the less likely individuals are to connect with everyone else, making the network more dispersed and less fluid. In a developing context where individual emancipation is more limited, this channel can be especially sizable and undermine specialization and effective technology diffusion.

Testing our hypothesis requires two empirical objects. First, we need to solve one of the fundamental identification problem of studying teams, which is the selection: team members have likely sorted themselves based on their skills and other characteristics that are unobservable to the econometrician. Selection also affects the extent to which team members interact and learn from each other during and outside the production process, which in turn poses challenges to correctly identifying the determinants of their performance – and that of the team as a whole. Second, measuring communication frictions within a team is problematic as it requires knowing the bilateral interaction of non self-selected team members.

We overcome the above-mentioned identification challenges by running the first randomized controlled trial (RCT) that creates and tracks groups – "production teams" henceforth – of exogenously assigned size and the receiptf a new technology to be diffused. The RCT helps us isolate the effects of team members' interaction from selection into networks of co-workers. This allows us to: *i)* identify the impact of the production team size on productivity and other production-related outcomes; *ii)* study how a new technology interacts with production teams' size to increase productivity, *iii)* study how size shapes the outcomes of individuals that work together and their interactions; *iv)* estimate a structural model using the micro-elasticities identified in the RCT to understand the impact of internal network on teams' performance and growth.

Contrary to what theories of human capital spillovers predict – i.e., higher productivity to be found in larger groups – our core result is that larger teams are *less* productive, hinting at a negative

role of team size on performance. Thanks to the richness of our data, we explore the mechanisms underlying these results. Firstly, we study how production teams of different sizes allocate tasks and divide management roles. Secondly, we look at how a new technology interacts with the team size. Consistent with a theoretical framework in which coordination and communication costs are very high, we find that in large teams, tasks are assigned in a less efficient way, with co-workers less likely to specialize in tasks. On the contrary, in smaller teams, tasks are more likely to be assigned following individual preferences and consensus. When a new technology is introduced, its effects on productivity are less pronounced in larger production teams. Once more, this result speaks to the role of coordination costs in preventing larger teams from reaping the benefits of technology advancement.

We set up our study in the Indian State of Chhattisgarh, in collaboration with the Chhattisgarh State Rural Livelihood Mission (CGSRLM). CGSRLM is part of the Indian National Rural Livelihood Mission, a poverty alleviation project implemented by the Ministry of Rural Development, Government of India. The program promotes the creation of economic collectives among poor women. These collectives have similar characteristics to saving and borrowing groups: women who are part of the collectives typically meet once a week or twice a month to pool money together (Dupas and Robinson, 2013). The size of these collectives varies considerably across villages, although, on average, there are approximately 12-15 members in each group. The women's collective context in India is particularly suitable for testing theories of social interactions and human capital spillovers in a production context. In fact, through these collectives, women typically engage in small-scale, joint production activities, including dairy farming and handicraft (Dhake and Narkhede, 2019), hence these collectives can be considered as small "teams", and their members as "coworkers".

Related Literature This paper relates to several strands of the literature. First, the literature on organizational economics which has emphasized the relationship between specialization and performance (Chandler et al., 2009, Becker and Murphy, 1992, Garicano and Rossi-Hansberg, 2006, Bassi et al., 2023, Freund, 2022). To the best of our knowledge, we present the first causal evidence of the impact of team size on performance. And, thanks to the richness of individual, group, and cross-individual information, we can empirically disentangle different mechanisms and highlight how difficulty coordinating in larger teams might be a strong internal organization barrier. At the same time, in small teams, it is easier for the leader to stand out, and less monitoring is needed. The closest study to ours is Bassi et al., 2023 that, through novel survey data and a brand-new structural model, find that Ugandan firms are size-constrained by being unable to specialize tasks since the demand is very customized. By encouraging our study population to produce a homogenous

good, our study "shuts down" the demand customization channel, thus complementing their study by delving into internal communication costs and their interaction with new technologies.

Our study also contributes to the literature on team performance, starting with the seminal work of Holmstrom (1982). More recent papers empirically evaluate team formation and peer effects on productivity or identify how single team members affect team output (Agha et al. 2018, Bonhomme 2021, Bandiera et al. 2010, Devereux 2018 and Mas and Moretti 2009). To the best of our knowledge, this is the first study that provides experimental and causal evidence on team size and performance.

Finally, our paper adds to a growing literature that uses field experiments to uncover and explain key characteristics of firms such as Cai and Szeidl (2018), Brooks et al. (2018) and Brooks et al. (2020). In contrast to this set of papers that focus on connecting firms with other firms, we study the internal dynamics of teams rather than the external ones.

The rest of the paper is divided as follows. In Section 2, we describe the experimental design and sample. Section 3 describes the data, the empirical strategy and the main outcomes. Section 4 unbundles the core mechanism and the alternative ones. In section 5, we report the effects of the differential treatments on women that belonged to active groups. Section 6 concludes. Additional results are in the Online Appendix.

2   Setting and Experimental Design

We leveraged the partnership with CGSRLM to create 206 new women's collectives ("production teams" or "teams" henceforth) in rural Chhattisgarh, involving a total of 1,582 women across 120 villages. At the creation stage, all teams received business training focusing on soap production techniques to promote female entrepreneurship under the creation of small-scale joint businesses. The reason for choosing soap over other possible products is the relative simplicity of the techniques and the low cost of raw material to produce soap bars. In addition, to maximize soap bars production, women should work in groups with a clear division of labor, mirroring the functioning of more complex production systems. This holds the potential to help women set the foundations for profitable business activities. As shown in Figure 1, the experiment was initiated at the end of November 2021 and was completed at the end of April 2022.

The RCT leverages two interventions that promote interactions and knowledge sharing among co-workers. The first intervention consists in exogenously varying the size of the production teams –

either 5 ("Small-size teams" treatment), or 8 ("Medium-size teams" treatment), or 12 ("Large-size teams" treatment) women per team. This is used to understand how the size of teams of coworkers affects learning and productivity. The second intervention consists in randomly "seeding" additional technology-related information ("Info Seed" treatment) in the teams to test how coworkers learn from each other, change the internal organization of labor to adopt it and whether their performance improves. This is used to study how a new technology might affect the performance of groups of varying size and their internal organization.

Sampling The RCT target population consists of women living in rural villages in the Mahasamund district of the Indian State of Chhattisgarh. The main inclusion criteria for women to be part of this study and to receive the business training are: i) they should be members of a Self-Help Group (SHG) affiliated to CGSRLM at the time of the recruitment; ii) they should be between 18 and 65 years of age; iii) they should own or have access to a smartphone (this is because the business training was administered via smartphones).

Randomization Randomization first occurred at the village level. Study villages were randomly assigned to three treatment arms, corresponding to creating production teams of 5, 8 and 12 women, respectively. A total of 218 groups were created, of which 47 consist of 12 members, 63 of 8 members and 96 of 5 members, respectively.[1] This yielded a sample size of 1,582 women in 120 villages. We then selected women to be part of the group(s) in a given village by randomly drawing from the lists of female SHG members in that village. Figure 2 shows the design of the experiment and how many women and villages were assigned to each treatment and cross-treatment arm.

Interventions After being administered a baseline survey individually, women were invited to attend a business training event to be held on the same day with other women from the same village. The training was held in central locations in the village and lasted approximately two hours. The main objectives of the training were to: *i)* introduce women to soap bars' production and *ii)* encourage them to start working with the other women who also attended the training event. The training was designed in partnership with CGSRLM. The training followed a "blended" approach: at the event, one of our enumerators explained to participants the aims of the business training, including the importance of acquiring business skills to increase their agency. A video was shared with women via their smartphones, wherever available, providing a detailed guide on how to produce soap bars. Importantly, the video stressed the group dimension of the production process, and encouraged trainees to engage in a joint soap-business activity. Every woman who attended the

---

[1]A village could be randomized to have at most two groups

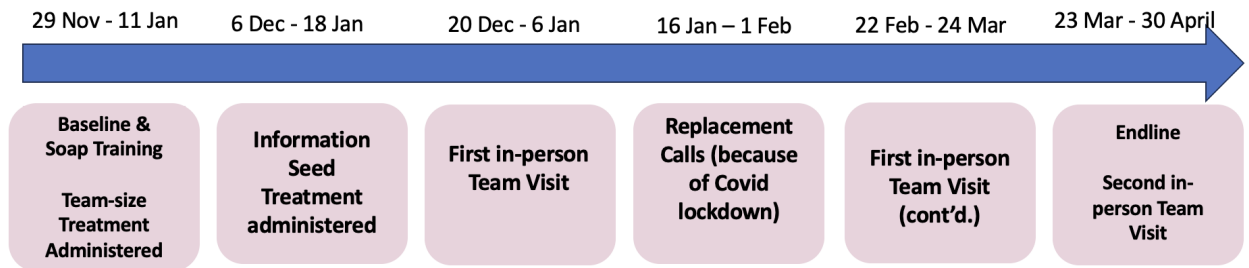| 29 Nov - 11 Jan	6 Dec - 18 Jan	20 Dec - 6 Jan	16 Jan – 1 Feb	22 Feb - 24 Mar	23 Mar - 30 April
Baseline & Soap Training Team-size Treatment Administered	Information Seed Treatment administered	First in-person Team Visit	Replacement Calls (because of Covid lockdown)	First in-person Team Visit (cont'd.)	Endline Second in-person Team Visit

Figure 1: RCT and Data Collection Timeline

training event was offered the same business skill training, irrespective of the randomly-assigned size of the production group they ended up being assigned to.

