ECONOMIC DEVELOPMENT & INSTITUTIONS

EDI WORKING PAPER SERIES

LEARNING WHAT TO LOOK FOR: HARD MEASURES OF SOFT SKILLS IN PROMOTION

Hannah Uckat
University of Oxford
Christopher Woodruff
University of Oxford

November 2019

Abstract

We report the results of a field experiment designed to promote women to supervisory positions in Bangladesh's garment factories, with which participating factories have little prior experience. We show that formal diagnostic tests lead factories to choose candidates that are more likely to be promoted and who, according to their subordinates, perform better after promotion. Diagnostics measuring 'soft skills' are particularly relevant for managers. Randomised timing of supervisory training for the selected candidates shows that providing the training earlier leads to higher compliance with the standard factory training protocol, but has only marginal effects on the performance of the trainees following promotion.

About Economic Development & Institutions
Institutions matter for growth and inclusive development. But despite increasing awareness of the importance of institutions on economic outcomes, there is little evidence on how positive institutional change can be achieved. The Economic Development and Institutions – EDI – research programme aims to fill this knowledge gap by working with some of the finest economic thinkers and social scientists across the globe.

The programme was launched in 2015 and will run until 2021. It is made up of four parallel research activities: path-finding papers, institutional diagnostic, coordinated randomised control trials, and case studies. The programme is funded with UK aid from the UK government. For more information see http://edi.opml.co.uk.

UKaid from the British people

1. Introduction

Supervisors matter. Lazear et al. (2015) find that replacing a supervisor in the 10^{th} percentile with one in the 90^{th} percentile increases output of a work team by more than 10 percent. However, identifying the best candidates for supervisory positions is challenging. Bensen et al. (2019) show that one potential simple decision rule – the best workers make the best managers – does not always hold. In the face of a more complex decision process, Hoffman et al. (2018) find that managers relying on intuition rather than hard tests make inferior hiring decisions (see also Autor and Scarborough 2008).

We conduct an experiment to understand the difficulties in selecting candidates for supervisory roles in a particularly challenging context. Garment factories in Bangladesh are increasingly interested in promoting women to supervisory positions. The factories face increasing competition for male workers in domestic labor markets and pressure from international brands to provide opportunities for women. More than three quarters of workers in the sewing sections of factories are women, but female supervisors remain very rare; more than 90 percent of supervisors are male (Menzel and Woodruff 2019).

Previous research highlights several factors that make identifying the best female candidates for supervisory positions challenging. Candidates are typically recommended by existing supervisors and higher-level managers. Beaman et al. (2018) report experimental evidence that men are more likely to refer other men for jobs, even when they are aware of better-qualified women. Moreover, the characteristics of the best female candidates may differ from those of the most qualified males. Managers may overlook "obvious" characteristics because they lack experience choosing female candidates. Finally, the feedback to decision-makers whether the decisions were correct is slow. Kahneman and Klein (2009) posit that intuition-oriented experts are likely to outperform harder test-based measures when they receive frequent and rapid feedback on the outcomes of their choices. However, in this context, all new supervisors typically spend many weeks in on-the-job training, and then often underperform initially before growing into the job. For women, the initial performance is further clouded by norms and biases (Macchiavello et al. 2016).

With this in mind, we conduct an experiment in two acts with 30 garment factories in Bangladesh. The factories were asked to nominate a number of women for a training program designed to prepare them for supervisory positions. At the beginning of the project, one-third of the participating factories had no female supervisors, and there was no factory where more than 20 percent of supervisors were female. We asked managers to rank the women in order of their expected performance as supervisors. The number of nominees in each factory was determined through conversations with factory management and was based on the number of supervisory positions they expected to open in the factory in the few months following the start of the training. If we anticipated N supervisory openings, we asked factories to nominate 1.67N candidates for training.

In the first intervention, we divided participating factories into two groups: half of them received a two-hour "management training" session introducing them to diagnostic tests as a means of selecting candidates for promotions. At the end of the training, managers in the management training arm were shown scores from seven diagnostics conducted with the women they had nominated for the supervisory training program. Four of the diagnostics related to cognitive ability or technical skills: tests of literacy, numeracy, fluid reasoning and garments knowledge. We call this 'aptitude'. The other three were hard measures of soft skills that previous research showed were important to the success of female supervisors: confidence in the ability to work as a supervisor, interest in the position, and support from the family for being a supervisor. We call this 'attitude'. After reviewing the diagnostic scores for all of the factory's nominees, the managers were invited to re-rank the candidates if they desired to do so.

