Improving Gender Norms in the Workplace *
Pre-Results Paper

Emily A. Beam[†] Joshua D. Merfeld[‡] Naveen Wickremeratne[§]

March 17, 2025

Abstract

Gender norms shape women's access to employment opportunities and their experiences in the workplace, with potential implications for firm productivity. We conduct a randomized controlled trial in Bangladesh, involving 5,000 workers and firm owners across 1,900 small and medium-sized enterprises (SMEs), to explore the impact of promoting gender-equitable norms. Workers and firm owners in randomly selected markets participate in an intensive three-day gender norms training program. We assess the impact of this intervention on gender attitudes, women's workplace experiences, inter-employee interactions, and firm productivity using a combination of self-reported survey data and lab-in-the-field experiments. By examining the malleability of gender norms and their influence on workplace dynamics and productivity, this study contributes to the design of policies and interventions aimed at fostering more equitable work environments and enhancing firm performance.

Keywords: gender norms, productivity, women's employment, discrimination, field experiment
JEL Codes: O15, J16, J24, J71, L25

*This work is supported by the International Development Research Centre (IDRC), Canada, and the Private Enterprise Development in Low-Income Countries (PEDL) Major Research Grant. The Institutional Review Board (IRB) of the BRAC James P. Grant School of Public Health, BRAC University, provided the ethics review for this study, with IRB approval number IRB-13 February'23-003. This study would not have been possible without the excellent research assistance from Nuzaira Binte Neaz and Raisa Rahman. We thank Sukanto Roy, Ehsun Newaz, and the entire GDRI team for their support throughout the intervention and analysis, and Sanaiyya Ansari for her expertise in developing the intervention. We are also grateful for the support of Narayan C. Das, Sheikh Touhidul Haque, Imran Matin, and Munshi Sulaiman at the BRAC Institute of Governance and Development (BIGD). We are grateful to the BRAC Skill Development Program (SDP) for their support, including listing firms for the training and surveys.

[†]Department of Economics, University of Vermont; IZA
[‡]KDI School of Public Policy and Management, IZA
[§]Department of Economics, Monash University

1 Introduction

Constraints on women's economic opportunities impede their active contribution to economic development. Restrictive gender norms affect women's access to employment by limiting avenues for employment or advancement and creating unpleasant or unsafe workplace environments (Jayachandran, 2021; Merfeld, 2023). However, these norms are also malleable and responsive to direct training or information provision (Bursztyn et al., 2020; Dhar et al., 2022), as well as broader societal changes (Seguino, 2007).

Barriers to women's labor force participation can harm the well-being of women who would otherwise want to engage in market work, which in turn can have larger-scale impacts by limiting women's decision-making power in the home and society more broadly (Sen, 1990; Kessler-Harris, 2003; Jayachandran, 2021). Additionally, increasing work opportunities and income for women directly improves outcomes for women and children (Rosenzweig and Schultz, 1982; Duflo, 2003; Qian, 2008; Jensen, 2012; Heath and Jayachandran, 2017).

Gender norms can also directly affect firm productivity. Bias in hiring can prevent firms from identifying high-quality female employees (e.g. Hsieh et al., 2019). Strained or limited interactions between male and female employees may affect productivity through reduced knowledge sharing and teamwork (Adhvaryu et al., 2023). Poor working conditions (Blattman and Dercon, 2018), hostile or unwelcoming work environments (Ilmakunnas et al., 2005), and barriers to advancement (Hersch, 1991) increase employee turnover, generating additional recruitment and training costs. High turnover rates are associated with lower levels of productivity and overall firm growth (Hancock et al., 2013), and these effects may be largest for small firms (Li et al., 2022), which are the focus of our study.

The promotion of egalitarian gender norms in the workplace can improve the lives of women and increase the efficiency of the firms in which they work. To this end, we conduct a randomized experiment with 5,000 owners and workers at 1,900 small and medium-sized enterprises (SMEs) across Bangladesh to understand the role of gender norms on women's

workplace experiences, inter-employee interactions, and firm productivity. We use surveys and lab-in-the-field experiments to measure manager and employee gender attitudes, and we develop and implement an intensive gender sensitivity curriculum with owners and employees within randomly selected markets. We then measure the impact of this intervention on gender attitudes, workplace practices, and productivity in the short and longer term.

We developed a three-day gender sensitivity workshop by collaborating with local gender experts and drawing elements from BRAC's current gender equality and skills development programming and current best practices (CARE, 2014; Dhar et al., 2022). The program comprises two full-day workshops that bring together workers and owners from the area. Roughly 82% of the 2,500 invited workers and owners participated in the first workshop, held in March 2024, and 87% joined the second workshop, held in May and June. A third half-day workshop consists of firm visits and separate meetings with owners and workers to discuss practical issues specific to each firm. During these meetings, facilitators work with participants to explore potential solutions, review steps taken following the previous two trainings, and address challenges encountered.

We anticipate that training will promote equitable gender attitudes among workers and owners, particularly men. We measure attitudes in three ways: self-reported general gender attitudes, following Dhar et al. (2022), self-reported attitudes about women's productivity in the workplace, and an incentivized activity. We anticipate that the training will lead owners to pursue more equitable hiring practices and take steps to improve workplace conditions for women, such as increasing the availability of nearby toilets or enacting women-friendly workplace policies.

We hypothesize that changes in gender attitudes will improve workplace productivity, and we outline a conceptual framework that demonstrates this relationship. We measure the impact of training on productivity through firms' reported revenues and profits as well as through an incentivized envelope-making activity. In this activity, participants work in same-gender or

mixed-gender pairs to cut, fold, and seal envelopes. This activity's high returns to effective collaboration will enable us to measure differentials in productivity by gender composition while holding the activity and measurement constant.

We use both survey measures and lab-in-the-field experiments to investigate mechanisms driving our results. We measure impacts on mixed-gender employee interactions, perspective-taking toward women's workplace experiences, as well as impacts on trust, altruism, and cooperation in mixed-gender groups. We also test alternative channels, such as if the training improves cooperation and employee relationships more generally, increases empathy, improves women's empowerment and productivity, or enhances worker-manager relationships.

Our study makes two main contributions to the growing body of knowledge on gender and firms in developing countries. First, we provide novel evidence on the malleability of gender norms in the workplace and assess the impact of changing gender norms on downstream outcomes such as firm hiring practices and trust. We build on a broad literature that shows training programs can directly improve women's empowerment and outcomes (e.g. Bandiera et al., 2020; Ashraf et al., 2020). Additionally, programs that target and support women—such as microcredit for female entrepreneurs (De Mel et al., 2013) or initiatives that enhance women's control over their earnings (Field et al., 2021)—can also promote empowerment and egalitarian gender norms, although there may be risks of backlash (Angelucci, 2008).[1]

A smaller body of literature has examined training programs for men or couples, with more mixed results. For example, couples training aimed at increasing men's engagement in reproductive health and caregiving has shown positive effects on joint decision-making and reductions in intimate partner violence (Doyle et al., 2018), while other programs focused on training men on topics like intimate partner violence have been less effective (Angelucci et al., 2023). An intervention similar to ours, targeting gender attitudes among adolescent boys and girls, produced lasting changes in gender attitudes (Dhar et al., 2022). However,

[1]See Chang et al. (2020) for a more comprehensive review.

our study is distinct in its workplace setting, where changes in gender attitudes can directly affect women's work experiences and firm productivity, and in its focus on an older adult population.

Our second contribution is investigating the causal link between gender norms and productivity and unpacking the nature of this relationship. We will measure the impact of training on worker productivity through lab-in-the-field experiments, as well as through changes in firm revenues and profits. Additionally, we will explore alternative channels, such as improved trust, altruism, and workplace communication. For example, Alan et al. (2021) found that a workplace climate improvement program strengthened relationships between leaders and subordinates and reduced employee separation. By including both workers and firm owners in our study, we will be able to measure the impacts on both worker behavior and management practices, which are critical to working conditions and firm productivity.

The results of this study will be valuable to private firms looking to understand and improve gender attitudes among employees, helping to identify best practices for fostering positive inter-employee collaboration and reducing worker turnover. For policymakers and NGOs in Bangladesh and beyond, this study will document gender attitudes in the workplace and offer pathways to enhance women's economic agency and promote gender equality. These objectives are crucial in their own right and key to improving women's agency and enhancing the well-being of women and children (Duflo, 2003; Heath and Jayachandran, 2017; Jensen, 2012; Qian, 2008; Rosenzweig and Schultz, 1982).

2 Background and intervention

2.1 Background

Despite Bangladesh's rapid economic progress over recent decades, improvements in gender equality has lagged (Aregu et al., 2018; Chandramohan et al., 2023). The UN Gender Social

Norms Index (GSNI) indicates that societal gender biases in Bangladesh hinder women's access to greater economic opportunities (UNDP, 2023). These biases are deeply rooted in long-standing social norms (Ahmed and Sen, 2018; Haque et al., 2022), further reinforced by labor market disparities, particularly outside the country's ready-made garment (RMG) sector (Balk, 1997; Blunch and Das, 2015; UNDP, 2023). These norms not only affect women's ability to seek work outside the home (Jayachandran, 2021), but also they influence employers' willingness to hire them (Buchmann et al., 2023b). Additionally, norms may affect women's experiences within the workplace. Boudreau et al. (2023) find that harassment is widespread in large garment factories and that current estimates may be underreported.

In Bangladesh, female labor force participation remains strikingly low at 36.3 percent, compared to 80.5 percent for men (Bangladesh Bureau of Statistics, 2018). Although women comprise 61% of the RMG workforce (Matsuura and Teng, 2020), their representation is much lower across other sectors, including SMEs, which comprise nearly all businesses in Bangladesh and account for 70–80% of non-agricultural employment (Hossin et al., 2023). Against this backdrop, Bangladesh offers a unique context to examine the relationship between gender attitudes, workplace interactions, and firm productivity.

Several recent studies have examined efforts to shift gender norms within the social and labor market spheres in Bangladesh. Buchmann et al. (2023a) evaluated a program that provided economic incentives and empowerment programs to promote egalitarian gender norms and reduce child marriage. Although economic incentives reduce underage marriage, the authors find no impacts of the empowerment program on adolescent marriage. In the garment sector, Macchiavello et al. (2020) find that women are sub-optimally less likely to be promoted to managerial roles. Although female managers are initially less productive and evaluated more poorly than male managers, this gap is driven by male workers' negative beliefs about the abilities of their female supervisors, and it resolves quickly.

2.2 Intervention

The intervention jointly engaged men and women working in employee and managerial roles to shape gender attitudes and enhance workplace communication. This hands-on curriculum was developed by a local expert consultant specializing in gender issues and rights in Bangladesh. She collaborated closely with the research team, incorporating elements from BRAC's current gender equality and skills development programming, along with best practices from other sources (CARE, 2014; Dhar et al., 2022). Appendix E outlines the training curriculum.

The training consisted of three days of workshops lasting approximately 16 hours. The curriculum began with a full-day interactive workshop for employees and owners, covering key topics such as gender equality and stereotyping, promoting teamwork and cooperation, improving communication, and fostering a woman-friendly workplace to improve the recruitment, retention, and advancement of female workers. During this workshop, participants developed action plans to increase the representation of female workers and create an environment for reporting and discussing challenges faced by women to create a gender-sensitive workplace. The second day of training took place approximately three months later and focused on specific strategies and challenges to cultivating a women-friendly workplace. Participants reviewed their progress toward their initial commitments and revised their goals for the coming months.

On both workshop days, we presented two 4–5 minute videos that we developed to showcase the stories of two successful women in male-dominated SME sectors, followed by a debriefing and discussion. The vignettes highlighted their experiences and competence, challenged societal norms about suitable jobs for women, emphasized the importance of recognizing individual abilities regardless of gender, and underscored the crucial role of societal support in creating a gender-inclusive workplace.[2]

[2]The first video portrays Saleha, challenging traditional gender roles by working as a clothes ironer in a shop. Her story sparks discussions about gender stereotypes and biases among the laundry owner and others. The

The third day of training involved workplace visits held approximately two months after the second day of training. The trainer met separately with owners and workers to discuss the steps they had taken to meet their previous commitments and to address potential challenges via a set of case studies.

After each training, the trainers participated in additional refresher sessions to address any identified issues and integrate lessons learned from their participation in the training. On the third day, we used the revised training materials based on feedback from the pilot and the group-based training in the first two days.