In addition to the initial set of information women received as part of the business training, half of the women in our study sample were also randomly exposed to additional information about a new technology that were delivered to them via phone calls. Specifically, we provided information on how to use sari cloth to wrap the soap bars instead of plastic foil. Overall, the aim of this second intervention was to enhance the quality of soap bars –especially in terms of innovation and environmental sustainability of the production process.

3   Data and Reduced Form Empirical Strategy

We collected two sets of data: at the team level, and at the individual level.

Production Team Data We envisioned the data collection for team-level outcomes to consist of in-person visits and weekly diaries to be completed by the teams, with the first in-person visit to take place approximately one month, and the second one four months after the training, respectively. Unfortunately, seven weeks after the start of the RCT, all field activities had to be halted because of the Omicron wave of Covid in India, forcing us to switch from in-person to phone-based data collection. Table A.1 provides an overview of how the production teams were tracked across time: after the start of the intervention, we managed to visit 49 teams in person to elicit production-related outcomes before Covid hit. For the remainder of the teams that we could not meet, the in-person visits were replaced by calls. We managed to reach by phone 153 more teams.[2] The Covid-induced lockdown lasted approximately 6 weeks; when restrictions were lifted, at the end of February 2022,

---

[2]To cross-validate outcomes, two team members were called from each team to answer the same survey.
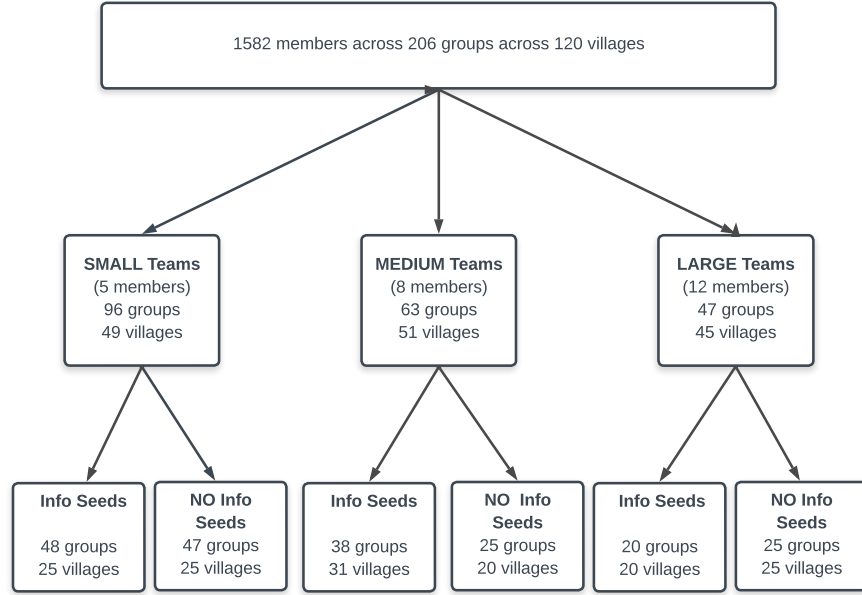
Figure 2: Experimental Design: This figure shows the experimental design and the type of groups, number of subjects involved, and number of members and villages in each treatment arm.

we decided to meet in person all the groups we had interviewed by phone – we successfully met 144 out of 153 groups. Four months after the start of the intervention, we met all the teams again in person and carried out a team-level survey with at least 50% of team members being present. In all the visits, we asked teams whether they met and produced in the previous month. Column 2 of Table A.1 shows how many groups reported having produced within that time frame. Over the entire time period, 116 groups reported having produced soap within the whole duration, and 165 reported meeting, even if this does not necessarily mean that they produced soap.

Individual-level Data Two rounds of socio-economic surveys were carried out throughout the

7

study: one at the start of the business training (baseline); the other one after four months (endline).[3] These surveys assessed women's socioeconomic profile, business skills, entrepreneurship, as well as decision-making power and wellbeing. Since the business training entailed a strong digital component, the surveys were also used to capture women's use of smartphones and digital technologies. We measured women's social connectedness by eliciting the extent of their social interactions. The endline survey contained questions to identify social interactions among training participants, both in person and through smartphones: how many times they met in person since the training was administered; whether and how many times they interacted via phone; what type of information they exchanged during in-person and digital interactions (e.g., purely business-related information, or also personal information). This allows us to track interactions within teams over time.

3.1 Summary Statistics and Balance Checks

3.1.1 Team Characteristics

Table A.2 presents summary statistics and balancing checks for the production teams included in the study. At the onset of the study, groups consisted of women of 33 years on average; more than 90% of group members were married; women had on average two children, all these characteristics being balanced across treatment arms. Overall, women in our sample appear fairly educated, with more than 60% of them having completed at least secondary school. In terms of economic activities, half of the women in our sample ran a business at the beginning of the study – this is not surprising as all the women involved in our study are part of women's collectives, through which they carry out small entrepreneurial activities –, while around 40% of them worked for a wage (either in agri or non-agri sector). Again these characteristics are all balanced at baseline. We instead find a slight imbalance in terms of per-capita household income and average business sales in the past 30 days. To account for these imbalances, we include these variables as controls in our main analysis.

3.2 Empirical Analysis

We begin our analysis by estimating the following regression at the group level:

$$y_{pv} = \beta_0 + \beta_1 T_{1v} + \beta_2 T_{2v} + X_{pv} + \epsilon_{pv} \tag{1}$$

where y_{pv} is our team-level outcome of interest (e.g., number of soap bars produced; per capita productivity) for production team p in village v. T_1 and T_2 are dummies indicating whether the

---

[3]The team-level survey was carried out at the end of the individual-level endline.

group is of small or medium size – the omitted dummy is for large-size teams.[4] We also include a vector of team-level characteristics X_{pv}. ϵ_{pv} is the error term. Standard errors are clustered at the village level.

Table 1 presents our main set of results in terms of duration of meetings and the production process, as well as the number of soap bars produced in total and per capita (PC) productivity. Panel A shows the main results for the full set of teams, irrespective of whether they met at least once to start production. Panel B instead presents results restricting the analysis to the groups who met at least once (but not necessarily ended up producing). Two main results stand out from this table: first, in terms of overall number of soap pieces produced, we find a positive relationship between group size and production (column 3). This is not surprising, as it simply reflects the larger dimension of production in larger groups. On the contrary, when we look at per-capita productivity in Column 4, we observe higher per-capita productivity in smaller teams than in larger teams: in the latter, women produce on average 1 soap bar per hour, whereas in small and medium-size teams they produce 62% and 33% more – the difference between small teams and large teams being also statistically significant. Results are similar, both in magnitude and significance, when we restrict our analysis to teams that met at least once (Panel B). In appendix C, we also include the results restricting the sample to teams that reported a positive production amount.

Overall, results from Table 1 suggest that the net effect between net gains from size are negative in this context. This could be completely mechanical if the production did not entail any advantage of large size. Yet, large teams could have specialized more or ripped the benefits of the information spread to one person being shared with more and going faster in the production. As a matter of fact, the most successful prodcution team was indeed a large group, named Muskan. We present their case study in Appendix section B.

In what follows, we investigate the mechanisms behind our main finding and we put forward our new hypothesis about size-biased network formation. On top of that, we analyze more broad differences in terms of the organization of the production process and monitoring systems the teams put in place.

---

[4]Note that we do not have a *pure* control group where no training was administered. Given our main research questions – the impact of team size on productivity –all teams received the business training and we look at treatment effects across different group sizes.

Table 1: Treatment Effects on Production Process

	Meeting Duration	Production Duration	Quantity Produced	PC Productivity (per hour)
	(1)	(2)	(3)	(4)
Panel A: Full Sample				
Small	-67.274**	-26.083*	-11.836**	0.512*
	(28.829)	(13.589)	(4.951)	(0.288)
Medium	-47.386	-17.268	-7.199	0.263
	(29.761)	(14.272)	(5.372)	(0.276)
Small = Medium	0.214	0.211	0.045	0.444
Mean (Large)	130.372	52.979	17.691	.963
N	206	206	206	206
Panel B: Conditional on Meeting				
Small	-83.956**	-32.717*	-14.974**	0.660**
	(34.711)	(16.692)	(6.101)	(0.328)
Medium	-56.338	-20.554	-8.920	0.395
	(34.724)	(16.993)	(6.483)	(0.314)
Small = Medium	0.120	0.120	0.021	0.461
Mean (Large)	165.608	67.297	22.473	1.224
N	167	167	167	167

Note: ***, **, * indicates significance at the 1%, 5%, and 10% level respectively. In all columns, controls are selected through LASSO. Robust standard errors reported in parentheses below the coefficients are clustered at the village level. Outcome variable definitions are available in C.6. Meeting duration, production duration, and quantity produced are calculated by summing over all group meetings for each group. PC Productivity per hour is calculated by taking the ratios of total quantity produced to meeting duration (in hours), and dividing it by the original group size. Sample size in Panel B corresponds to the number of groups that met, but did not necessarily produce soap.

4 Disentangling the Mechanisms

Our main empirical results show that larger teams have lower per-capita productivity than medium and small teams. We hypothesise that these differences could be driven by more efficient production processes in smaller vs. larger groups. Hsieh and Klenow (2009) show that, in the context of India and China, large firms face more organizational constraints than small ones – implying that smaller groups achieve a better workers' specialization than larger ones. Following this line of reasoning, it could be that the coordination and communication costs further reduce the benefits of large-scale production. Also, it might be the case that with additional organization layers in larger teams, monitoring the production process becomes more difficult. In light of this, one may wonder how the introduction of a new technology that could potentially improve the production process would affect groups of different size.

To test which of these hypotheses dominates, we exploit the rich set of data we collected through our surveys, which contain specific targeted questions on team organization and labor division.

4.1 Testing for Specialization: Task Divisions

We study how teams organize themselves for the production process by asking them how they decide who-does-what: whether they follow members' ability, their preferences; whether they reach a consensus on how tasks should be performed. We also ask questions about how team members do actually perform tasks: whether they do any task, irrespective of their ability and skills, and whether they work in subgroups. Results for differences across group size in terms of task allocation are shown in Table 2. All the dependent variables are dummies. We do not observe any statistical difference across team sizes when we look at whether tasks were assigned by ability (Column 1). However, in Column 2, results show that, while no large teams followed individual preferences in assigning tasks to individual workers, small and medium-size teams did so – the difference between small and large groups being also statistically significant. Similarly, Column 3 shows that smaller groups are 8 percentage points significantly more likely to assign tasks following the group consensus. Column 4 shows that the small groups are, respectively, 31 p.p. less likely to assign tasks such that "everyone does all tasks". This result is striking as it hints at the difficulty in organizing tasks and letting members specialize, particularly in larger groups. Importantly, this finding sharply contrasts with evidence that, in larger groups, workers are able to specialize to a higher extent. All in all, results from Table 2 indicate that larger groups face coordination and communication problems. This, in turn, may explain why their productivity and overall performance are lower than in smaller

groups.

Table 2: Treatment Effects on Production Organization

	Ability	Individual Preferences	Consensus	Everyone Does all Tasks	Work in Subgroups
	(1)	(2)	(3)	(4)	(5)
Panel A: Full Sample					
Small	0.055	0.076**	0.076**	-0.330***	-0.060
	(0.038)	(0.035)	(0.030)	(0.092)	(0.050)
Medium	0.033	0.010	-0.005	-0.039	0.071
	(0.042)	(0.021)	(0.026)	(0.103)	(0.065)
Small = Medium	0.606	0.133	0.048	0.001	0.038
Mean (Large)	.021	0	0	.66	.067
N	206	206	206	206	192
Panel B: Conditional on Meeting					
Small	0.067	0.085**	0.092**	-0.433***	-0.093
	(0.052)	(0.042)	(0.039)	(0.100)	(0.062)
Medium	0.044	-0.004	-0.007	-0.075	0.067
	(0.057)	(0.030)	(0.031)	(0.100)	(0.076)
Small = Medium	0.665	0.115	0.049	0.000	0.035
Mean (Large)	.027	0	0	.838	.081
N	167	167	167	167	163

Note: ***, **, * indicates significance at the 1%, 5%, and 10% level respectively. In all columns, controls are selected through LASSO. Robust standard errors reported in parentheses below the coefficients are clustered at the village level. Variables in columns (1)-(4) take on the value 1 if a group divides tasks according to ability, individual preferences, consensus, or by assigning to all tasks to everyone, respectively, and 0 otherwise. Column (5)'s "Work in Subgroups" takes the value 1 if the group divided in order to produce. Outcome variable definitions are available in C.6. Sample size in Panel B corresponds to the number of groups that met, but did not necessarily produce soap.

4.2 Organizational Layers: Monitoring and Leadership

An alternative and potentially complementary hypothesis that may explain why larger production teams perform worse than smaller ones is that, given their size, in larger teams it is hard to set up organizational layers. We look at the extent of monitoring in group production in Table 3. We find weak support for this hypothesis. Panel A of column (1) shows that small groups are less likely than larger groups to use monitoring. Column (2) adds to findings in column (1) by showing that numbers of monitors per capita are higher in medium groups than in large groups. This leads us to rule out the hypothesis that smaller groups perform better because their members are easier to monitor.

Table 3: Treatment Effects on Monitoring Production

	Use Monitoring	Monitors by Group Size	Has Group Leader
	(1)	(2)	(3)
Panel A: Full Sample			
Small	-0.040	0.024	0.015
	(0.080)	(0.017)	(0.049)
Medium	0.091	0.036*	-0.001
	(0.082)	(0.020)	(0.057)
Small = Medium	0.088	0.643	0.723
Mean (Large)	.191	.019	.9330000000000001
N	206	201	193
Panel B: Conditional on Meeting			
Small	-0.012	0.026	0.026
	(0.095)	(0.021)	(0.047)
Medium	0.123	0.043*	-0.011
	(0.094)	(0.023)	(0.059)
Small = Medium	0.128	0.563	0.342
Mean (Large)	.189	.023	.9460000000000001
N	167	166	164

Note: ***, **, * indicates significance at the 1%, 5%, and 10% level respectively. In all columns, controls are selected through LASSO. Robust standard errors reported in parentheses below the coefficients are clustered at the village level. Outcome variable definitions are available in C.6. Sample size in Panel B corresponds to the number of groups that met, but did not necessarily produce soap.

4.3 How a New Technology Interacts with Team Size

The evidence we have shown so far is consistent with the hypothesis that coordination and communication costs prevented further labor specialization and, to some extent, functional organizational layers, in larger groups. According to classical theories like Becker and Murphy (1992), access to new technology helps to increase the gains from specialization and, thus, the costs of coordination and communication. We create an entire randomization leg to test this hypothesis and further validate this mechanism. As described above, via phone, we explained a new technology for wrapping soap bars to two team members. To test our hypothesis, we run the following specification:

$$y_{pv} = \beta_0 + \beta_1 T_{1v} + \beta_2 T_{2v} + \beta_3 InfoSeeds_{pv} +$$
$$\beta_4 T_{1v} \times InfoSeeds_{pv} + \beta_5 T_{2v} \times InfoSeeds_{pv} + \beta_6 T_{3v} \times InfoSeeds_{pv} + \epsilon_{pv} \quad (2)$$

where $InfoSeeds_{pv}$ denotes the random assignment to receive additional information related to cloth wrapping, and $T_{1v} \times InfoSeeds_{pv}$, $T_{2v} \times InfoSeeds_{pv}$ and $T_{3v} \times InfoSeeds_{pv}$ represent the interaction between receiving the information seed and different group sizes, respectively. Table 4 reports the results of specification 2. Column (4) shows that among teams that did not receive the additional information, smaller teams produced more soap bars per capita and time. Among those that received the information, medium-size teams did significantly better than medium-size teams that did not receive it. We do not finds similar results for large groups.

Table 4: Treatment Effects on Production Process- Technology Diffusion

	Meeting Duration	Production Duration	Quantity Produced	PC Productivity (per hour)
	(1)	(2)	(3)	(4)
Small	-66.330	-27.712	-12.440	0.754*
	(43.754)	(21.501)	(7.803)	(0.382)
Medium	-99.840**	-40.681*	-15.269*	-0.201
	(45.447)	(22.950)	(8.580)	(0.310)
Small × Info Seeds	-15.277	-9.420	-2.730	-0.165
	(17.471)	(7.663)	(2.077)	(0.451)
Medium × Info Seeds	76.391***	27.900***	10.144***	1.093***
	(23.161)	(10.428)	(3.296)	(0.338)
Large × Info Seeds	-35.386	-21.214	-6.421	0.192
	(51.106)	(23.936)	(9.034)	(0.359)
Small = Medium	0.098	0.170	0.327	0.018
Small × Info Seeds = Medium × Info Seeds	0.002	0.006	0.002	0.023
Small × Info Seeds = Large × Info Seeds	0.711	0.635	0.683	0.531
Medium × Info Seeds = Large × Info Seeds	0.050	0.068	0.094	0.065
N	205	205	205	205

Note: ***, **, * indicates significance at the 1%, 5%, and 10% level respectively. In all columns, controls are selected through LASSO. Robust standard errors reported in parentheses below the coefficients are clustered at the village level. Outcome variable definitions are available in C.6. Sample size $N = 205$ corresponds to the full sample.