We anticipate that gender training will directly influence gender attitudes and workplace interactions, while the monthly coaching will increase women's empowerment. Additionally, these interventions may lead to greater cooperation and productivity either directly or through the gender attitude and workplace interaction channels. Section 3 discusses mechanisms in more detail.

2.2.1 Workshop implementation

The first two training days were held in central locations within the study upazilas, selected for the convenience of the invited owners and workers. We targeted 25–35 participants per session and delivered the initial training invitations in person to the targeted owners and workers. Workers received 500 Bangladesh taka (BDT), or about \$4.50[3] per day of training, and owners received 800 BDT, or about \$7.15, to account for travel costs and lost wages. All participants were provided with lunch and snacks. Additionally, participants received a small token (jute bag) for their participation in the third day of training, which was held at their firm.[4]

second video features Monira, a carpenter in a traditionally male-dominated field. Despite facing societal barriers and criticism, she finds encouragement and support from her mentor.

[3]This and all subsequent references to BDT converted to USD at a rate of 1 USD = 111.82 BDT, the average rate from February–June 2024 (OANDA, 2024).

[4]Other workers were permitted to attend the training, but they were not compensated.

Prior to each workshop day, we held training sessions for the trainers, followed by a brief pilot. We refined materials based on these pilots and then held a second refresher training before launching the workshops.

The intensity of the training aligns with recently published studies that found detectable impacts on gender attitudes. The key outcomes of interest are gender attitudes, job satisfaction, reports of harassment and discrimination, employee turnover, and cooperation and productivity. The workshops are designed to be interactive and participatory, with participants organized into small groups to ensure active engagement. Trainers facilitate discussions aimed at encouraging participants to share their knowledge and experiences regarding gender, societal norms, and discrimination. Through dialogue, participants gained insights into the experiences and perspectives of women in both the workplace and society at large. Additionally, they are expected to gain skills to effectively communicate across diverse backgrounds and viewpoints.

2.3 Sampling frame

We derive our sample based on a conducted a firm-level listing exercise (census) with 4,754 firms in two rounds from April–May 2023. We added to this frame 864 firms participating in BRAC's skill development program (SDP).[5] This survey spanned 21 districts and 88 sub-districts, providing a broad snapshot of the diverse economic landscape in Bangladesh. We chose these 21 districts based on the location of operations of BRAC's SDP current cohorts to ensure operational and logistical support from BRAC during delivery of the intervention. Enumerators canvassed each market within the selected sub-districts, listing all shops or enterprises that (1) had a permanent structure, (2) had at least one employee, and (3) fell into a set of pre-identified firm types, including tailoring and garment making, canteens, retail or wholesale stores, and repair shops, among others. Additional details on firm selection are

[5]BRAC's SDP is a six-month apprenticeship program designed for adolescents (aged 14–18) who have dropped out or are at risk of dropping out of formal schooling.

in Appendix B.1.

After categorizing the firms, we selected 2,000 enterprises for our baseline survey and kept an additional 200 on the waitlist, with a focus on maximizing the number of BRAC SDP firms, maximizing the number of manufacturing firms, and ensuring high representation of mixed-gender firms, as well as a balance of male-only and female-only firms. We have a high share of tailoring and garment-making firms, as well as small food shops, restaurants, or canteens in our sample. Due to the peri-urban nature of the sample, we do not include agriculture-oriented firms.

The baseline survey was conducted with 1,888 out of the selected 2,200 firms from 16 September–20 October 2023. During the baseline survey, we interviewed up to two male and two female randomly selected employees in each firm, as well as the manager, for a total of 1,888 owners and 3,207 workers. We only surveyed employees who are (1) not members of the owner's household and (2) work regularly at the firm (versus temporarily). Appendix B.2 provides additional details on worker selection.

2.4 Randomization

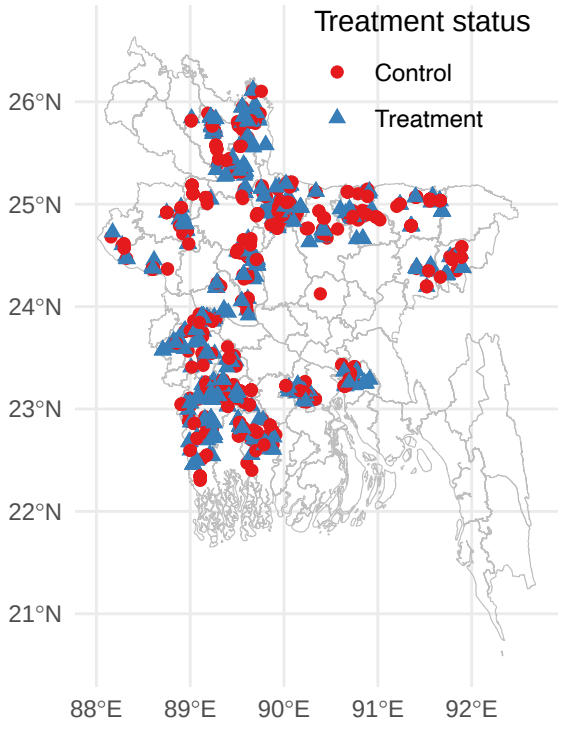
We randomize firms into treatment and control groups at the market level.[6] Out of a total of 803 markets surveyed, 403 markets were randomly allocated to the control group and 400 markets to the treatment group. This resulted in 971 firms in the control group and 917 firms in the treatment group. The workshops were held in locations that would be central for multiple firms and markets, with 96 workshops scheduled to accommodate the 917 invited firms across the country.

Figure 1 presents the location of treatment and control firms throughout the country, showing we have broad geographic coverage outside of the capital city, Dhaka, and the second-largest, Chittagong. SMEs in these locations were excluded because they are more likely to have

[6]The cluster is "market,", defined as the set of firms located in the same geographic space.

access to better resources, infrastructure, and support services, which can influence the outcomes of the intervention.[7] By focusing on SMEs outside the biggest two metropolitan centers of the country, we aim to better understand the intervention's impact in settings that more closely resemble the majority of the country's SME landscape.

Figure 1: Map of treatment and control firms



2.5 Power calculations

Our primary outcome is gender attitudes, and we calculate the minimum detectable effect size with a 5% significance level at 80% power, based on an index of gender attitudes standardized to the control group, our primary outcomes. We calculate the intra-cluster correlation coefficient (ICC) to be 0.1014 (for firm owners) and 0.1858 (for workers) based on the baseline gender attitudes index using the `loneway` command in Stata.

With 403 markets in the control group and 404 markets in the treatment group, there is

[7]Additionally, these areas have higher levels of economic activity and competition, which could affect the generalizability of the study results to other regions.

an average of 2.41 firms per market in the control group and 2.27 firms per market in the treatment group. On this basis, we have an MDE of 0.14 standard deviations (s.d.) at 80% power for owners. Moreover, we have around 4.1 workers per market in the control group and 3.9 workers per market in the treatment group. Thus, we are powered for a MDE of 0.12 standard deviations at 80% power for workers in our sample.

Compared with the documented 0.18 s.d. change in gender attitudes generated by an intervention of similar intensity but with adolescents in schools Dhar et al. (2022), this study is well-powered to detect impacts on gender attitudes, although we note that we may expect smaller effect sizes because we are working with adults, whose beliefs may be less malleable.

We are well-powered for our primary productivity outcome, the number of acceptable quality envelopes produced in mixed-gender pairs. To ensure gender balance, we adjust our sample sizes to align the proportions of firm owners and workers. This results in a balanced sample comprising 628 firm owners and 2,720 workers, distributed across treatment and control groups. On average, there are 0.87 firm owners per market in the control group and 0.69 firm owners per market in the treatment group. Similarly, the control group has 3.5 workers per market, compared to 3.3 workers in the treatment group. In the absence of baseline productivity experiment measures, we use an intra-cluster correlation coefficient (ICC) based on the average ICC values from related survey-based baseline measures. This yields an ICC of 0.13 for firm owners and 0.19 for workers. We have 80% power to detect an MDE of 0.22 standard deviations (s.d.) for owners and an MDE of 0.13 standard deviations (s.d.) at 80% power for workers in our sample.

2.6 Balance

Table 1 shows the mean characteristics of owners for the treatment and control groups, as well as mean differences between them. We present simple means for the treatment and control groups in the first two columns. Since we stratified based on the market-level gender

composition of firms and division, with 22 total strata, we include stratification-cell fixed effects when testing for balance for individual covariates and overall. For p-values of the F-test, we perform randomization inference, following issues around over-rejection of the null reported in Cattaneo et al. (2018) and recent suggestions on remedies in Kerwin et al. (2024).[8]

We find no significant differences in any individual outcome in Table 1 between firms in treatment and control markets at the five-percent level, though secondary education is significant at the 10-percent level. We also note that since a sub-sample of our firms come from BRAC's Skill Development Program, we explicitly test for balance on this variable in the last row of the table. We find no significant or qualitatively large difference. When calculating an F-test for joint significance of all outcomes, we cannot reject equality ($p = 0.194$).

Table 2 presents balance tests for workers. Table 2 shows that we have balanced in all the characteristics at the workers level with the exception of one outcome: the belonging index. This leads to a marginally significant F-test, with a p-value of 0.103.

To ensure robustness, we will include a specification that controls for these baseline covariates in our analysis, allowing us to account for any potential biases introduced by these differences. This approach will help verify whether our results hold even after adjusting for any baseline imbalances.

2.7 Training attendance

We hold a total of three separate workshops with invitees. Attendance on the first day was 82%, with 93% of all firms sending at least one person, and attendance on the second day was 87%. As of submission, the third day of training has not been held. Based on detailed attendance data analyzed for day one, Table 3 shows that women attend more often than men, and owners attend more often than workers.

[8]Since randomization is done at the market level, we randomly assign treatment at the same level, across 5,000 replications.

Table 1: Balance table (owners)

	Means			
	Treatment	Control	Diff.	p-value (diff.)
Female employees (perm.)	1.104	1.148	-0.032	0.773
	(0.067)	(0.106)	(0.109)	
Male employees (perm.)	2.081	2.019	0.063	0.717
	(0.129)	(0.144)	(0.174)	
Total employees (perm.)	3.184	3.167	0.032	0.890
	(0.154)	(0.197)	(0.227)	
Gender attitude index	-0.043	0.000	-0.032	0.542
	(0.04)	(0.04)	(0.053)	
Gender productivity index	-0.067	0.000	-0.063	0.240
	(0.039)	(0.041)	(0.053)	
Woman-friendly index	0.042	0.000	0.048	0.370
	(0.039)	(0.041)	(0.054)	
Job satisfaction index	-0.022	0.000	-0.037	0.549
	(0.045)	(0.044)	(0.061)	
Belonging index	-0.037	0.000	-0.038	0.481
	(0.038)	(0.042)	(0.053)	
Trust index	0.007	0.000	0.011	0.852
	(0.045)	(0.049)	(0.058)	
Women hired (12 months)	0.01	0.008	0.002	0.729
	(0.005)	(0.003)	(0.006)	
Female-specific toilet	0.033	0.035	-0.003	0.764
	(0.006)	(0.007)	(0.009)	
Any leave	0.281	0.262	0.019	0.449
	(0.019)	(0.021)	(0.026)	
Male owner	0.848	0.82	0.024	0.306
	(0.018)	(0.019)	(0.024)	
Age	40.99	40.886	0.126	0.804
	(0.4)	(0.411)	(0.507)	
Married	0.904	0.895	0.009	0.571
	(0.012)	(0.014)	(0.016)	
Muslim	0.843	0.83	0.017	0.391
	(0.014)	(0.015)	(0.02)	
Primary completed	0.802	0.826	-0.025	0.194
	(0.014)	(0.015)	(0.019)	
Secondary completed	0.206	0.256	-0.048	0.059
	(0.017)	(0.02)	(0.025)	
Experience	16.882	17.619	-0.695	0.248
	(0.44)	(0.434)	(0.601)	
Square feet (log)	5.472	5.483	-0.014	0.874
	(0.051)	(0.08)	(0.089)	
Average profits	77,116	91,718	-15,023	0.520
	(5,510)	(25,025)	(23,325)	
SDP firm	0.217	0.234	-0.017	0.470
	(0.028)	(0.029)	(0.023)	
Firms (N)	917	971	1,888	
F (joint)				1.259
p-value				0.194

Note: Standard errors are in parentheses and are clustered at the market level, which is also the level of randomization. The number in brackets is the number of firms with non-missing values for each variable. Strata fixed effects are included when calculating the differences and p-values but not when calculating the pure means, meaning the stratified difference does not equal the difference in means. We calculate the joint p-value using randomization inference, as suggested by Kerwin et al. (2024).

Table 2: Balance table (workers)

	Means			
	Treatment	Control	Diff.	p-value (diff.)
Gender attitude index	-0.048	-0.006	-0.031	0.566
	(0.042)	(0.042)	(0.053)	
Gender productivity index	-0.054	0.002	-0.047	0.346
	(0.037)	(0.041)	(0.05)	
Job satisfaction index	-0.01	-0.005	-0.02	0.71
	(0.042)	(0.042)	(0.054)	
Belonging index	-0.175	-0.031	-0.148	0.004
	(0.04)	(0.041)	(0.052)	
Trust index	0.027	0.011	0.01	0.877
	(0.044)	(0.06)	(0.062)	
Male worker	0.575	0.576	-0.006	0.806
	(0.019)	(0.023)	(0.025)	
Age	30.544	30.077	0.54	0.277
	(0.461)	(0.387)	(0.496)	
Married	0.611	0.598	0.016	0.424
	(0.019)	(0.016)	(0.021)	
Muslim	0.873	0.851	0.024	0.206
	(0.014)	(0.014)	(0.019)	
Primary completed	0.718	0.748	-0.032	0.137
	(0.017)	(0.015)	(0.022)	
Secondary completed	0.134	0.147	-0.013	0.497
	(0.014)	(0.013)	(0.019)	
Experience	9.786	9.529	0.279	0.54
	(0.366)	(0.315)	(0.455)	
Monthly salary (log)	8.796	8.762	0.037	0.432
	(0.037)	(0.038)	(0.046)	
Firms (N)	800	834	1,634	
Workers (N)	1,564	1,643	3,207	
F (joint)				2.196
p-value				0.103

Note: Standard errors are in parentheses and are clustered at the market level, which is also the level of randomization. Since the firm is our unit of analysis, we reweight observations such that each firm receives equal weight, which also means the mean in the control group is not zero for the indices. Strata fixed effects are included when calculating the differences and p-values but not when calculating the pure means, meaning the stratified difference does not equal the difference in means. We calculate the joint p-value using randomization inference, as suggested by Kerwin et al. (2024).

Table 3: Training attendance, day 1

	Owners	Workers	All
Women	0.935	0.825	0.844
Men	0.833	0.790	0.810
All	0.848	0.804	0.821

Note: The values are the share of invitees who attended the first round of workshops. These do not include attendance at the second and third workshops.

2.8 Timeline

Figure 2 shows the timeline of project activities. We completed the baseline survey of firms in late 2023 and conducted the first day of training in February and March 2024. The second day of training took place in May and June. The third day is scheduled for August and September, but at the time of submitting this PAP, training has been postponed due to countrywide student protests against the government and a curfew in Bangladesh. We will conduct the first endline survey three months after the end of the workshops, projected for December 2024 and January 2025. We will conduct the second endline (follow-up) survey 12 months later, which will be 15 months after completion of the workshops, in order to assess whether any changes are temporary or more long-lasting.

Figure 2: Project Timeline

3 Conceptual framework

The relationship between gender norms and productivity is not necessarily direct, and gender-norms training may also affect productivity through alternative channels. Our conceptual framework maps the potential linkages between training, gender norms, and productivity. We anticipate three primary potential channels: changes in the perceived productivity of women, cross-gender within-firm interactions, and workforce composition (through hiring and retention).

In this framework, workers collaborate during production, choosing their effort level to maximize utility, which is the difference between the benefit from what they produce and a convex effort cost. Their production is based on their own effort, the quality of their effort, and the effort and effort quality of their collaborators. While individuals know their own effort quality, they only observe the quality of their collaborators' effort in expectation. Restrictive gender norms affect production in two ways: first, individuals perceive women's effort to be less effective than men's. Second, collaborating with someone of a different gender imposes an additional effort cost. Both frictions reduce individual effort, and in turn, productivity.

Gender norms training may influence individual effort in several ways. First, it promotes more egalitarian gender norms, increasing perceptions of women's productivity, especially among men. This increase in perceived effort quality leads to higher returns to effort, increasing individual effort and productivity. Second, it may improve mixed-gender interactions. This could occur through a gender norms channel or due to improvements in men's understanding of women's perspectives. This channel would reduce the cost of working with members of the opposite gender, increasing individual optimal effort and productivity. Finally, training may affect workforce composition by increasing the hiring and retention of women. This could be a direct impact, by influencing managers' perceptions of women's productivity or reducing their male managers' effort cost of working with women. Additionally, increases in empowerment may lead women to remain at their jobs. Increased retention

could also result indirectly, such as if workplace conditions for women improve due to better interactions with co-workers or if owners implement specific policies or infrastructure changes, like improving access to bathrooms for women. Greater hiring and retention could enhance productivity due to lower turnover rates and greater worker experience (Maranto and Rodgers, 1984), shifting the distribution of effort quality among employees.

3.1 Alternative mechanisms

The training may also influence firm productivity through alternative channels: improvements in general trust and cooperation and improvements in management leadership.

By bringing employees and firm owners together and facilitating discussions and activities, the intervention could promote general trust and cooperation independent of any change in gender norms. In this case, we would see increases in giving and cooperation among same-gender pairs in the public goods, ultimatum, prisoner's dilemma, and trust game. If both mechanisms are at work, we would see treatment increase altruism, trust, and cooperation among all pairs, with a more pronounced increase in mixed-gender pairs.

The training could improve manager leadership and worker-manager relationships, independent of gender norms. We will measure the impact of training on both a leadership index (Alan et al., 2021), which reflects workers' perception of firm owners, as well as an index of owner attitudes toward workers, provided in Section D.4.

4 Hypotheses

Here, we discuss our primary hypotheses based on the conceptual framework above.

4.1 H1 Attitudes: Workplace gender norms training will lead to more gender-egalitarian attitudes among employees and owners.

To measure gender-egalitarian attitudes, we will analyze the effects of the intervention on the gender attitudes index and the gender productivity index (as defined in Table 4). We hypothesize that the intervention will increase gender-egalitarian attitudes, meaning the indices will increase in the treatment group relative to the control group.

We will report the following outcomes, as well as a domain-specific index that takes a variance-weighted average of each outcome:

- Gender attitudes index, owners
- Gender attitudes index, workers
- Gender productivity index, owners
- Gender productivity index, workers
- Incentivized productivity perceptions, workers
- Incentivized productivity perceptions, owners

4.2 H2 Recruitment: Workplace gender norms training will lead to increased efforts to recruit women.

We will analyze whether training affects the number of women interviewed, offered jobs, and hired.

We will report the following outcomes, as well as a domain-specific index that takes a variance-weighted average of each outcome:

- Number of women interviewed, past 6 months

- Number of women offered a job, past 6 months
- Number of women hired, past 6 months
- Employer demand index, resume evaluation exercise

The final outcome will be an index of the three outcomes obtained from the incentivized resume evaluation exercise described in Section 5.2.4: total applicant rank for female applicants, share of acceptable candidates that are women, and average perceived productivity of female applicants relative to make applicants. We have included this final outcome since some firms may hire only occasionally.

4.3 H3 Workplace conditions: Provision of gender norms training to owners and employees will improve workplace conditions for women.

We measure workplace conditions using a constructed woman-friendly workplace index. The elements of this index are listed in Table 4, which reflects working conditions for women, including the availability of toilet facilities, flexible working hours, different types of leave (paid, unpaid, and maternity), and whether there are formal support groups for women.

4.4 H4 Trust and Cooperation: Provision of gender norms training to owners and employees will improve trust and cooperation between men and women.

We will report the following outcomes, as well as a domain-specific index that takes a variance-weighted average of each outcome:

- Self-reported trust index (workers)
- Lab-in-field results, mixed-gender pairs/groups

- Contributions, public goods game
- Agree to cooperate, prisoner's dilemma game
- Amount sent, trust game
- Amount returned, trust game

We hypothesize that trust will increase across all dimensions, but that mixed-gender trust will increase more than same-gender trust.

4.5 H5 Productivity: Gender norms training will improve productivity.

We hypothesize that gender norms training will enhance productivity by increasing perceptions of women's productivity, improving inter-employee interactions, and increasing the hiring and retention of female employees, as outlined in Section 3 .

We measure productivity through firm-reported profits, revenue, and investments. We will use log transformations for revenue and investment, but not profits, which can have negative values. We will also conduct a productivity experiment, described in Section 5.2.5, to generate a consistent measure of worker- and manager-specific productivity across firms.

We will report the following outcomes, as well as a domain-specific index that takes a variance-weighted average of each outcome, dividing each by the number of staff (workers plus owner):

- Profits per employee
- Log revenue per employee
- Log investment per employee
- Acceptable quality envelopes produced (mixed-gender pairs)

5 Outcomes

5.1 Primary survey outcomes

5.1.1 Gender attitudes and gender productivity indices

In the survey, we collect information on gender attitudes to create two variance-weighted indices.[9] The two indices are a gender attitudes index and a gender productivity index. We list the questions included for each index in Table 4. In all cases, we recode variables such that higher values reflect more gender egalitarian attitudes, and we calculate indices separately for owners/managers and workers, normalizing to control-group owners.

We will also measure incentivized productivity perceptions, the ratio of the number of envelopes that a respondent believes that a randomly selected woman vs. man could complete, as described in Section 5.2.5.

We present the baseline distribution of these two main indices in Figure 3, splitting the sample by the gender composition of each worker/owner pair. There appears to be sorting across the different gender composition categories. For example, female workers with a female owner tend to have more progressive gender attitudes (higher index values) on both indices, while male workers with a male owner tend to have less progressive attitudes. At least part of this is likely due to sorting across industries; certain industries are dominated by men—both owners and workers. For example, restaurants, IT support, and tailoring/garments for men have over 90% male owners, while wood furniture firms do not have a single female owner in our sample. Restaurants, wood furniture, and tailoring/garments for men are the three firm types[10] with the lowest gender indices for owners, while wood furniture is the firm type with the lowest gender index for workers.

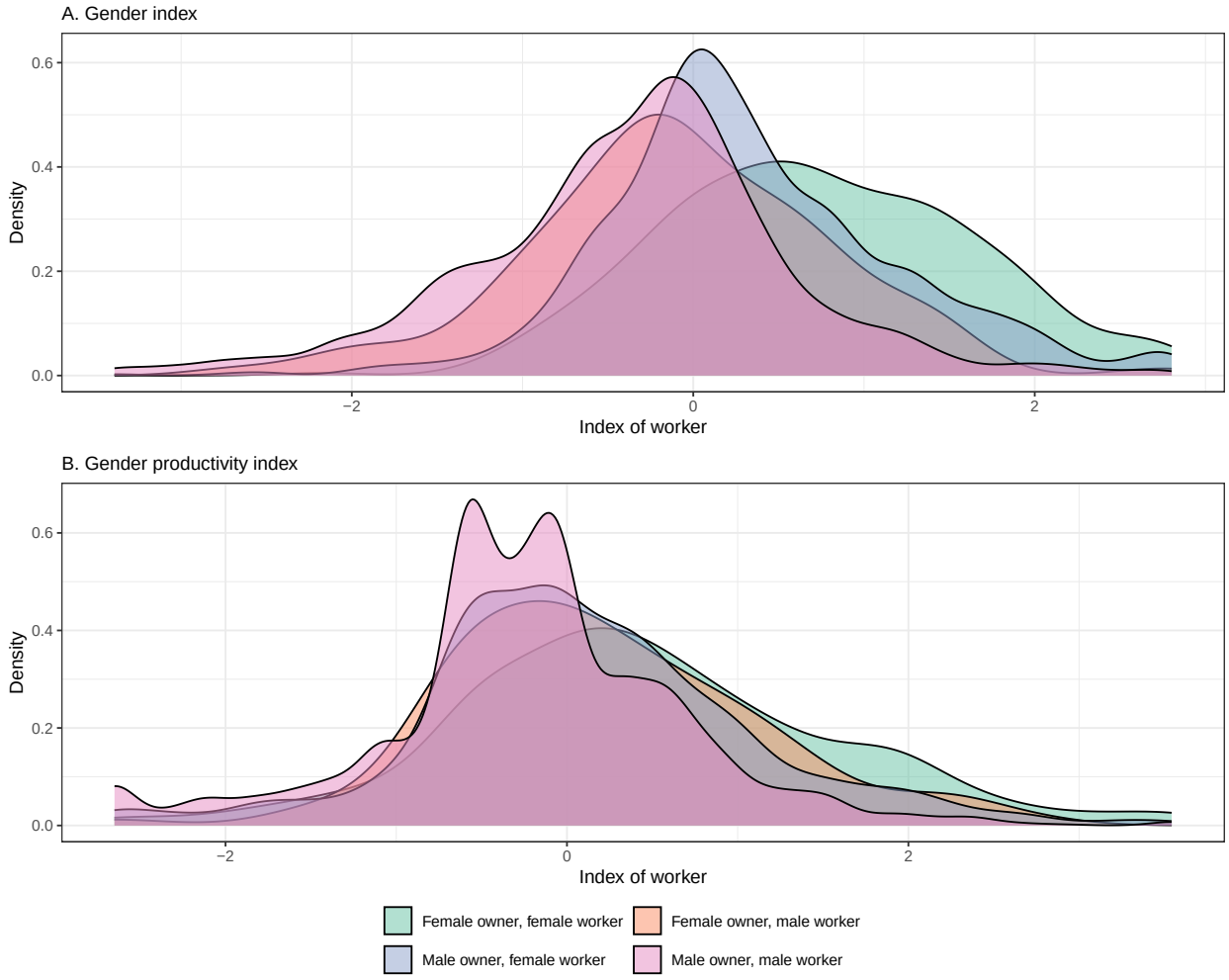
[9] We use the *swindex* function in Stata (Schwab et al., 2021) to calculate indices, which calculates standardized weighted indices based on Anderson (2008).