We can test our mechanism even more directly to observe whether among groups that received the new technology, medium groups tend to organize the task more efficiently. The results are striking as shown in table 5. When the info seed is given, the relationships with the division of labor stay the same for the groups that did not receive it. But it reverts for small and medium groups that did receive, and it improves for the large groups that did receive it. This confirms our current hypothesis that when technology improves, coordination costs get relatively smaller than the gains of specialization, expanding the size of the groups as a result.

Table 5: New Technology and Task Division among Different Size Teams

	Ability	Individual Preferences	Consensus	Everyone Does all Tasks	Work in Subgroups
	(1)	(2)	(3)	(4)	(5)
Small	0.059	0.071	0.166*	-0.319***	-0.076
	(0.059)	(0.066)	(0.087)	(0.108)	(0.079)
Medium	-0.025	-0.042	0.026	-0.141	0.222
	(0.053)	(0.032)	(0.045)	(0.143)	(0.137)
Small × Info Seeds	0.036	-0.095	-0.016	-0.152	0.098*
	(0.094)	(0.063)	(0.087)	(0.136)	(0.057)
Medium × Info Seeds	0.052	0.045	-0.000	0.095	-0.137
	(0.079)	(0.051)	(0.038)	(0.157)	(0.138)
Large × Info Seeds	0.087	-0.010	-0.028	-0.117	0.153
	(0.083)	(0.026)	(0.037)	(0.110)	(0.129)
Small = Medium	0.241	0.089	0.104	0.288	0.039
Small × Info Seeds = Medium × Info Seeds	0.896	0.112	0.870	0.254	0.142
Small × Info Seeds = Large × Info Seeds	0.687	0.244	0.898	0.842	0.694
Medium × Info Seeds = Large × Info Seeds	0.778	0.357	0.584	0.264	0.137
N	116	116	116	116	115

Note: ***, **, * indicates significance at the 1%, 5%, and 10% level respectively. In all columns, controls are selected through LASSO. Robust standard errors reported in parentheses below the coefficients are clustered at the village level. Outcome variable definitions are available in C.6. Sample size $N = 116$ corresponds to groups that recorded producing soap.

4.4 Evidence on Team Interactions

To gain a better understanding of how team interactions change with team size, we utilize our micro data on interactions within-group. In the survey, we ask team members to list with which other members they *i)* communicate over the phone, *ii)* communicate in person, *iii)* discuss production, and *iv)* ask for support during production. Using this data, we can construct a snapshot of the network of interactions of a group. Although not all group members responded to the questionnaire, we find no differential attrition in responses by group size (see Table C.17 in the appendix). Tables 6 and 7 display the number of group members that a member lists to be in contact with for each category, with the former displaying the raw number of connections, and the latter the number of connections divided by the team size. We can see that even though the number of connections mechanically increases in group size, once we study the normalized connections in Table 7, we obtain that the network density is *decreasing* in group size, meaning that on average individuals in smaller groups interact more with each other, consistent with the idea that communication and coordination become more difficult as team size increases. By studying the constructed network of team members,

we can also infer about team interactions using the shape of the network. Sparrowe et al. (2001) posit evidence that denser networks in the context of support are positively related to higher levels of group performance. They also predict that in groups with more centralized networks, performance in complex tasks is likely to be hindered compared to decentralized communication networks, as is also predicted in the seminal works of Shaw (1964).

Table 6: Number of (Raw) Team Connections

	Talk on Phone	Meet in Person	Talk Production	Support in Production
	(1)	(2)	(3)	(4)
Small	-0.192**	-1.030***	-0.856**	-0.253
	(0.096)	(0.167)	(0.396)	(0.157)
Medium	0.156	-0.380*	-0.420	0.043
	(0.110)	(0.199)	(0.419)	(0.190)
Small = Medium	0.002	0.000	0.081	0.061
Mean (Large)	1.179	2.85	1.44	.589
N	1044	1063	821	812

Note: ***, **, * indicates significance at the 1%, 5%, and 10% level respectively. In all columns, controls are selected through LASSO. Robust standard errors reported in parentheses below the coefficients are clustered at the village level. Each column represents the number of group members each group member has interacted with. (talked to on the phone, in person, talked during production, or has helped during production) Outcome variable definitions are available in C.6. Sample size differs by individuals that received the question in the survey.

4.5 Alternative Hypothesis: External Barriers

So far, our results speak to the idea that organizational constraints may offset the gains derived from working in larger teams. The negative relationship between group size and productivity appears to be driven by the ability of smaller groups to organize their work better, leading to workers' specialization and, hence, to a more efficient production process. Yet, one alternative explanation for our results could be that our treatments were differentially exposed to marketing strategies, or experienced input procurement or soap quality issues differently. While we tried to neutralize some of these

Table 7: Number of (Normalized) Team Connections

	Talk on Phone	Meet in Person	Talk Production	Support in Production
	(1)	(2)	(3)	(4)
Small	0.101***	0.126***	-0.005	0.020
	(0.015)	(0.017)	(0.038)	(0.019)
Medium	0.070***	0.072***	0.004	0.029
	(0.013)	(0.020)	(0.040)	(0.020)
Small = Medium	0.084	0.007	0.790	0.672
Mean (Large)	.098	.237	.12	.049
N	1044	1063	821	812

Note: ***, **, * indicates significance at the 1%, 5%, and 10% level respectively. In all columns, controls are selected through LASSO. Robust standard errors reported in parentheses below the coefficients are clustered at the village level. Each column represents the number of group members each group member has interacted with (talked to on the phone, in person, talked during production, or has helped during production), divided by the group size. Outcome variable definitions are available in C.6. Sample size differs by individuals that received the question in the survey.

issues "by construction" in the experiment (i.e., we provided inputs proportionally to group size), we can't, in principle, fully rule out that group size could correlate with any of the above-mentioned issues. In fact, groups could have potentially have bought more inputs by themselves from other providers (we collect these details as well).

We study the effects of groups size on external barriers more rigorously through a regression analysis whose results are displayed in Table 8. Overall, we don't see any differential treatment effects on any of these outcomes. This rules out that our results may be driven by how external barriers interact with group size.

Table 8: Treatment Effects on External Barriers

	Input Procurement Issues	Marketing Issues	Soap Quality Issues
	(1)	(2)	(3)
Small	-0.093	-0.084	-0.140
	(0.124)	(0.122)	(0.102)
Medium	-0.239**	-0.037	-0.051
	(0.120)	(0.121)	(0.109)
Small = Medium	0.191	0.653	0.326
Mean (Large)	.438	.406	.25
N	144	146	146

Note: ***, **, * indicates significance at the 1%, 5%, and 10% level respectively. In all columns, controls are selected through LASSO. Robust standard errors reported in parentheses below the coefficients are clustered at the village level. A smaller sample size is being used due to the question being introduced in the final stages of the survey.

5   Wellbeing and Empowerment Impact on Team Members

In this section, we study whether team size had differential impacts on team members' wellbeing. The hypothesis is that if teams were more productive and organized their production process more efficiently, there might be also positive spillovers to individual-level outcomes like financial empowerment or mental health. To assess the extent of spillover effects from the group to individual members, we estimate the following regression equation at the individual level:

$$y_{ipv} = \beta_0 + \beta_1 T_{1ipv} + \beta_2 T_{2ipv} + X_{ipv} + \epsilon_{ipv} \tag{3}$$

where y_{ipv} represents individual-level outcomes, T_{1ipv} and T_{2ipv} are dummies for whether woman i in production team p in village v was assigned to either a large or medium group. We also include

a vector of baseline individual-level characteristics X_{ipv}. While 1452 women were involved in the experiment, we restrict the individual-level analysis only to women who were part of active teams (1092 women) to make sure we capture the effect on team production.

Table 9 shows that there is no differential impact on autonomy indicators (grouping autonomy in financial, health-related, education-related, and consumption decision-making) across members belonging to different groups. One potential reason for this lack of result is the relatively short time frame of the intervention. On the contrary, we observe that members of larger teams have less phone access than members in small and medium teams at endline.

Table 9: Treatment Effects on Empowerment

	Autonomy (KLK)	Phone Access
	(1)	(2)
Small	0.022	0.026**
	(0.057)	(0.011)
Medium	0.083	0.033***
	(0.053)	(0.011)
P-values		
Small = Medium	0.296	0.413
Mean (Large)	.076	.964
N	1092	1092

Note: ***, **, * indicates significance at the 1%, 5%, and 10% level respectively. Sample includes all women whose groups met up, but did not necessarily produce.

We then turn to study treatment effects on physical and mental health. Results are reported in table 10. Overall, we do not find any significant treatment effects on mental or physical health. Again, one potential explanation for the lack of findings is the relatively short time span of the intervention – improvements in outcomes like mental health or expectations about the future may take longer to

realise.

Table 10: Treatment Effects on Mental Health, Expectations and Physical Health

	Kessler Scale	Worry About Money	Future Expectations	Illness
	(1)	(2)	(3)	(4)
Small	0.283	-0.008	0.018	0.043
	(0.305)	(0.030)	(0.019)	(0.033)
Medium	0.199	-0.019	-0.013	0.040
	(0.297)	(0.032)	(0.017)	(0.031)
P-values				
Small = Medium	0.768	0.732	0.170	0.923
Mean (Large)	3.288	.48	.053	.193
N	1092	1087	1092	1087

Note: ***, **, * indicates significance at the 1%, 5%, and 10% level respectively.

Finally, we study the impact of belonging to different group sizes on borrowing and saving as reported in table 11. We do not observe the differential impact on saving. However, we find at the endline that women belonging to large groups have borrowed 6,000 and 5,000 Indian Rupees less than women in small and medium groups, respectively. This finding is consistent with the hypothesis that members belonging to more productive groups wanted to finance the business and make it more sustainable.

Table 11: Treatment Effects on Saving and Borrowing

	Savings			Borrowing	
	Owns Savings Account (1)	Account Usage (Frequency) (2)	Savings Amount (3)	Borrowed this Year (4)	Amount Borrowed (5)
Small	0.001	-0.101	185.143	0.037	6000.368**
	(0.012)	(0.099)	(255.522)	(0.034)	(2527.242)
Medium	-0.008	-0.025	-165.286	0.038	5555.297**
	(0.015)	(0.107)	(152.752)	(0.037)	(2557.697)
P-values					
Small = Medium	0.524	0.390	0.188	0.973	0.878
Mean (Large)	.978	4.350	813.835	.325	8833.358
N	1092	1039	1092	1084	1056

Note: This table reports saving and borrowing indicators at the individual level, including in the sample only members that belong to active teams. Note: ***, **, * indicates significance at the 1%, 5%, and 10% level respectively. In all columns, controls are selected through LASSO. Robust standard errors reported in parentheses below the coefficients are clustered at the village level.

6 Future Work and Conclusions

As you can see from our current draft, this is still a preliminary version in which we have highlighted some core results. However, a lot more can be extracted from the richness of our data. In fact, understanding at a much deeper extern the origin of this communication costs can be fundamental to unpack the mechanism even further. To do so, we created extended modules of team members bilateral interactions and assessments of other members cognitive and non-cognitive skills. We will test further hypotheses of how larger groups increase communication costs and do not allow for the best members to be recognized and to speak out. This might be a particularly strong mechanism in developing countries, especially, among women where empowerment and awareness of oneself skills is limited.

Despite our best efforts in this first RCT, a long list of questions on the internal organization of

teams still remain unanswered – for example, how is learning affected by the size of one's coworkers network? And who learns from whom? Do people learn from the *top*, the *bottom*, or the *average* coworkers? Finally, how does this learning translate into larger productivity and firm growth? What are the characteristics of a group that allow for innovation? We hope that future work tracking groups over an extended period of time will be able to address them.

References

Agha, L., K. M. Ericson, K. H. Geissler, and J. B. Rebitzer (2018). Team formation and performance: Evidence from healthcare referral networks. Technical report, National Bureau of Economic Research.

Bandiera, O., I. Barankay, and I. Rasul (2010). Social incentives in the workplace. *The review of economic studies 77*(2), 417–458.

Bassi, V., J. H. Lee, A. Peter, T. Porzio, R. Sen, and E. Tugume (2023). Self-employment within the firm. Technical report, National Bureau of Economic Research.

Becker, G. S. and K. M. Murphy (1992). The division of labor, coordination costs, and knowledge. *The Quarterly journal of economics 107*(4), 1137–1160.

Bonhomme, S. (2021). Teams: Heterogeneity, sorting, and complementarity. *arXiv preprint arXiv:2102.01802*.

Brooks, W., K. Donovan, and T. R. Johnson (2018). Mentors or teachers? microenterprise training in kenya. *American Economic Journal: Applied Economics 10*(4), 196–221.

Brooks, W., K. Donovan, and T. R. Johnson (2020). From micro to macro in an equilibrium diffusion model. *Globalization*.

Cai, J. and A. Szeidl (2018). Interfirm relationships and business performance. *The Quarterly Journal of Economics 133*(3), 1229–1282.

Chandler, A. D., T. Hikino, and A. D. Chandler (2009). *Scale and scope: The dynamics of industrial capitalism*. Harvard University Press.

Deloitte, B. (2016). Global human capital trends 2016, the new organization: different by design.

Devereux, K. (2018). Identifying the value of teamwork: Application to professional tennis. Technical report, Working Paper Series.

Dhake, S. P. and S. P. Narkhede (2019). Income-generating activity: An avenue to improve the sustainability of shgs. *Annals of public and cooperative economics 90*(4), 737–759.

Dupas, P. and J. Robinson (2013). Why don't the poor save more? evidence from health savings experiments. *The American Economic Review 103*(4), 1138–1171.

Freund, L. (2022). Superstar teams: The micro origins and macro implications of coworker complementarities. *Available at SSRN 4312245*.

Garicano, L. and E. Rossi-Hansberg (2006). Organization and inequality in a knowledge economy. *The Quarterly journal of economics 121*(4), 1383–1435.

Holmstrom, B. (1982). Moral hazard in teams. *The Bell journal of economics*, 324–340.

Hsieh, C.-T. and P. J. Klenow (2009). Misallocation and manufacturing tfp in china and india. *The Quarterly journal of economics 124*(4), 1403–1448.

Mas, A. and E. Moretti (2009). Peers at work. *American Economic Review 99*(1), 112–45.

OECD (2000). Small and medium-sized enterprises: local strength, global reach. *Policy Brief*.

Shaw, M. E. (1964). Communication networks. In *Advances in experimental social psychology*, Volume 1, pp. 111–147. Elsevier.

Sparrowe, R. T., R. C. Liden, S. J. Wayne, and M. L. Kraimer (2001). Social networks and the performance of individuals and groups. *Academy of management journal 44*(2), 316–325.

Williamson, O. E. (1967). Hierarchical control and optimum firm size. *Journal of political economy 75*(2), 123–138.

A Data Description

A.1 Data Collection and Organization

There was a gap of 10 days (6th Jan - 16th Jan) for making necessary changes to the survey instrument to shift from In-person Visits to Replacement Calls. Replacement Calls was launched on 16th Jan. In these 10 days, no group received In-person Visits or Calls.

We called any two women from the group we had of baseline - similar to the protocol followed in the call involving the information seed treatment.

Given that there are two respondents from each group, information on each meeting had to be harmonized to be collapsed into one observation per group.

Steps:

1. Exclude single-response observations and variables that are identifiers, member lists and categorical variables, and reshape them separately

2. Reshape the data to wide at TG level to have responses from 2 respondents in wide format

3. **Correcting inconsistencies** in responses between respondents for numeric and dummy variables

   (a) **For numerical variables:**

      i. Impute missing values based on responses of respondent that reported more group meetings

      ii. If both respondents reported meeting equally, take the means of the non-missing response of either respondent based on mismatch response in numeric variables

   (b) **For categorical variables:**

4. Merge back the identifiers, categorical variables and member lists to the corrected dataset. Then merge back single-response observations

Once the replacement call data was collapsed to one observation per group, we reshaped all the meeting data to wide to compare values of respondents, pick out any inconsistencies with meeting dates across replacement calls, in-person visits, and the second in-person visit, and merge all 3 datasets together.

Table A.1: Team Visits

	N. teams	N. teams that met at least once
First Visit (one month after BL)	49	29
Replacement Calls	*153*	*60*
First Visit (two and half months after BL)	144	13
Second Visit (four months after BL)	201	59

Note: Column 1 reports the number of teams surveyed in each visit; Column 2 reports the number of teams that reported meeting for production in the relevant preceding time period in each round, respectively. We managed to visit only 49 teams one month after baseline because of the Covid Omicron outbreak. During the Covid lockdown, we replaced in-person visits with phone-based interviews, managing to reach out remotely to 153 teams. Once the Covid restrictions were lifted, we met in person 144 out of the 153 teams previously reached out via phone. The second visit was performed in person to all teams (we successfully met 201 teams).