[10] Of firm types with more than a handful of observations.

Table 4: List of variables in gender indices

(1) Gender attitude index	(2) Gender productivity index	(3) Woman-friendly index
1. Wives should be less educated than their husbands.	1. In general, women are less productive workers than men.	1. Are there toilet facilities within the enterprise?
2. Boys should be allowed to get more opportunities and resources for education.	2. Women and men are equally likely to miss work for family responsibilities.	2. Are there toilet facilities specifically for women in the factory/workshop/enterprise?
3. A man should have the final word about decisions in his home.	3. It is best when men and women do their jobs separately in a workplace.	3. Does your firm allow flexible work hours?
4. Parents should maintain stricter control over their daughters than their sons.	4. It is more difficult to give feedback to women at work than men.	4. Does your firm have paid leave?
5. A woman's most important role is to take care of her home, feeding kids and cook for her family.	5. Men are better suited than women to work outside the house.	5. Does your firm have unpaid leave?
6. Daughters should have a similar right to inherited property as sons.		6. Does your firm provide maternity leave?
7. It would be a good idea to elect a woman as the Chairman of your village committee.		7. Does your establishment provide any formal support groups for women?
8. Girls should be allowed to study as far as they want.		
9. Men and women should get equal opportunities in all spheres of life.		

Figure 3: Distribution of main indices



5.1.2 Women-friendly workplace index

We also collect information about workplace amenities and policies, including amenities that may make the workplace more "female friendly." We include these variables in the final column of Table 4. These include questions on toilet facilities, and whether there are female-only toilet facilities; whether the firm has flexible working hours; whether the firm provides leave, including paid, unpaid, and maternity leave; and whether the firm has any formal support groups for women. We will use these variables to create an index, using the same methodology above, with higher values indicating more female-friendly policies. We call this index the "woman-friendly index." We note that some of these policies—like paid

leave—are not necessarily only advantageous for women. However, given social norms around women and household responsibilities, we nonetheless see these policies as being particularly beneficial for women.

We will also conduct in-person observational assessments by enumerators and fieldworkers, in addition to collecting employee feedback, to gauge improvements in the physical and social work environment. These assessments will include evaluating policies that support work-life balance and gender equality.

5.1.3 Trust index

We will create a trust scale based on the following four statements:

- How much do you trust your co-workers?
- How much do you trust your male co-workers?
- How much do you trust your female co-workers?
- How much do you trust your boss or manager?

The responses are provided on a five-point scale from 0 to 4 (4 = completely trust). The trust scale is the sum of the following four items with higher values representing more trusting attitudes. We will standardize this scale to the control group.

5.2 Lab-in-the-field experiments

Immediately following the endline survey, we will implement several incentivized lab-in-the-field experiments to validate self-reported answers and test for potential mechanisms. For all experiments, we will randomly select male and female owners and workers who were surveyed at baseline. Each selected owner or worker will be invited to participate in all five games, summarized in Table 5. Our pilot results indicate that these experiments elicit meaningful variation within our population of interest. Additionally, similar experimental

games have been extensively tested in developing countries similar to Bangladesh, though in different contexts. For instance, Gangadharan et al. (2016) tested the trust game and public goods game in India, while Gangadharan et al. (2022) applied both the trust and dictator games in Cambodia. Alan et al. (2021) implemented the trust game and prisoners' dilemma in Turkish schools, and Rao (2019) utilized the dictator game in Delhi schools. These studies consistently elicited significant variation, indicating the robustness of these incentivized lab-in-the-field experiments across various developing country environments.

As with the workshops, we will conduct these lab-in-the-field experiments in central locations within the study upazilas. To ensure effective field logistics and the practical implementation of our experimental games, we will randomly divide the total sample into roughly equal halves. Each participant will play two out of four games, and all will participate in the envelope-making experiment. Additionally, we will assign pairs such that we have owner-worker pairs and worker-worker pairs in our experimental games. Each owner will play with a worker, and a worker will randomly play with an owner or a worker. With the exception of the trust game and productivity experiment, participants will play each game twice: once in a same-gender pair or group, and once in a mixed-gender pair or group. The order and parings will be selected randomly.

For the asynchronous games, the identity of the players will be kept anonymous, but they will initially be informed of the gender of their fellow player via a gender-specific pseudonym. Participants will receive their earnings via mobile money two to three weeks after their participation. All participants will receive 500 BDT for participation plus the earnings from one randomly selected game as payment, compensating them for their time while minimizing the potential influence of a wealth effect in experimental games. All participants will be paired with a worker outside their own firm, and they will be compensated following the payment rules.

Table 5: Summary of lab-in-the-field experiments

	Public Goods	**Ultimatum**	**Prisoner's dilemma**	**Trust game**	**Envelope-making**
Outcome	Cooperation	Equality and fairness	Cooperation/trust	Trust	Productivity
Sample	BL men and women	BL men and women	BL men and women	BL men and women	BL men and women
Group size	4	2	2	2	2
Parings	Same- **and** mixed-gender groups	Same- **and** mixed-gender pairs	Same- **and** mixed-gender pairs	Same- **or** mixed-gender pairs	Same- **or** mixed-gender pairs
Timing	Asynchronous	Asynchronous	Asynchronous	Asynchronous	Synchronous
Decisions/game	2	2	2	∼8	Effort

5.2.1 Ultimatum game

We adopt a version of the ultimatum game with a known outside option (see e.g. Camerer and Thaler (1995) and Hennig-Schmidt et al. (2018)) to test how notions of equality and fairness influence decision-making in the context of improving gender norms. People often reject unfair offers even if doing so means they receive nothing, suggesting a preference for equitable outcomes over personal gain. The presence of a favorable outside option for the proposer amplifies their dominance concerning the proposal. This increased control subsequently augments the proposer's capacity to autonomously determine the offer's magnitude without concern for the responder's decision. We will randomly match male and female workers to assume the role of sender and receiver.

Moreover, both the players are informed about the consequences of the game in advance. The first player (proposer) is given an initial endowment of 100 BDT (approximately 1 USD) and asked to split it with the second player (responder), who knows about the initial endowment) The proposer then suggests a division of the money between the two players, and the responder can either accept or reject the proposal.

If the responder accepts the proposal, then the enumerator splits money between the players as the proposer indicated. If the responder rejects the proposal, then the proposer will receive 30 BDT and the responder will receive nothing. We will randomize the order of

the player type. Furthermore, we intend to employ the strategy method for this game by asking the responder, "what is the smallest offer you would accept" giving participants the opportunity to formulate a strategic approach or plan prior to rendering their decisions. Thus, participants can consider various factors such as fairness, the likelihood of acceptance, and potential counteroffers.

5.2.2 Prisoner's dilemma game

We will implement a one-shot prisoner's dilemma game based on Alan et al. (2021) to measure cooperation between male and female employees. In this game, there are two players who are paired randomly across firms. The enumerator will ask the two players individually whether they would like to choose blue or green among two cards. The color card chosen by both plays will determine their monetary earnings, ranging from 0–90 BDT. The enumerators explain the payoff options and offer a practice round so players understand the nature and potential consequences of the game in advance.

Table 6: Payoff Matrix: Prisoner's Dilemma game

		P2	
		Blue	Green
P1	Blue	30, 30	90, 0
	Green	0, 90	60, 60

As the payoff matrix in Table 6 shows, if both players pick the blue card, then each player will receive 30 BDT. On the other hand, if both players pick the green card, then each player will receive 60 BDT. However, if player 1 picks blue and player 2 picks green (or vice versa), then player 1 will receive 90 BDT, and player 2 will receive nothing. And if player 1 picks green and player 2 picks blue, then player 1 will receive 0 tokens (0 BDT) and player 2 will receive 9 tokens (90 BDT)

5.2.3 Trust game

In addition to self-reports of trust, we will implement a trust game based on Berg et al. (1995). We will randomly match workers in same or mixed-gender pairs. Each will take turns being a sender and a receiver.

Both players are informed about the consequences of the game in advance. The sender receives an initial endowment of 100 BDT can transfer pre-specified amounts of 0 BDT, 20 BDT, 40 BDT, 60 BDT, 80 BDT, or 100 BDT to the receiver. The experimenters triple the money that is transferred. The second player is also given an endowment of 100 BDT and can choose to transfer back any of the same 6 pre-specified options. As before, the experimenters triple the money that is transferred back to the first player from the second player. The share of money sent to the fellow player is our measure of trust in this setting. This game is designed using a strategy method such that both players choose how much to send back (reciprocate) if they assume the role of a sender.

If the receiver gets any money from the sender, they then decide how much of the money to return by selecting one of six options, reflecting the tripling: 0 BDT, 60 BDT, 120 BDT, 180 BDT, 240 BDT, or 300 BDT). We will measure reciprocity based on the average value sent back.

The trust game will complement the trust questions we ask in the survey. We hypothesize that the intervention will increase trust in the treatment group, especially among mixed-gender participants.

5.2.4 Firm hiring preferences

Because firms are small and hiring may be infrequent, our two follow-up surveys will not be able to detect changes in hiring preferences among owners who have not yet needed to recruit new workers. To understand firms' revealed preferences for hiring men versus women, we implement a hypothetical resume evaluation exercise in the spirit of Kessler

et al. (2019). Specifically, in the second follow-up survey, we will present owners with a set of eight potential (hypothetical) workers, including their gender, marital status, age (20–25 vs. 26–30), highest education (primary completed vs. secondary completed), location, and past work experience (none vs. 1–2 years). The specific characteristics will be randomized, but each owner will see the characteristics of four men and four women. We will ask them to (1) rank their candidates based on their order of preference; (2) indicate which candidates would be "acceptable" to their firm; and (3) evaluate the worker on a series of potential characteristics: whether the worker would be likely to accept the job, whether the worker would be likely to be a productive employee, whether the worker would be likely to get along with other employees, whether the worker would be likely to be employed at the firm a year from now.

The key outcome variables for each owner from this activity will be the following: (1) Total applicant rank for female applicants (on average, employers indifferent to employee gender will rank women as 4.5); (2) Share of acceptable candidates that are women; and (3) Average perceived productivity of female applicants relative to male applicants. The other outcome measures will provide additional insight into how employers think about the characteristics of male and female applicants.

5.2.5 Productivity experiment

We adapt a productivity experiment involving envelope stuffing by DellaVigna et al. (2022) to our distinct context. In rural areas of South Asia, envelope production is a common form of self-employment, as it requires minimal skills and inexpensive materials.