A.2 Balance Checks and Attrition Tables

Table A.2: Balance Checks - Group Level

Variable	N	(1) Small Mean/SE	N	(2) Medium Mean/SE	N	(3) Large Mean/SE	T-test Difference (1)-(2)	(1)-(3)	(2)-(3)
Share married	96	90.729 (1.504)	62	91.729 (1.154)	47	92.561 (0.949)	-1.000	-1.832	-0.832
Average respondent's age	96	33.960 (0.367)	62	33.185 (0.391)	47	33.887 (0.473)	0.775	0.074	-0.701
Average HH size	96	5.875 (0.107)	62	5.738 (0.131)	47	6.054 (0.198)	0.136	-0.180	-0.316
Average nb children	96	1.904 (0.057)	62	1.981 (0.053)	47	2.004 (0.064)	-0.077	-0.100	-0.023
Share general caste	96	3.958 (0.977)	62	2.110 (0.609)	47	3.766 (0.941)	1.848	0.192	-1.656
Share scheduled tribe and caste	96	38.229 (3.248)	62	34.884 (2.635)	47	36.729 (3.644)	3.346	1.500	-1.846
Share other backward classes	96	57.604 (2.835)	62	62.827 (2.835)	47	58.263 (3.893)	-5.223	-0.659	4.564
Completed at least secondary edu	96	63.646 (2.390)	62	65.654 (2.437)	47	63.265 (2.421)	-2.008	0.381	2.389
Share working for wage	96	44.306 (3.063)	62	46.756 (2.664)	47	44.780 (2.861)	-2.451	-0.475	1.976
Share running business	96	53.785 (2.513)	62	59.243 (3.165)	47	53.468 (3.422)	-5.458	0.317	5.775
Mean PC HH monthly income	96	2046.593 (197.254)	62	1419.656 (85.755)	47	1622.172 (150.250)	626.937***	424.420	-202.516
Mean wage last month	96	698.944 (61.951)	62	618.538 (57.446)	47	637.975 (61.004)	80.406	60.970	-19.437
Mean farm income last season	96	21564.250 (3509.945)	62	20893.859 (2096.889)	47	25762.989 (2868.252)	670.391	-4198.739	-4869.130
Mean business sales last month	96	1381.253 (212.778)	62	2346.692 (450.458)	47	1125.683 (211.969)	-965.439**	255.570	1221.009**
Mean PC HH overall monthly expenditure	96	7.028 (0.065)	62	6.966 (0.038)	47	6.947 (0.054)	0.061	0.081	0.020
Mean amount saved last month	96	1076.624 (105.031)	62	1016.465 (135.703)	47	987.206 (136.909)	60.160	89.419	29.259
Mean borrowed amount last 12m	96	29740.123 (2981.706)	62	26950.303 (2803.970)	47	28449.263 (2922.331)	2789.820	1290.860	-1498.961
Mean owning land	96	95.243 (1.046)	62	93.844 (1.152)	47	96.384 (0.831)	1.399	-1.141	-2.540**
Mean asset index KLK	96	-0.001 (0.034)	62	-0.061 (0.034)	47	-0.034 (0.038)	0.060	0.033	-0.027

Table A.3: Differences across Group, by likelihood of meeting

	(1) Did not Meet		(2) Met		(1)-(2) Pairwise t-test	
Variable	N/Clusters	Mean/(SE)	N/Clusters	Mean/(SE)	N/Clusters	Mean difference
Share married	39	89.808	165	91.929	204	-2.122
	36	(1.746)	124	(1.025)	145	
Share never married	39	1.154	165	1.707	204	-0.553
	36	(0.672)	124	(0.582)	145	
Share divorced or separeted	39	1.175	165	0.566	204	0.610
	36	(0.577)	124	(0.235)	145	
Share widowed	39	7.863	165	5.798	204	2.065
	36	(1.513)	124	(0.745)	145	
Share general caste	39	3.226	165	3.460	204	-0.233
	36	(1.333)	124	(0.787)	145	
Share schedules caste	39	11.004	165	12.384	204	-1.380
	36	(2.448)	124	(1.739)	145	
Share schedules tribe	39	23.462	165	24.863	204	-1.401
	36	(3.927)	124	(2.046)	145	
Share OBC	39	62.094	165	58.794	204	3.300
	36	(4.419)	124	(2.406)	145	
Mean age	39	33.468	165	33.816	204	-0.348
	36	(0.622)	124	(0.304)	145	
Mean household size	39	5.767	165	5.895	204	-0.128
	36	(0.190)	124	(0.097)	145	
Mean Nb children	39	1.921	165	1.943	204	-0.022
	36	(0.098)	124	(0.044)	145	

Table A.4: Balance Checks - Group Meeting

Variable	(1) Did not Meet N/Clusters	(1) Did not Meet Mean/(SE)	(2) Met N/Clusters	(2) Met Mean/(SE)	(1)-(2) Pairwise t-test N/Clusters	(1)-(2) Pairwise t-test Mean difference
Share no education	39	8.120	165	7.611	204	0.509
	36	(2.048)	124	(1.067)	145	
Share primary education	39	17.137	165	18.362	204	-1.226
	36	(2.582)	124	(1.369)	145	
Share secondary education	39	28.654	165	27.868	204	0.786
	36	(2.441)	124	(1.472)	145	
Share high School	39	41.368	165	39.124	204	2.243
	36	(3.978)	124	(1.729)	145	
Share graduate education	39	3.654	165	4.550	204	-0.896
	36	(1.112)	124	(0.635)	145	
Share post grad. education	39	1.068	165	2.409	204	-1.341
	36	(0.611)	124	(0.555)	145	
Share works for wage	39	42.714	165	45.494	204	-2.780
	36	(3.064)	124	(1.807)	145	
Share runs business	39	50.940	165	56.441	204	-5.501
	36	(3.594)	124	(2.002)	145	
Mean income top coded at p99	39	6894.324	165	7991.354	204	-1097.030
	36	(889.693)	124	(438.876)	145	
Mean farm income uncdt	39	21754.908	165	25273.529	204	-3518.620
	36	(4557.164)	124	(2875.550)	145	
Mean business sales uncdt	39	6190.304	165	2401.504	204	3788.799
	36	(4814.742)	124	(454.896)	145	

If the table includes missing values (.n, .o, .v etc.) see the Missing values section in the help file for the Stata command iebaltab for definitions of these values. Significance: ***=.01, **=.05, *=.1. Errors are clustered at the village level.

Table A.5: Differential Attrition at Endline: Ever met

VARIABLES	(1) Model 1	(2) Model 2
Small	-0.085	-0.090
	(0.071)	(0.073)
Large	-0.006	-0.006
	(0.082)	(0.083)
Mean grade completed		0.006
		(0.018)
Mean age		0.003
		(0.010)
Mean HH income		0.000
		(0.000)
Observations	201	201
R-squared	0.009	0.010
F-test Model 1	0.309	
F-test Model 2		0.290

Note: This table reports the attrition rates at endline comparing groups of different size. We also control for average of team's members main demographics. Robust standard errors are in parentheses. ***, **, * indicates significance at the 1%, 5%, and 10% level respectively.

B A Success Case Study in a Large Team: Muskan

Mudhipar, a village in Pithora, that is coloured by the crimson flames of mahua, is the place where the story of Muskan unfolds. It is located 50 kilometres east of the district's headquarters. As per the Census of 2011, the sex ratio of Mudhipar village is around 1067 compared to 991 which is the average in Chhattisgarh state. The literacy rate of Mudhipar village is 67.74% out of which 83.33% of males are literate and 53.13% of females are literate. The main source of livelihood in the village is agriculture and its allied sector. What makes Muskan stand apart from other ventures started by other training groups is their working structure.

Muskan was able to secure a spot in the local market in a short period of time due to its high quality. This was mainly because of their diligent follow-up of the guidelines and procedures mentioned in the training sessions. The soap-making video, which was shared on their smartphone during the training, was very helpful for them to watch whenever they were facing some issues in any phase in the soap-making process. Hitherto, the group has made a total of 852 soaps, earning an income of nearly Rs.10,600.

Wide usage of digital technology to support their production activities was another remarkable practice followed by the training group members, which has in fact become the cornerstone of Muskan's success. When technology becomes the linchpin of the production process, there will be more efficiency, participation and transparency within the group. This was very evident in this training group (TG), where digital tools became a key factor in arranging production meetings. A whatsapp group was started by the TG members immediately after the training. Meeting dates were decided after discussing in the WhatsApp group and the date was chosen based on the consensus. When it was difficult to arrange an in-person meeting because of covid restrictions, decisions were taken primarily through discussion in the WhatsApp group and phone conversations. Mobile phones were also commonly utilised to contact dealers and purchasers. Besides, whatsapp and Facebook were primarily used for advertising. The majority of the soap orders they received came from folks in other villages. As a result, digital advertising enabled them to reach a larger customer base while also being cost effective.

Muskan was able to effortlessly overcome all of the initial challenges that every new business confronted by employing cost-effective business methods. The group employed cost-effective business practices such as self-packaging of soaps using plastic food wraps, advertisement through WhatsApp and Facebook, cheaper stickers for labelling, moulds of smaller grams and different shapes for customisation.

Strict adherence to the group's regulations by all members has aided in the seamless operation of all group activities, which has had a significant impact on the muskan's development into the business enterprise of current stature. One such is levying fines, which is the most well appreciated and unique step taken by the group. A fine of Rs.100 is imposed on a member for her absence in any of the group meetings. When free-riding is a problem that affects every social institution, this group has successfully addressed the issue by levying a fine. Secondly, in order to sell soaps, each group member must purchase soaps from the group at the agreed-upon price and then sell them to clients as needed at the agreed-upon price—which cannot be raised or lowered.

All transactions, meeting details and group activities are recorded in a diary, which ensures transparency and accountability. This practice, introduced by the Mahasamund team, has been well-followed by the entire group since training.

The growth of Muskan was remarkable and it would not have happened without proper support and guidance. The group, throughout its journey, availed of all the support systems provided by the Mahasamund team, right from soap-making to its selling. For instance, When raw materials were in

short supply, the Mahasamund team helped them locate vendors in Raipur and Uttarakhand.

As of today, Muskan's journey is entering a new phase with new hopes and dreams. As part of their marketing strategy, they had planned to use innovative marketing tools such as switching from paper to box packages, advertising on all social media platforms, and circulating pamphlets.

C   Data Analysis Appendix

C.1   Data Construction

Initial data related to group production was collected at the meeting level, meaning that groups were asked to fill in a log of every meeting that they had, even if they did not end up producing. Production variables Meeting Duration, Production Duration, and Quantity Produced were summed across all the meetings a group had, whereas attendance measures were taken as the average number of people that attended all meetings. Productivity measures were calculated by taking the ratio of these totals, and in the case of attendance, dividing by average attendance instead of original group size.

Table C.6: Variable Definitions

Variable	Definition
A. Production Variables	
Meeting duration	Group meeting duration (in minutes), total of all group meetings
Production duration	Group production duration (in minutes), total of all group meetings
Quantity produced	Soap quantity produced by group during production meeting, total of all group meetings
PC Productivity (per hour)	Hourly rate of soap production, divided by original group size
B. Organization Variables	
Ability	(=1) if group divides tasks by ability
Individual preferences	(=1) if group divides tasks according to individual level preferences
Consensus	(=1) if group divides tasks by consensus
Everyone does all tasks	(=1) if all tasks are done by everyone
Work in subgroups	(=1) if group split into subgroups to produce
Use Monitoring	(=1) if group has women monitoring others during production to ensure productivity and good soap quality
Monitors by group size	Number of monitors divided by group size
Has group leader	(=1) if group has a leader
Work in subgroups	(=1) if groups split into subgroups to produce
C. Productivity Variables	
Hourly productivity	Total soap bars produced divided by production duration, in hours
Production PC (Size)	Total soap bars produced divided by original group size
Production PC (Attendance)	Total soap bars produced divided by average number of attending members
Productivity PC (Attended)	Total soap bars produced divided by average number of attending members, per hour of production duration
Share of high quality soaps	Share of high quality soaps produced out of the total
PC high quality production	Share of high quality soaps produced per capita
D. External Barriers Variables	
Input procurement issues	(=1) if group faced any problems related to procuring more raw materials
Marketing issues	(=1) if group faced any problems related to marketing of the soap product
Soap quality issues	(=1) if group faced any problems related to the quality of soaps
E. Individual-level Variables	
Owns savings account	(=1) if owns a personal savings account under their name in any financial institution
Account usage (frequency)	(=1) if savings account is used more than once a week, (=5) if less tham once a month
Savings amount	Money saved in the past 30 days
Borrowed recently	(=1) if HH borrowed money from someone since Baseline date
Amount borrowed	Amount HH borrowed since Baseline date
Phone access	(=1) if had access to any mobile phone at home in the past 30 days
Autonomy KLK	KLK index of the degree of autonomy in household financial, health, education, and food-consumption decisions
Illness	(=0) if never feels ill, (=1) if has felt ill
Worry about money	(=0) if during last week never worried about providing basic needs, (=1) if during last week has worried
Kessler Scale	Kessler scale to indicate wellbeing, composed of how often members felt nervous, depressed, worthless, restless, and worn out. A higher score indicates worse mental health
Future expectations	In one year from now expects that life will be better(=0), will be more or less the same or worse (=1)

Table C.7: Treatment Effects on Production Conditional on Producing

	Meeting Duration	Production Duration	Quantity Produced	PC Productivity (per hour)
	(1)	(2)	(3)	(4)
Small	-56.204**	-14.494	-9.489***	1.294***
	(26.785)	(9.880)	(3.345)	(0.346)
Medium	-38.696	-2.843	-2.871	0.518
	(26.721)	(10.753)	(3.905)	(0.315)
Small = Medium	0.386	0.203	0.025	0.050
Mean (Large)	188.646	72.083	21.896	1.701
N	114	114	114	114

Table C.8: Meeting Frequency and Attendance

	Times Met (1)	Attendance (2)	Attendance (Normalized) (3)
Panel A: Full Sample			
Small	-1.742	-2.451***	-0.036
	(1.076)	(0.556)	(0.056)
Medium	-1.813	-0.748	0.055
	(1.196)	(0.630)	(0.064)
Small = Medium	0.887	0.000	0.119
Mean (Large)	3.213	3.809	.317
N	206	201	201
Panel B: Conditional on Meeting			
Small	-2.274*	-2.851***	-0.024
	(1.337)	(0.605)	(0.062)
Medium	-2.297	-0.751	0.086
	(1.456)	(0.681)	(0.070)
Small = Medium	0.969	0.000	0.085
Mean (Large)	4.081	4.632	.386
N	167	166	166

	Productivity (per hour)	PC Production (Size)	PC Production (Attendance)	PC Productivity (per hour)	PC Productivity (by attendance, per hour)	Share of High Quality Soaps	High Q
	(1)	(2)	(3)	(4)	(5)	(6)	
Small	-5.963**	0.558	0.681	1.294***	2.208**	-0.067	
	(2.869)	(0.346)	(0.544)	(0.346)	(0.955)	(0.043)	
Medium	-2.644	0.479	0.405	0.518	0.913	-0.081*	
	(2.949)	(0.402)	(0.597)	(0.315)	(1.016)	(0.045)	
Small = Medium	0.159	0.851	0.661	0.050	0.275	0.784	
Mean (Large)	20.407	1.825	3.565	1.701	3.617	.928	
N	114	114	114	114	114	114	

	Quantity Produced	Production Duration	Soap Quality Rating	Number of Soap Quality	Visit Productivity	PC Visit Productivity (Attendance- Time)	Visit PC Productivity (Size- Time)	Vis PC Produ (Attend
	(1)	(2)	(3)	(4)	(4)	(5)	(6)	(7)
Small	-2.084***	-1.610	0.107	-0.350	-3.394***	1.021***	0.979***	0.480
	(0.488)	(1.581)	(0.274)	(0.564)	(0.997)	(0.206)	(0.135)	(0.06
Medium	0.234	2.510	0.266	0.328	-0.409	0.678***	0.565***	0.358
	(0.769)	(1.893)	(0.270)	(0.559)	(1.195)	(0.256)	(0.140)	(0.09
Small = Medium	0.005	0.035	0.545	0.102	0.020	0.231	0.023	0.26
Mean (Large)	7.721	33.7	2.256	2.977	14.069	1.8	1.172	.643
N	190	180	190	190	180	180	180	190

	Productivity (per hour)	PC Production (Size)	PC Production (Attendance)	PC Productivity (per hour)	PC Productivity (by attendance, per hour)	Share of High Quality Soaps	High Q
	(1)	(2)	(3)	(4)	(5)	(6)	
Small	-5.963**	0.558	0.681	1.294***	2.208**	-0.067	
	(2.869)	(0.346)	(0.544)	(0.346)	(0.955)	(0.043)	
Medium	-2.644	0.479	0.405	0.518	0.913	-0.081*	
	(2.949)	(0.402)	(0.597)	(0.315)	(1.016)	(0.045)	
Small = Medium	0.159	0.851	0.661	0.050	0.275	0.784	
Mean (Large)	20.407	1.825	3.565	1.701	3.617	.928	
N	114	114	114	114	114	114	

Table C.9: Treatment Effects on Mental Health, Expectations and Physical Health

	Kessler Scale	Worry About Money	Future Expectations	Illness
	(1)	(2)	(3)	(4)
Small	0.234	0.016	0.027*	0.011
	(0.260)	(0.027)	(0.016)	(0.027)
Medium	0.033	-0.023	-0.006	0.022
	(0.259)	(0.031)	(0.016)	(0.027)
P-values				
Small = Medium	0.402	0.238	0.075	0.688
Mean (Large)	3.33	.468	.055	.2
N	1443	1432	1443	1432

Note: ***, **, * indicates significance at the 1%, 5%, and 10% level respectively.

C.2 Controlling for Covid

In the previous regressions, we had a dummy variable controlling for whether a group was impacted by covid, as reported by that group. Here, we also present the tables as in the main text, but this time controlling for whether the group had to receive a replacement call because of the covid lockdown.