The task involves two roles: the assembler, who cuts and folds the paper into the correct envelope size and shape—precision here is paramount; and the sealer, who applies adhesive to the flap and ensures the envelope is flawlessly sealed, preventing any accidental openings. Each envelope that meets our quality standards will be purchased by the experimenters for 5 BDT (approximately 0.05 USD).

Prior to the paired activity, each participant will complete a 20-minute version of the experiment, in which they will act as assembler and sealer and also earn 5 BDT per accepted envelope. This will serve as a measure of individual productivity. Afterward, we will ask participants how many envelopes a randomly selected man and randomly selected woman completed during this activity. If their guess is close, they will receive a bonus equal to 10 percent of the revenue earned in the activity.

The paired manufacturing activity will take 40 minutes, preceded by a 20-minute training session. After randomly forming pairs, participants will randomly pick a chit from a basket, which will assign them to be the assembler or sealer. Before they blindly pick their roles, we will ask their actual preference of the role they prefer to undertake.

After pairing up, the participants will be allowed to discuss their roles with their partner, if they desire, to swap roles. Here, we attempt to exploit dominance in their decision-making as well as any variation in dominance of role assignment among the workers between treatment and control firms. Moreover, we will record their initial preferred roles, which will allow us to measure whether workers are more likely to compromise.

We will monitor the quality, quantity, and revenue generated by the envelopes produced within a specified time frame. Productivity will be measured by number of acceptable envelopes each team produces within the time frame.

5.3 Heterogeneity

The conceptual framework highlights that the gender is likely to be the primary source of treatment effect heterogeneity We observed in baseline that men have greater biases toward women's productivity (Figure 3) and we hypothesize a non-zero cost to working in mixed-gender pairs. We also anticipate heterogeneity in impacts based on individuals' baseline beliefs about women's productivity. Those with especially negative beliefs about womens' productivity and role in the workplace may have the largest margin for change.

Additionally, we aim to understand the extent to which managers' baseline attitudes influence employee outcomes and whether there is heterogeneity in this effect. This analysis could provide valuable insight into how the pre-existing gender norms held by firm owners and managers mediate the intervention's effectiveness across different types of firms. We further attempt to check another aspect of heterogeneity by examining how the baseline characteristics and attitudes of coworkers interact with the treatment, particularly in firms that have more than one worker. Finally, we will explore heterogeneity along other key dimensions, including the marital status of firm owners and workers, the age of both employees and managers, and their educational backgrounds.

6 Analysis

6.1 Empirical specification

We will estimate the intent-to-treat (ITT) effects of the program on our key experimental and survey outcomes using the two main empirical specifications below.

Firm/owner-level analysis

$$Y_{jc} = \beta_1 Treated_c + \gamma X_{jc} + \delta_b + \varepsilon_c, \tag{1}$$

where Y_{jc} is outcome Y for firm j in market c; $Treated$ is the market-level treatment assignment; X_{jc} is a vector of baseline controls; and δ_b is a vector of strata fixed effects.

Worker-level analysis

$$Y_{ijc} = \beta_1 Treated_c + \gamma X_{ijc} + \delta_b + \varepsilon_c, \tag{2}$$

where Y_{jc} is outcome Y for worker i at firm j in market c; $Treated$ is the market-level treatment assignment; X_{ijc} is a vector of baseline controls; and δ_b is a vector of strata

fixed effects. Because each firm may vary in the number of worker-respondents, we weight observations such that each firm is equally weighted in the sample.

All standard errors will be clustered at the market level. We will select the vector of baseline controls using the post-double-selection lasso procedure introduced by Belloni et al. (2014).[11] $\hat{\beta}_1$ is the coefficient of interest and is the ITT effect.

We will estimate outcomes at the 3-month endline and 15-month follow-up separately, allowing us to separate the short-run and medium-run impacts of the intervention.

We will explore mechanisms by comparing the results of our lab-in-the-field experiments when participants form same-gender versus mixed-gender groups. For these tests, we have individual-level decisions and will use a difference-in-differences specification:

$$Y_{ijc} = \beta_1 Treated_c + \beta_2 MixedGender_{ijc} + \beta_3 Treated_c * MixedGender_{ijc} + \gamma X_{ijc} + \delta_b + \varepsilon_c, \tag{3}$$

where $MixedGender_{ijc}$ is a binary indicator equal to one if respondent i is in a mixed-gender pair or group. The estimated $\hat{\beta}_1$ reflects the average impact of treatment across all participants, and $\hat{\beta}_3$ reflects the average differential impact of treatment between same-gender and mixed-gender groups. As before, we cluster at the market (c) level.

6.2 Inference and multiple hypotheses testing

We will utilize two methods to correct p-values. We will use Westfall-Young adjustments with 1,000 bootstrap resampling (Westfall and Young, 1993) and randomized-based inference (RI) p-values with market-level randomization permuted, and 1,000 replications as proposed by Young (2019). We will report both types of p-values. However, if the RI p-values and

[11] We will include as candidate predictors only variables for which we have baseline values.

unadjusted p-values are nearly indistinguishable, we will only report the FWER p-values alongside the unadjusted p-values in the main tables of results.

6.3 Robustness

6.3.1 Social desirability bias

Given the nature of the study and intervention, there is a possibility that respondents may provide socially desirable answers and that the intervention may influence the perceived desirability of certain answers, biasing our results. We take several steps to reduce this risk. We introduce our endline and follow-up surveys as measuring workers' and owners' workplace experiences, without any specific reference to gender. Additionally, we pair survey outcomes with incentivized activities, such as asking participants to predict how many envelopes men and women will produce in the productivity experiment and asking owners to complete a resume rating activity, both of which are described below.

Additionally, we test whether differences between treatment and control groups persist when examining the subset of individuals with a higher baseline likelihood of providing socially desirable answers, following Dhar et al. (2022). We measure this propensity using a sub-scale of the Marlowe-Crowne Social Desirability Scale at baseline (Crowne and Marlowe, 1960). This sub-scale is based on eight binary questions listed in Appendix C. We will sum all the socially desirable responses to construct a social desirability index which ranges from 0 to 8, with a higher score indicating a greater tendency to provide socially desirable answers.

We also present relationships between calculated social desirability bias and responses on indices in subsection A.1 of the appendix. All relationships except one[12] are in the expected directions, with the indices correlating positively with the social desirability bias. Reassuringly, the overall correlation is relatively low; the highest correlation across eight separate coefficients is 0.133, meaning that, at most, social desirability explains just 1.8% of the main

[12]This one exception is driven by a small sample size in the extremes of the social desirability index.

indices we use in this paper (Table A1).

6.3.2 Attrition

We do not anticipate differential attrition by treatment arm at the firm level, as we do not anticipate the intervention will affect the likelihood of firm survival. Additionally, the owners had worked at their firm for an average of 12 years, and 90% had worked there for at least 3 years, suggesting that both firm and manager turnover are likely to be low, especially prior to the endline.

However, if the intervention itself or its downstream effects impact retention, this could lead to differential attrition by treatment assignment for workers. To reduce this risk, we collect up to three phone numbers from each worker, including those from family members, in case they are no longer employed at follow-up. We also will update these phone numbers during the endline.

We will follow up with all workers and managers even if they quit their job or business. An attrition analysis will be conducted for workers and managers who cannot be tracked in the endline and follow-up surveys. If we are able to track them, we will consider them separately in our analysis and examine if there is differential turnover or quitting across treatment and control groups. These individuals who quit their job or firm are a subject of interest as an outcome of our intervention. However, it is possible that we might not observe significant effects due to (1) workers' long tenure at these firms; (2) the lack of alternatives within their own geographic or commuting region. Additionally, for women, changing jobs often entails significant challenges at both the family and workplace levels.

For attrition from the survey, we will test for differential attrition by treatment arm. In the event of differential attrition by treatment, we will test whether baseline characteristics remain balanced conditional on response. Additionally, we will adjust for differential non-response using inverse probability weighting and Lee bounds (Lee, 2009).

6.4 Analysis of textual data using Natural Language Processing (NLP)

With the rapid advancement of natural language processing (NLP) algorithms and their increasing application in economics research (Gentzkow et al., 2019; Ash and Hansen, 2023), we utilize rich textual data from our gender norms training program to gain deeper insights into the mechanisms underlying our intervention. During the training program, participants develop action plans to increase the representation of female workers and create a gender-sensitive workplace, and they complete a feedback survey at the end of each day to reflect on their learning and experiences. We utilize these action plan data and the feedback data regarding the training to conduct a comprehensive textual analysis that will assist us in understanding the mechanisms that may drive our expected results. We will use Natural Language Processing (NLP) techniques, such as those provided by the Natural Language Toolkit (NLTK), to tokenize the text data into individual words and establish thematic content. Following initial text processing, we will apply feature extraction techniques like Term Frequency (TF) and Term Frequency-Inverse Document Frequency (TF-IDF) to count the occurrences of each word or theme. TF-IDF adjusts these counts by the inverse document frequency, assigning more weight to unique terms within the corpus. Utilizing the Scikit-learn library in Python, we will vectorize our processed text data to systematically quantify the frequency of themes. This textual data analysis will not only assist in understanding the mechanisms of our intervention, but also it will provide a scientific foundation to evaluate the strengths and weaknesses of our intervention, informing future policy interventions in this context.

7 Cost-effectiveness analysis

We will conduct a comprehensive cost-effectiveness analysis based on our primary outcomes of interest and the overall costs of the intervention. This analysis will help quantify the finan-

cial efficiency of the gender norms training program by assessing the relationship between costs incurred and the observed outcomes. To ensure accuracy, we will work closely with GDRI to track detailed costing data for both the development and implementation phases of the intervention. This includes costs associated with curriculum development, workshop facilitation, participant recruitment, training materials, and any additional logistical expenses such as travel and venue rentals.

The analysis will provide insights into the cost of achieving various intermediate outputs, such as the cost per participant trained or cost per firm involved in the program. These measures will allow us to understand the financial investment required to engage participants and deliver the training effectively.

Furthermore, we will extend the analysis to measure the cost-effectiveness of achieving key intermediate and final outcomes. This includes calculating the cost per additional worker-year retained, following the framework of (Alan et al., 2021), which captures the cost of keeping a worker employed for an additional year as a result of the intervention. Additionally, we will assess the cost per standard deviation increase in worker productivity, as measured through our productivity experiments. This metric will allow us to determine the financial efficiency of boosting productivity through gender norms training.

By comparing our findings with those of similar studies, we will be able to benchmark the cost-effectiveness of this intervention against other programs aimed at improving workplace gender dynamics and productivity. This comparison will provide a broader understanding of how cost-effective gender norm interventions can be in various contexts, helping policymakers and organizations make informed decisions about investing in similar programs.

8 Concluding remarks

This study aims to assess the impact of a gender norms training program on workplace dynamics and firm productivity in SMEs in Bangladesh. By engaging both male and female workers and firm owners in an intensive training program, we seek to address key issues such as gender bias, cooperation, and mixed-gender interactions in the workplace. While the results of the study are not yet available, our conceptual framework suggests that improvements in gender norms, hiring practices, and retention could lead to enhanced workplace productivity and better working conditions for women. The study will explore alternative mechanisms, such as improved trust, cooperation, and management leadership, which may also contribute to these outcomes. The findings will provide valuable insights for policymakers and organizations seeking to promote gender equality in the workplace, offering evidence on how targeted interventions can improve both firm outcomes and gender dynamics.

9 Administrative information

Funding: This work is supported by the International Development Research Centre (IDRC), No. 109842 and the Private Enterprise Development in Low-Income Countries (Major Research Grant, MRG009_9362)

Ethical review: The Institutional Review Board (IRB) of the BRAC James P Grant School of Public Health, BRAC University, has provided ethics review for this study, IRB approval number IRB-13 February'23-003

Declaration of interest: None.

Acknowledgments: This study would not have been possible without excellent research assistance from Nuzaira Binte Neaz and Raisa Rahman. We thank Sukanto Roy, Ehsun Newaz, and the entire GDRI team for their support throughout the intervention and analysis, and we thank Sanaiyya Ansari for her expertise in developing the intervention.

References

Adhvaryu, A., Howell, P., Nyshadham, A., and Tamayo, J. A. (2023). *Rapport in Organizations: Evidence from Fast Food*. Harvard Business School.

Ahmed, T. and Sen, B. (2018). Conservative outlook, gender norms and female wellbeing: Evidence from rural Bangladesh. *World Development*, 111:41–58.