Table C.10: Treatment Effects on Production Process

	Meeting Duration	Production Duration	Quantity Produced	PC Productivity (per hour)
	(1)	(2)	(3)	(4)
Panel A: Full Sample				
Small	-34.056	-8.601	-5.286**	0.624**
	(20.635)	(7.650)	(2.542)	(0.290)
Medium	-14.727	2.120	-0.366	0.375
	(23.215)	(9.010)	(3.189)	(0.295)
Small = Medium	0.245	0.140	0.029	0.444
Mean (Large)	103.278	38.444	11.678	.907
N	204	204	204	204
Panel B: Conditional on Meeting				
Small	-44.448*	-11.552	-7.248**	0.673**
	(24.082)	(9.143)	(3.055)	(0.334)
Medium	-17.741	2.847	-0.943	0.395
	(25.803)	(9.914)	(3.720)	(0.368)
Small = Medium	0.141	0.071	0.014	0.439
Mean (Large)	132.786	49.429	15.014	1.166
N	165	165	165	165

Note: ***, **, * indicates significance at the 1%, 5%, and 10% level respectively. In all columns, controls are selected through LASSO. Robust standard errors reported in parentheses below the coefficients are clustered at the village level. Outcome variable definitions are available in C.6. Meeting duration, production duration, and quantity produced are calculated by summing over all group meetings for each group. PC Productivity per hour is calculated by taking the ratios of total quantity produced to meeting duration (in hours), and dividing it by the original group size. Sample size in Panel B corresponds to the number of groups that met, but did not necessarily produce soap.

Table C.11: Treatment Effects on Production Organization

	Ability	Individual Preferences	Consensus	Everyone Does all Tasks	Work in Subgroups
	(1)	(2)	(3)	(4)	(5)
Panel A: Full Sample					
Small	0.044	0.076**	0.088***	-0.287***	-0.061
	(0.037)	(0.035)	(0.033)	(0.093)	(0.049)
Medium	0.025	0.011	0.007	0.001	0.065
	(0.042)	(0.020)	(0.023)	(0.105)	(0.063)
Small = Medium	0.633	0.142	0.056	0.002	0.050
Mean (Large)	.022	0	0	.644	.07
N	204	204	204	204	190
Panel B: Conditional on Meeting					
Small	0.043	0.081*	0.113**	-0.386***	-0.089
	(0.049)	(0.043)	(0.045)	(0.101)	(0.059)
Medium	0.021	-0.008	0.013	-0.029	0.067
	(0.056)	(0.027)	(0.028)	(0.101)	(0.072)
Small = Medium	0.673	0.118	0.054	0.000	0.041
Mean (Large)	.029	0	0	.829	.086
N	165	165	165	165	161

Note: ***, **, * indicates significance at the 1%, 5%, and 10% level respectively. In all columns, controls are selected through LASSO. Robust standard errors reported in parentheses below the coefficients are clustered at the village level. Variables in columns (1)-(4) take on the value 1 if a group divides tasks according to ability, individual preferences, consensus, or by assigning to all tasks to everyone, respectively, and 0 otherwise. Column (5)'s "Work in Subgroups" takes the value 1 if the group divided in order to produce. Outcome variable definitions are available in C.6. Sample size in Panel B corresponds to the number of groups that met, but did not necessarily produce soap.

Table C.12: Treatment Effects on Monitoring Production

	Use Monitoring	Monitors by Group Size	Has Group Leader
	(1)	(2)	(3)
Panel A: Full Sample			
Small	-0.047	0.025	0.020
	(0.082)	(0.018)	(0.052)
Medium	0.081	0.038*	0.006
	(0.084)	(0.019)	(0.058)
Small = Medium	0.096	0.629	0.756
Mean (Large)	.2	.019	.93
N	204	199	191
Panel B: Conditional on Meeting			
Small	-0.010	0.026	0.029
	(0.097)	(0.023)	(0.053)
Medium	0.121	0.044*	-0.007
	(0.096)	(0.023)	(0.065)
Small = Medium	0.135	0.561	0.348
Mean (Large)	.2	.024	.9430000000000001
N	165	164	162

Note: ***, **, * indicates significance at the 1%, 5%, and 10% level respectively. In all columns, controls are selected through LASSO. Robust standard errors reported in parentheses below the coefficients are clustered at the village level. Outcome variable definitions are available in C.6. Sample size in Panel B corresponds to the number of groups that met, but did not necessarily produce soap.

Table C.13: Treatment Effects on Production Process- Technology Diffusion

	Meeting Duration	Production Duration	Quantity Produced	PC Productivity (per hour)
	(1)	(2)	(3)	(4)
Small	-17.387	-0.825	-2.020	0.855**
	(25.256)	(10.251)	(2.874)	(0.392)
Medium	-48.492*	-7.921	-3.708	-0.131
	(27.060)	(12.341)	(3.137)	(0.303)
Small × Info Seeds	-13.543	-8.062	-2.039	-0.153
	(16.987)	(7.258)	(1.807)	(0.449)
Medium × Info Seeds	74.614***	23.289**	9.642***	1.136***
	(22.896)	(11.536)	(3.251)	(0.332)
Large × Info Seeds	19.148	7.188	4.767	0.307
	(36.809)	(13.568)	(4.752)	(0.363)
Small = Medium	0.099	0.477	0.474	0.013
Small × Info Seeds = Medium × Info Seeds	0.003	0.025	0.002	0.019
Small × Info Seeds = Large × Info Seeds	0.429	0.331	0.197	0.427
Medium × Info Seeds = Large × Info Seeds	0.203	0.371	0.391	0.089
N	203	203	203	203

Note: ***, **, * indicates significance at the 1%, 5%, and 10% level respectively. In all columns, controls are selected through LASSO. Robust standard errors reported in parentheses below the coefficients are clustered at the village level. Outcome variable definitions are available in C.6. Sample size $N = 114$ corresponds to groups that recorded producing soap.

Table C.15: Treatment Effects on External Barriers

	Input Procurement Issues	Marketing Issues	Soap Quality Issues
	(1)	(2)	(3)
Small	-0.119	-0.090	-0.143
	(0.123)	(0.121)	(0.104)
Medium	-0.259**	-0.043	-0.053
	(0.119) 44	(0.121)	(0.111)

Table C.14: New Technology and Task Division among Different Size Teams

	Ability	Individual Preferences	Consensus	Everyone Does all Tasks	Work in Subgroups
	(1)	(2)	(3)	(4)	(5)
Small	0.093	0.075	0.131	-0.323***	-0.069
	(0.064)	(0.062)	(0.083)	(0.110)	(0.088)
Medium	0.028	-0.037	-0.026	-0.150	0.240
	(0.058)	(0.029)	(0.054)	(0.146)	(0.149)
Small × Info Seeds	0.053	-0.093	-0.031	-0.158	0.109*
	(0.091)	(0.066)	(0.084)	(0.140)	(0.055)
Medium × Info Seeds	0.040	0.046	0.010	0.098	-0.145
	(0.079)	(0.052)	(0.044)	(0.158)	(0.138)
Large × Info Seeds	0.125	-0.003	-0.069	-0.124	0.162
	(0.087)	(0.023)	(0.043)	(0.114)	(0.136)
Small = Medium	0.377	0.105	0.079	0.312	0.035
Small × Info Seeds = Medium × Info Seeds	0.913	0.135	0.661	0.250	0.113
Small × Info Seeds = Large × Info Seeds	0.561	0.209	0.689	0.850	0.713
Medium × Info Seeds = Large × Info Seeds	0.486	0.378	0.196	0.243	0.131
N	114	114	114	114	113

Note: ***, **, * indicates significance at the 1%, 5%, and 10% level respectively. In all columns, controls are selected through LASSO. Robust standard errors reported in parentheses below the coefficients are clustered at the village level. Outcome variable definitions are available in C.6. Sample size N = 114 corresponds to groups that recorded producing soap.

C.3   Team Interaction Data

Table C.16:   Team Connections

Variable: Select the best at	Mean	Std	Min	Max	More than 3	All Team	N
Organizing Meetings	1.34199	1.14764	1	12	40	14	1155
Being active & motivated	2.08997	2.36618	1	12	156	102	1156
Creativity	1.63754	1.71147	1	12	88	53	1156
Being hardworking	2.41696	2.68151	1	12	231	123	1156
Communicating with (you)	1.82526	1.45902	1	12	79	19	1156
(You) Learning from	1.4282	1.54333	1	12	55	28	1156

Note: The survey asked group members to list the members who had the best organization skills, were most active in the group, were most creative, hardworking, the ones whom they talked to most and learned from the most.

Table C.17: Number of (Normalized) Responses

	Talk on Phone	Meet in Person	Talk Production	Support in Production
	(1)	(2)	(3)	(4)
Small	-0.008	0.031	-0.023	-0.022
	(0.048)	(0.031)	(0.044)	(0.031)
Medium	0.055	0.043	-0.016	-0.001
	(0.044)	(0.034)	(0.046)	(0.034)
Small = Medium	0.165	0.678	0.852	0.448
Mean (Large)	.472	.736	.169	.11
N	1388	1388	1388	1388

Note: ***, **, * indicates significance at the 1%, 5%, and 10% level respectively. In all columns, controls are selected through LASSO. Robust standard errors reported in parentheses below the coefficients are clustered at the village level. Each column represents the number of group members each group member has interacted with (talked to on the phone, in person, talked during production, or has helped during production), divided by the group size. Outcome variable definitions are available in C.6. Sample size differs by individuals that received the question in the survey.