Alan, S., Baysan, C., Gumren, M., and Kubilay, E. (2021). Building social cohesion in ethnically mixed schools: An intervention on perspective taking. *Quarterly Journal of Economics*, 136(4):2147–2194.

Anderson, M. L. (2008). Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects. *Journal of the American Statistical Association*, 103(484):1481–1495.

Anderson-Butcher, D. and Conroy, D. E. (2002). Factorial and criterion validity of scores of a measure of belonging in youth development programs. *Educational and Psychological Measurement*, 62(5):857–876.

Angelucci, M. (2008). Love on the rocks: Domestic violence and alcohol abuse in rural Mexico. *BE Journal of Economic Analysis & Policy*, 8(1).

Angelucci, M., Heath, R., and Noble, E. (2023). Multifaceted programs targeting women in fragile settings: Evidence from the Democratic Republic of Congo. *Journal of Development Economics*, 164:103146.

Aregu, L., Choudhury, A., Rajaratnam, S., Locke, C., and McDougall, C. (2018). Gender Norms and Agricultural Innovation: Insights from Six Villages in Bangladesh. *Journal of Sustainable Development*, 11(4):270.

Ash, E. and Hansen, S. (2023). Text algorithms in economics. *Annual Review of Economics*, 15:659–688.

Ashraf, N., Bau, N., Low, C., and McGinn, K. (2020). Negotiating a better future: How interpersonal skills facilitate intergenerational investment. *Quarterly Journal of Economics*, 135(2):1095–1151.

Balk, D. (1997). Defying Gender Norms in Rural Bangladesh: A Social Demographic Analysis. *Population Studies*, 51(2):153–172.

Bandiera, O., Buehren, N., Burgess, R., Goldstein, M., Gulesci, S., Rasul, I., and Sulaiman, M. (2020). Women's empowerment in action: Evidence from a randomized control trial in Africa. *American Economic Journal: Applied Economics*, 12(1):210–259.

Bangladesh Bureau of Statistics (2018). Labour Force Survey 2017. Accessed: 2018-12-31.

Belloni, A., Chernozhukov, V., and Hansen, C. (2014). Inference on treatment effects after selection among high-dimensional controls. *Review of Economic Studies*, 81(2):608–650.

Berg, J., Dickhaut, J., and McCabe, K. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior*, 10(1):122–142.

Blattman, C. and Dercon, S. (2018). The impacts of industrial and entrepreneurial work on income and health: Experimental evidence from Ethiopia. *American Economic Journal: Applied Economics*, 10(3):1–38.

Blunch, N. H. and Das, M. B. (2015). Changing norms about gender inequality in education: Evidence from Bangladesh. *Demographic Research*, 32:183–218.

Boudreau, L. E., Chassang, S., Gonzalez-Torres, A., and Heath, R. M. (2023). Monitoring harassment in organizations. Technical report, National Bureau of Economic Research.

Buchmann, N., Field, E., Glennerster, R., Nazneen, S., and Wang, X. Y. (2023a). A Signal to End Child Marriage: Theory and Experimental Evidence from Bangladesh. *American Economic Review*, 113(10):2645–2688.

Buchmann, N., Meyer, C., and Sullivan, C. D. (2023b). Paternalistic discrimination. Technical report, Working Paper.

Bursztyn, L., González, A. L., and Yanagizawa-Drott, D. (2020). Misperceived social norms: Women working outside the home in Saudi Arabia. *American Economic Review*, 110(10):2997–3029.

Camerer, C. and Thaler, R. H. (1995). Anomalies: Ultimatums, dictators and manners. *Journal of Economic Perspectives*, 9(2):209–219.

CARE (2014). *CARE Gender, Equity, and Diversity Training Materials: Module 4, Gender Training*. CARE.

Cattaneo, M. D., Jansson, M., and Newey, W. K. (2018). Inference in linear regression models with many covariates and heteroscedasticity. *Journal of the American Statistical Association*, 113(523):1350–1361.

Chandramohan, S., Salinger, A. P., Wendt, A. S., Waid, J. L., Kalam, M. A., Delea, M. G., Comeau, D. L., Sobhan, S., Gabrysch, S., and Sinharoy, S. (2023). Diagnosing norms and norm change in rural Bangladesh: an exploration of gendered social norms and women's empowerment. *BMC Public Health*, 23(1).

Chang, W., Diaz-Martin, L., Gopalan, A., Guarnieri, E., Jayachandran, S., and Walsh, C. (2020). What works to enhance women's agency: Cross-cutting lessons from experimental and quasi-experimental studies. *J-PAL Working Paper*, 87.

Crowne, D. P. and Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology*, 24(4):349.

De Mel, S., McKenzie, D., and Woodruff, C. (2013). The demand for, and consequences of, formalization among informal firms in Sri Lanka. *American Economic Journal: Applied Economics*, 5(2):122–150.

DellaVigna, S., List, J. A., Malmendier, U., and Rao, G. (2022). Estimating social preferences and gift exchange at work. *American Economic Review*, 112(3):1038–1074.

Dhar, D., Jain, T., and Jayachandran, S. (2022). Reshaping adolescents' gender attitudes: Evidence from a school-based experiment in India. *American Economic Review*, 112(3):899–927.

Doyle, K., Levtov, R. G., Barker, G., Bastian, G. G., Bingenheimer, J. B., Kazimbaya, S., Nzabonimpa, A., Pulerwitz, J., Sayinzoga, F., Sharma, V., et al. (2018). Gender-transformative Bandebereho couples' intervention to promote male engagement in reproductive and maternal health and violence prevention in Rwanda: Findings from a randomized controlled trial. *PloS one*, 13(4):e0192756.

Duflo, E. (2003). Grandmothers and granddaughters: Old-age pensions and intrahousehold allocation in South Africa. *World Bank Economic Review*, 17(1):1–25.

Field, E., Pande, R., Rigol, N., Schaner, S., and Troyer Moore, C. (2021). On her own account: How strengthening women's financial control impacts labor supply and gender norms. *American Economic Review*, 111(7):2342–2375.

Gangadharan, L., Islam, A., Ouch, C., and Wang, L. C. (2022). The long-term effects of genocide on antisocial preferences. *World Development*, 160:106068.

Gangadharan, L., Jain, T., Maitra, P., and Vecci, J. (2016). Social identity and governance: The behavioral response to female leaders. *European Economic Review*, 90:302–325.

Gentzkow, M., Kelly, B., and Taddy, M. (2019). Text as data. *Journal of Economic Literature*, 57(3):535–574.

Hancock, J. I., Allen, D. G., Bosco, F. A., McDaniel, K. R., and Pierce, C. A. (2013). Meta-analytic review of employee turnover as a predictor of firm performance. *Journal of Management*, 39(3):573–603.

Haque, M., Chowdhury, S., Islam, A., Anjum, I., and Miah, A. S. (2022). *Exploring Attitude towards Gender Norms among the Youth Population in Bangladesh*. BRAC Advocacy for Social Change, Dhaka.

Heath, R. and Jayachandran, S. (2017). The causes and consequences of increased female education and labor force participation in developing countries. In *The Oxford Handbook of Women and the Economy*, pages 345–367. Oxford University Press.

Hennig-Schmidt, H., Irlenbusch, B., Rilke, R. M., and Walkowitz, G. (2018). Asymmetric outside options in ultimatum bargaining: a systematic analysis. *International Journal of Game Theory*, 47:301–329.

Hersch, J. (1991). Education match and job match. *The Review of Economics and Statistics*, pages 140–144.

Hossin, M. M., Azam, M. S., and Hossain, M. S. (2023). Understanding the Concept of SMEs in Driving Economic Growth and Development in Bangladesh. *International Journal of Finance, Economics and Business*, 2(3):195–204.

Hsieh, C.-T., Hurst, E., Jones, C. I., and Klenow, P. J. (2019). The allocation of talent and U.S. economic growth. *Econometrica*, 87(5):1439–1474.

Ilmakunnas, P., Maliranta, M., and Vainiomäki, J. (2005). Worker turnover and productivity growth. *Applied Economics Letters*, 12(7):395–398.

Jayachandran, S. (2021). Social norms as a barrier to women's employment in developing countries. *IMF Economic Review*, 69(3):576–595.

Jensen, R. (2012). Do labor market opportunities affect young women's work and family decisions? Experimental evidence from India. *Quarterly Journal of Economics*, 127(2):753–792.

Kerwin, J., Rostom, N., and Sterck, O. (2024). Striking the Right Balance: Why Standard

Balance Tests Over-Reject the Null, and How to Fix It. Technical report, Institute of Labor Economics (IZA).

Kessler, J. B., Low, C., and Sullivan, C. D. (2019). Incentivized resume rating: Eliciting employer preferences without deception. *American Economic Review*, 109(11):3713–3744.

Kessler, R. C., Barker, P. R., Colpe, L. J., Epstein, J. F., Gfroerer, J. C., Hiripi, E., Howes, M. J., Normand, S.-L. T., Manderscheid, R. W., Walters, E. E., et al. (2003). Screening for serious mental illness in the general population. *Archives of General Psychiatry*, 60(2):184–189.

Kessler-Harris, A. (2003). *Out to work: A history of wage-earning women in the United States*. Oxford University Press.

Lee, D. S. (2009). Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects. *Review of Economic Studies*, 76(3):1071–1102.

Li, Q., Lourie, B., Nekrasov, A., and Shevlin, T. (2022). Employee turnover and firm performance: Large-sample archival evidence. *Management Science*, 68(8):5667–5683.

Macchiavello, R., Menzel, A., Rabbani, A., and Woodruff, C. (2020). Challenges of change: An experiment promoting women to managerial roles in the bangladeshi garment sector. Working Paper 27606, National Bureau of Economic Research.

Maranto, C. L. and Rodgers, R. C. (1984). Does work experience increase productivity? A test of the on-the-job training hypothesis. *Journal of Human Resources*, 19(3):341–357.

Matsuura, A. and Teng, C. (2020). *Understanding the Gender Composition and Experience of Ready-Made Garment (RMG) Workers in Bangladesh*. International Labour Organization. Accessed: 2020-10-15.

Merfeld, J. D. (2023). Sectoral wage gaps and gender in rural India. *American Journal of Agricultural Economics*, 105(2):434–452.

OANDA (2024). Historical Exchange Rates. Technical report.

Qian, N. (2008). Missing women and the price of tea in China: The effect of sex-specific earnings on sex imbalance. *Quarterly Journal of Economics*, 123(3):1251–1285.

Rao, G. (2019). Familiarity does not breed contempt: Generosity, discrimination, and diversity in Delhi schools. *American Economic Review*, 109(3):774–809.

Rosenzweig, M. R. and Schultz, T. P. (1982). Market opportunities, genetic endowments, and intrafamily resource distribution: Child survival in rural India. *American Economic Review*, 72(4):803–815.

Schwab, B., Janzen, S., Magnan, N. P., and Thompson, W. M. (2021). SWINDEX: Stata module to create a standardized weighted index of multiple indicator variables.

Seguino, S. (2007). Plus ça change? Evidence on global trends in gender norms and stereotypes. *Feminist Economics*, 13(2):1–28.

Sen, A. (1990). Development as capability expansion. *The Community Development Reader*, 41:58.

UNDP (2023). *2023 Gender Social Norms Index (GSNI): Breaking down gender biases: Shifting social norms towards gender equality*. New York.

Westfall, P. H. and Young, S. S. (1993). *Resampling-based multiple testing: Examples and methods for p-value adjustment*, volume 279. John Wiley & Sons.

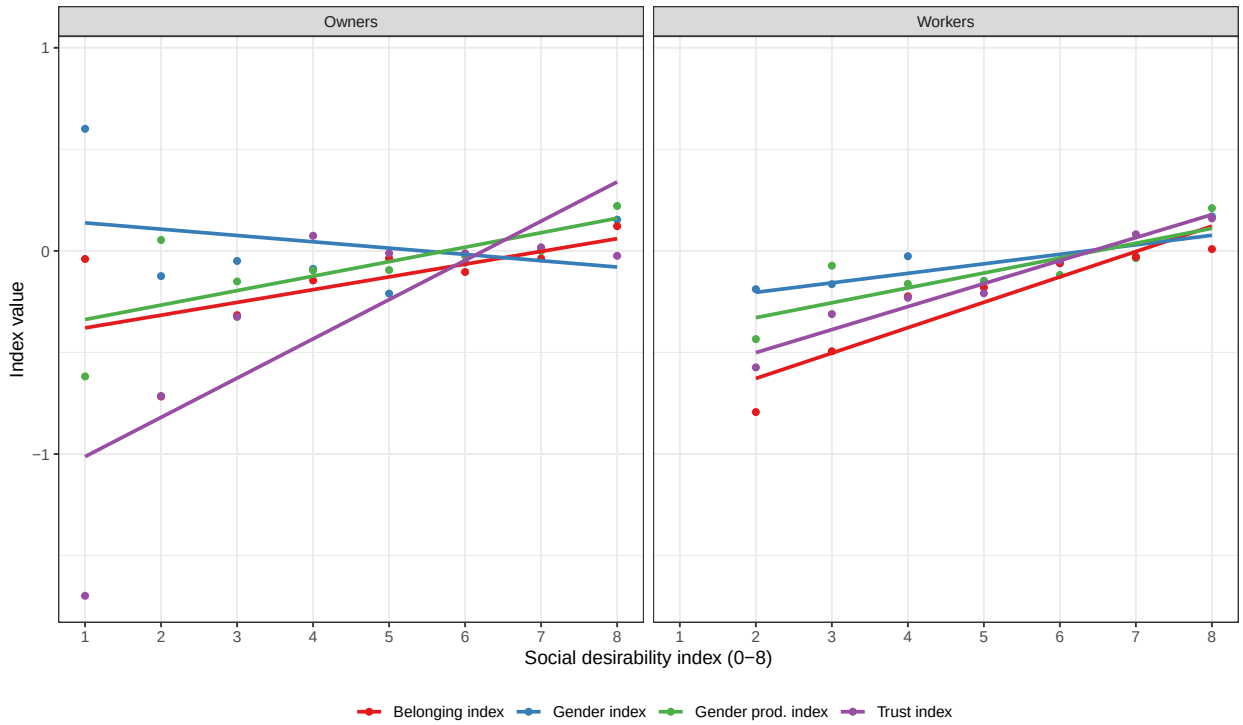
Young, A. (2019). Channeling fisher: Randomization tests and the statistical insignificance of seemingly significant experimental results. *Quarterly Journal of Economics*, 134(2):557–598.

A Additional results

A.1 Social desirability bias

We calculate social desirability bias following Crowne and Marlowe (1960). We use eight questions and sum up socially desirable answers. We then take the mean across five separate indices and plot these together in Figure A1. At the mean, there does appear to be a correlation between the social desirability index and responses on other indices; however, we note that much of this is driven by relatively few observations at the bottom and top of the social desirability scale.[A1]

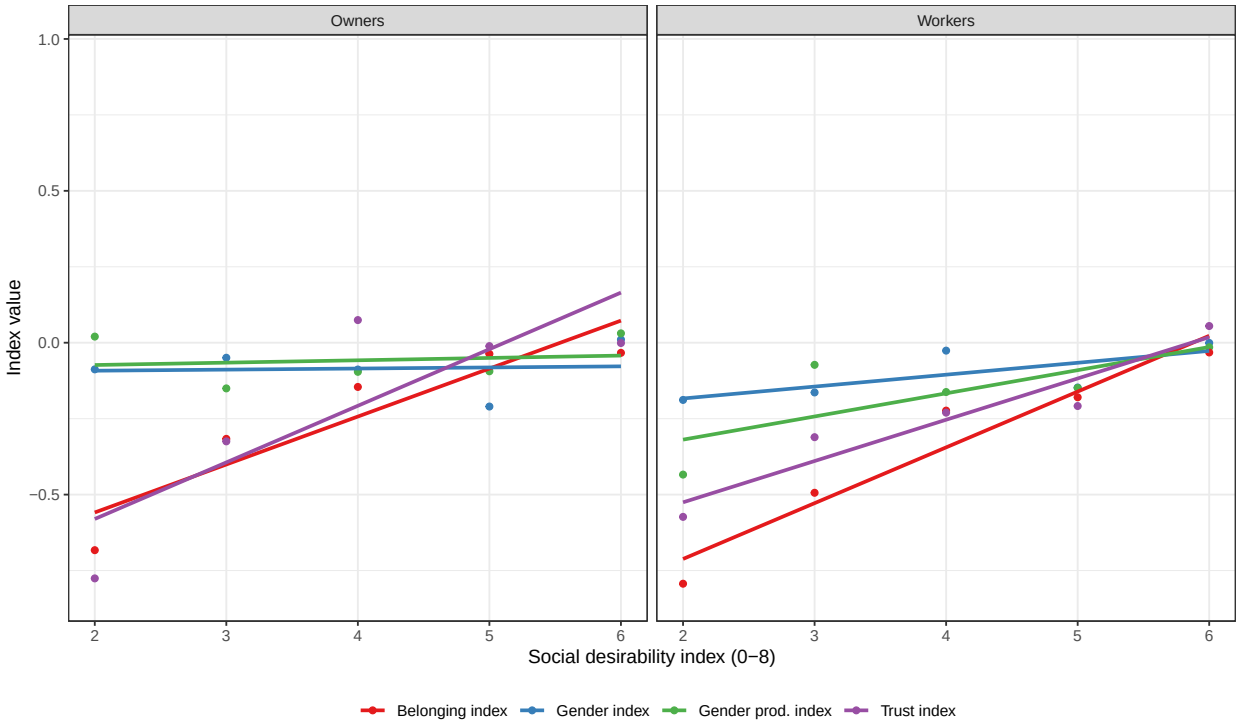
Figure A1: Mean indices by social desirability response



If we winsorize at two and six for both owners and workers, the resulting distribution shows relatively less correlation. We show this in Figure A2.

[A1] For owners, around 2.6% of responses are below two and 1.5% of responses are above six. For workers, around 1.4% have a social desirability bias of less than two and 2.5% have a social desirability bias of more than six.

Figure A2: Mean indices by social desirability - Winsorized desirability



As further evidence, we report simple pairwise correlations, at the individual level, between social desirability bias and different indices we use in the paper. We report these correlations in Table A1. While all the correlations are in the expected direction, overall correlations are quite small; the highest correlation is just 0.126, meaning that social desirability explains, at most, 1.6% of responses to the index questions.

Table A1: Correlation between social desirability index and indices

	Owners	Workers
Gender index	0.080	0.073
Gender productivity index	0.077	0.101
Belonging index	0.090	0.090
Trust index	0.034	0.133

Note: The social desirability index ranges from 0 to 8. Correlations are simple pairwise correlations.

A.2 Relationship between worker and owner characteristics

In this section, we present results of regressions of different worker-level indices on the index of the firm's owner. We include strata fixed effects to account for the randomization process. We present these results in Table A2.

Table A2: Worker and owner characteristics

	Gender index	Prod. index	Disc. index	Job sat. index	Belonging index	Trust index
Owner index value	0.2874***	0.3084***	0.1542***	0.4324***	0.4099***	0.2846***
	(0.0221)	(0.0236)	(0.0234)	(0.0205)	(0.0215)	(0.0205)
Observations	3,207	3,207	3,204	3,207	3,207	3,157

Note: Standard errors are in parentheses and are clustered at the market level, which is the level of randomization. Strata fixed effects are included in all regressions. The outcome is the value for each worker and the predictor is the value for the owner.

A.3 Power calculations for other primary outcomes

For the remaining primary outcome variables, we calculate the minimum detectable effect size with a 5% significance level at 80% power. As before, there are 403 markets in the control group and 404 markets in the treatment group, such that there is an average of 2.41 firms per market in the control group and 2.27 firms per market in the treatment group. Moreover, we have around 4.1 workers per market in the control group and 3.9 workers per market in the treatment group.

1. Perceived productivity index

   We calculate the intra-cluster correlation coefficient (ICC) to be 0.07605 (for firm owners) and 0.16511 (for workers) based on the baseline perceived gender productivity index using the loneway command in Stata. On this basis, we have an MDE of 0.135 standard deviations (s.d.) at 80% power for owners and MDE of 0.121 standard deviations (s.d.) at 80% power for workers in our sample.

2. Women-friendly workplace index

   We calculate the intra-cluster correlation coefficient (ICC) to be 0.14109 (for firm owners) and 0.22578 (for workers) based on the baseline women-friendly workplace index using the loneway command in Stata. On this basis, we have an MDE of 0.141 standard deviations (s.d.) at 80% power for owners and MDE of 0.127 standard deviations (s.d.) at 80% power for workers in our sample.

3. Job satisfaction index

   We calculate the intra-cluster correlation coefficient (ICC) to be 0.27592 (for firm owners) and 0.29992 (for workers) based on the baseline job satisfaction index using the loneway command in Stata. On this basis, we have an MDE of 0.151 standard deviations (s.d.) at 80% power for owners and MDE of 0.136 standard deviations (s.d.) at 80% power for workers in our sample.

4. Belonging index

   We calculate the intra-cluster correlation coefficient (ICC) to be 0.20051 (for firm owners) and 0.28018 (for workers) based on the baseline belonging index using the loneway command in Stata. On this basis, we have an MDE of 0.145 standard deviations (s.d.) at 80% power for owners and MDE of 0.134 standard deviations (s.d.) at 80% power for workers in our sample.

5. Trust index

   We calculate the intra-cluster correlation coefficient (ICC) to be 0.18867 (for firm owners) and 0.16706 (for workers) based on the baseline trust index using the loneway command in Stata. On this basis, we have an MDE of 0.144 standard deviations (s.d.) at 80% power for owners and MDE of 0.121 standard deviations (s.d.) at 80% power for workers in our sample.

B Additional sampling details

B.1 Firm listing details

For the firm-listing exercise, enumerators were provided with a list of markets within each included sub-district (upazila). A market is an informal designation reflecting an area with a dense concentration of firms. In an urban or peri-urban area, it could be a set of small shops comprising 2–3 floors of a building. In rural areas, it could span an area of 100–200 meters. They then proceeded with the following procedure to identify potential firms for the sample:

Enumerators started with the largest local market within each sub-district. They were instructed to canvas the entire market and list as many potentially eligible firms, following the indicated criteria:

- Firms needed to work in a set of specific shop types that were pre-determined to be popular and unlikely to employ exclusively women: these included mobile phone servicing, tailoring and garment making, block or batik work, fridge or AC repair, embroidery (handicrafts and machine), wooden furniture work, basic electrical work, graphics design, hardware technician services, IT support technician services, aluminum fabrication, medium grocery stores, retail or wholesale stores, and small hotels (which in Bangladesh often refer to food shops, restaurants, or canteens)
- Only shops with permanent structures should be included in the survey. Temporary shops with no permanent establishment should be excluded.
- The shop must have at least one employee working under the owner. Initially, we required that the shops have at least one female employee, but we relaxed this criteria early on because it was overly restrictive.

They then proceeded to the remaining markets, until the list was complete. While all firms had employees during the listing exercise and, as such, ended up in our final sample, some

of these firms have since cut ties with employees, leaving only the owner.

Among the listed firms, we excluded some firms based on the following criteria:

1. Firms in low-density upazilas/districts: Some geographic areas had too few firms to be logistically feasible to include
2. Missing total number of employees
3. The only employees were female household members of the owner
4. The manager stated that they were not interested in training or that they did not anticipate the firm would be open one year from now

B.2 Worker selection details

For this study, a regular employee is defined as someone who (1) works for pay for the company (2) on a continuing basis. However:

- They may or may not have a contract.
- The may or may not work full time, and the hours can change from week to week.
- They may be newly hired, but the expectation is that they will continue to work for the firm in the coming months.
- Anyone hired just for the month with no expectation of renewal or an employed hired on a day-to-day basis is considered a temporary – not regular – employee.

In the case there were more eligible employees than interview spots, we prioritized:

- Employees who have worked at the firm for six or more months.
- Employees aged 18-35.

C Social desirability bias questions

The socially desirable answer is given within brackets at the end of each following question:

1. There have been occasions when I took advantage of someone (Disagree).
2. I sometimes feel resentful when I don't get my way (Disagree).
3. I have deliberately said something that hurt someone's feelings (Disagree).
4. I'm always willing to admit it when I make a mistake (Agree).
5. I am always courteous, even to people who are disagreeable (Agree).
6. There have been times when I felt like rebelling against people in positions of authority even though I knew they were right (Disagree).
7. There have been times when I was quite jealous of the good fortune of others (Disagree).
8. I am sometimes irritated by people who ask favors of me (Disagree).

D Mechanism questions

D.1 Perspective-taking

To assess changes in participants' ability to understand others' viewpoints, we will measure the impact of the intervention on a perspective-taking index built using responses to the following questions, drawn from Alan et al. (2021). Respondents are asked whether each statement applies to them never, rarely, sometimes, often, or always.

1. I try to understand how others feel.
2. My friends talk to me about their problems.
3. I can put myself in someone else's shoes and understand how they feel.
4. I can tell if a friend of mine is upset.

The second component of our perspective-taking measure presents two hypothetical workplace scenarios to our respondents, followed by questions to assess their perspective on the situation.

1. I try to understand how others feel.
2. My friends talk to me about their problems.
3. I can put myself in someone else's shoes and understand how they feel.
4. I can tell if a friend of mine is upset.

Scenario 01: Fatima has been working at a small factory for three years. She works hard and gets good feedback from her boss. But when a promotion comes up, a male coworker who has been there for less time gets the job. When Fatima asks why, her boss says he thinks the male worker is more of a "leader."

1. Do you believe Fatima faced unfair treatment in her workplace? (Yes /No).
2. Fatima's experience reflects challenges commonly faced by women in the workplace. (Strongly Disagree / Disagree / Neutral / Agree / Strongly Agree).
3. The reasoning provided by the boss for promoting the male worker seems justified. (Strongly Disagree / Disagree / Neutral / Agree / Strongly Agree).
4. What is the most likely reason Fatima did not receive the promotion? (Lack of leadership skills / Gender bias / Seniority / Other / Don't know)

Scenario 02: Mr. Rahman owns a small fabric-making shop and needs to hire a new worker. He knows many skilled women who would be good at the job, but he worries that if he hires a woman, she might not stay long because of family obligations or pressure from her family not to work. Mr. Rahman is unsure whether he should hire a woman or look for a male worker instead.

1. Do you think Mr. Rahman's hesitation to hire a woman is influenced by gender stereo-

types? (Yes / No).

2. Mr. Rahman's concerns reflect common challenges that women face in the labor market. (Strongly Disagree / Disagree / Neutral / Agree / Strongly Agree).

3. I can understand why Mr. Rahman might be hesitant to hire a woman due to potential family pressures. (Strongly Disagree / Disagree / Neutral / Agree / Strongly Agree).

4. If you were in Mr. Rahman's position, how likely would you be to hire a woman despite your concerns? (Very unlikely / Unlikely / Neutral / Likely / Very likely).

D.2 Empathy

Similarly, we build an empathy index based on whether respondents say that each of the following statements applies to them never, rarely, sometimes, often, or always.

1. When I see someone being treated unfairly, I feel sorry for them.

2. When I see someone being taken advantage of, I become protective towards them.

3. I often have tender feelings for people less fortunate than me

4. I feel sorry for other people when they are having problems.

5. I would describe myself as a pretty soft-hearted person.

D.3 Employee interactions

To measure employee interactions and segregation in the workplace, we include two, one targeting workers and the other targeting owners:

D.3.1 Worker module

Think about your work over the past week:

- In a typical day, how many male employees do you interact with directly while at work?
- In a typical day, how many female employees do you interact with directly while at work?
- In a typical day, how many female customers or clients do you interact with directly for your work?
- In a typical day, how many male customers or clients do you interact with directly for your work?
- In a typical day, how many hours did you spend working on tasks alone
- In a typical day, how many hours did you spend working on tasks with other people?
  - Of that time, how many hours did you spend working on tasks only with people of the same gender?
  - Of that time, how many hours did you spend working on tasks in mixed-sex pairs or groups?

D.3.2 Owner module

- How often do male and female employees work together on the same task? All the time / most of the time /some of the time / rarely / never
- Do men and women have the same tasks and responsibilities, or do they differ? Same / Different
- Do men ever do the tasks and responsibilities typically assigned to women? All the time / most of the time /some of the time / rarely / never
- Do women ever do the tasks and responsibilities typically assigned to men? All the time / most of the time /some of the time / rarely / never

D.4 Manager-worker relationships

To evaluate improvements in communication, respect, and collaboration between managers and workers, we construct two indices of worker perception of manager professionalism and manager perceptions of worker professionalism:

The leadership index draws from Alan et al. (2021) and is adapted for our context:

Table D1: Leadership index

The following statements are related to your manager. Please use the following scale to state your opinion. (Never-Rarely-Sometimes-Often-Always)	
m1a	Our manager is a good listener.
m1b	Our manager has favorites and they are given favorable treatment
m1c	Our manager is modest and accepts her mistakes
m1d	I completely trust our manager's professionalism.
m1e	I receive regular and motivating feedback from our manager
m1f	Our manager takes credit for successes but blames mistakes on others
m1g	Our manager understands the challenges that workers face

Statements m1a–m1f taken from Alan et al. (2021), with m1f wording modestly adjusted

The worker relationship index captures owners' attitudes toward their employees:

Table D2: Worker relationship index

The following statements are related to the employees that work at your firm. Please use the following scale to state your opinion. (Never-Rarely-Sometimes-Often-Always
Employees at my firm are motivated and hard-working.
Workers respond well to feedback and try their best.
My employees take credit for their own successes, but they blame mistakes on others
My employees understand the challenges of owning and operating a firm
My employees won't do things well unless they are carefully supervised
Employees at our firm make reasonable requests for support or accommodation

Table D3: Social connection scale

G9a	How many different people at your workplace did you interact with at all in the past 7 days? This could include people with whom you said hello, had a conversation, worked together on a task, or had a meal or snack with.
Among those people that you interacted with in your workplace...	
G9c	About how many people could you ask for advice about your current job?
G9d	About how many people asked you for advice about their current job in the last three months?
G9e	About how many people could you ask for advice about salary or promotions?
G9f	About how many people asked you for advice about salary or promotions in the last three months?
G9g	With how many people could you discuss your personal and family matters?
G9h	About how many have you discussed your personal and family matters with in the past 3 months?
G9i	About how many have discussed their own personal and family matters with you in the past 3 months?

D.5 Social connections and networks

To identify changes in the strength and breadth of professional relationships and support networks within the firm, we will use the questions in Table D3 to construct our social connection scale, dividing each response by the total number of employees and owners at the firm.

D.6 Sense of belonging, mental health, and job satisfaction

We measure job satisfaction and participants' sense of belonging using the questions listed in columns 1 and 2 of Table D4. We ask respondents the extent to which they agree with each of 6 statements about their satisfaction at their workplace, using a 5-point Likert scale. The sense of belonging index is based on four statements developed by Anderson-Butcher and Conroy (2002), with respondents selecting from a four-item Likert scale. In both cases, we recode negative statements so that a higher number implies greater satisfaction or belonging.

We measure mental health using the Kessler Psychological Distress Scale (K10) (Kessler et al., 2003). For each question, respondents have five options: (0) none of the time, (1) a little of the time, (2) some of the time, (3) most of the time, and (4) all of the time. Following

Table D4: List of variables in satisfaction/mental health indices

(1) **Satisfaction index**	(2) **Belonging index**	(3) **Kessler Psychological Distress Scale (K10)** "In the past 4 weeks, about how often did you feel..."
1. I regret that I decided to become a [position].	1. I feel comfortable at work.	1. Tired out for no good reason?
2. I enjoy working at this firm.	2. I feel like I am an important member of the company.	2. Nervous?
3. I wonder whether it would have been better to choose another profession.	3. I don't have many friends at work.	3. So nervous that nothing could calm you down?
4. I would recommend my firm as a good place to work.	4. I am accepted at the company.	4. Hopeless?
5. I feel respected by my [co-workers/employees] at this firm.		5. Restless or fidgety?
6. All in all, I am satisfied with my job.		6. So restless you could not sit still?
		7. Depressed?
		8. That everything is an effort?
		9. So sad that nothing could cheer you up?
		10. Worthless?

Note: For the satisfaction and belonging indices, responses are on a Likert scale, from strongly agree to strongly disagree, with items coded so larger numbers indicate greater satisfaction and belonging. For the mental health index, respondents have five options: (1) none of the time, (2) a little of the time, (3) some of the time, (4) most of the time, and (5) all of the time.

the literature, we create the index by summing responses, resulting in an index that can take on values between 0 and 40, with higher values indicating higher levels of distress (i.e. poorer mental health).

E Gender-norms training outline

Participants

Female and male workers and managers (mobile phone servicing, tailoring and garment making, block or batik, refrigerator or AC repair, embroidery (handcraft and machine), wooden furniture work, electrical work/shop (basic), graphics design, hardware technician, IT support technician, aluminium manufacturer, medium grocery/retail/wholesale shops, clothing store, hotel, sweet shop, servicing (bike motors, automobile), medicine (factory, wholesale, retail), small factory, and small shop)

Objectives of the training

1. To enhance the importance of creating a gender-inclusive workplace and environment for owners and workers.
2. To inspire the owners to facilitate a gender-sensitive workplace, take initiative in ensuring protections for women, communicate with them, and support women to increase job retention and career advancement.
3. To encourage the SMEs to internalize the power within in challenging the gender norms at home and workplace, especially regarding washroom/toilet facilities, working time, use of non-abusive language, non-harassment environment, breastfeeding space, etc.
4. To find a way out or have a concrete action plan to increase female representation in the workplace and promote a gender-inclusive environment.
5. To create a space for check-ins at three-month intervals to discuss successes, challenges, and next steps.

Table E1: Day 1 Training Schedule

Day 1 (9:00 AM to 3:30 PM)				
Module	**Topic**	**Sub-topics and activities**	**Methodology**	**Duration**
1	Inauguration and Check in	- Welcome speech and formalities - Ice-breaking activities - Dos and don'ts during training	- Sharing and caring - Helium stick	45 minutes
2	Understanding rights	- Introduction to diversity and inclusions - Gender inequality, equality, equity and justice	- Mapping exercise - Picture puzzle - Roleplay - Slide show	1 hour
Tea break				30 minutes
3	Gender in everyday life	- Attitudes, behaviour and language used against women, men, and other gender identities by the society - Gender division of labour - Gender and sex	- Group work - Lecture discussion - Card game	1 hour and 45 minutes
Lunch break				1 hour
4	Gender in everyday life contd.	- Gender division of labour - Gender and sex	- Role play - Debate - Poster presentation on gender and sex - Quiz	1 hour
5	Let's see what we can remember	- Learning check	- Quiz	15 minutes
Check out				15 minutes

Table E2: Day 2 Training Schedule

Day 2 (9:00 AM to 4:00 PM)				
Module	**Topic**	**Sub-topics and activities**	**Methodology**	**Duration**
1	Welcome and self-reflection on the progress of the commitments made after the 1st day training	- Mood setting, welcome and introduction of participants - Emotional well being - Reflections of commitments	- Emotional well being exercise - brainstorming - group work - Story wall of success	1 hour and 20 minutes
Tea break				15 minutes
2	Challenges standing between us and success	- Challenges in the path of success - Who should meet the challenges	- Challenge hunt - Brainstorming	1 hour and 15 minutes
3	Redefine our strength	- Role of women and men in redesigning society in a positive way	- Video on Monira and Saleha - Group work on the gender role (Redefined by the participants)	1 hour and 10 minutes
Lunch break				1 hour
4	Measuring strengths and define new gender-sensitive norms	- New thoughts on gender norms - How to shape new norms	- Ship of new norms /Lamp of new norms	1 hour and 30 minutes
5	Closing	- Personal and professional commitments after the training - wrap-up of the day	- Sharing	30 minutes

Table E3: Day 3 Training Schedule

Day 3 (10:00 AM to 12:00 PM)				
Module	**Topic**	**Sub-topics and activities**	**Methodology**	**Duration**
1	Meet and greet	- Meet and greet with management and workers - Meet and greet with male/female workers - Tokens of appreciation	-Observation of the trainer	30 minutes
2	Observation of changes in KAS/3 H: Knowledge (Head) Attitude (Heart) and Skill (Hand)	- Introductory conversations /Motivational speech - Finding a way out - Feelings after finding a way out	- Observation - Case Study	1 hour and 15 minutes
3	Discussion	- Development of plans/commitment - Reflection of three days	- Discussion	15 minutes