In the second intervention, we provided supervisory training to the nominees who were ranked within the highest N positions. We randomized those selected for training into three groups. The first received a 4-day training program in soft skills – stress management, assertive communication and leadership – and a 5-day training program in technical skills – line balancing, quality control, etc. – immediately. The second group received the soft skills training immediately and the technical skills training about six months later. The third group received both training modules six months later. All of the nominees were assigned to trial immediately as assistant supervisors for at least two months, which is the *status quo* method of training candidates for supervisory positions. Assistant supervisors begin by shadowing a current supervisor. After some period of shadowing, they are typically given responsibility for a small number of sewing machine operators. The number of operators under their supervision is gradually increased until they are managing a full section of a line, after which they may be promoted to supervisor.[2]

We are interested in three questions following from the two interventions. First, do managers re-rank nominees following the management training? If they do, which parts of the diagnostic scores do they pay attention to? Second, does either the management training or the supervisory training increase rates of promotion to supervisor? Finally, does either the training or the characteristics that factory managers re-rank on after receiving the management training lead to better performance of the trainees working as supervisors? If the supervisory training affects outcomes, is soft skills training sufficient, or does the combination of soft- and technical skills training outperform soft skills alone? We assess these questions using a combination of surveys and administrative records.

On the first question, the data show that factory managers do re-rank nominees after seeing the diagnostic scores. They are particularly likely to re-rank on our attitudes scores: confidence in supervisory ability, interest in being a supervisor, and the support of family. There is less movement on the aptitude scores. We find that the attitude scores are also positively associated with persistence in the supervisory track, and some measures of performance as a supervisor among those who are promoted. We posit that the attitude diagnostics have a large effect on managers' ranking because they are characteristics that are

[2] Of course, the trainees either in the status quo or in our program may drop off of the supervisory ladder either because they themselves decide they do not want to be a supervisor, or because managers decided they are not qualified for the position.

not likely to be relevant for male supervisory candidates. Similar to Hanna et al. (2014), the attitude diagnostics lead the managers to noticing characteristics that they otherwise are not conditioned to notice.

With regard to the second question, we find that trainees assigned to receive training immediately - the soft skills training group and the soft and hard skills group - have a higher rate of compliance with the factory protocol. That is, they work a larger percentage of days as assistant supervisors during the trial period. Moreover, those receiving the soft-skills training immediately are more likely to be promoted to full supervisor by the time of our final data collection. The attitude measures are strongly correlated with these same outcomes, but the aptitude measures are not. Finally, we find weak support that both the nine-day training and the baseline aptitude scores improve some measures of trainee performance, but stronger support for a relationship between baseline attitude scores and trainee performance.

In addition to the potential to increase the pool of management talent in factories, increasing the rate of promotion of women may have transformative effects on women's lives as well. Women's labor force participation generates changes in development dynamics in lower-income countries, increasing female empowerment and educational attainment of children (e.g. Heath and Jayachandran 2018; Duflo 2012; Qian 2008; Heath and Mubarak, 2017). The garment sector provided an entry into wage work for women in Bangladesh. Even in 2017, garment factories were the workplace for 40 percent of female wage workers with less than tertiary education. As in other countries, employment in the sector is responsible for higher educational attainment and later age of marriage for women (Heath and Mubarak 2015). In a project that builds on the experiment we report here, Uckat (2019) examines the impact of promotion on women's position in the household, both for the women selected for training and for those working under their direction.

The paper proceeds as follows. We explain the two interventions in Section 2, and present results in Section 3. Section 4 discusses the results and concludes.

2. Study design

We implemented two distinct interventions sequentially: training for factory management on using standardised worker evaluations for promotions, and training for female operator to become line supervisors. Figure 1 provides an overview of the design.

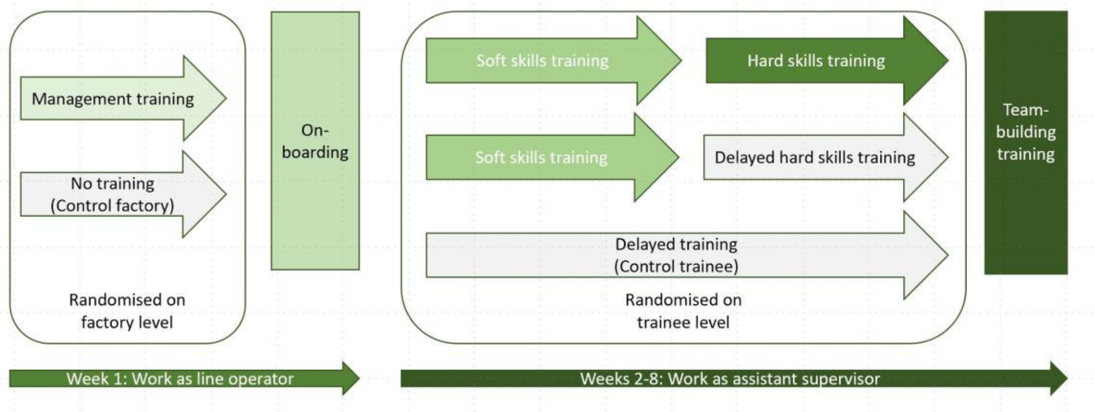

Figure 1: Overview of study design

At the beginning of the project, we determined the number of operators to train in each factory in conversations with the management of the 30 factories that initially agreed to participate. We advised them to select this number based on the number of openings for supervisory positions they expected to have in the following few months.[3] We then asked managers to nominate 1.67 times the number of workers they intended to train, and to rank all nominees based on their expected performance as supervisors. Thus, a factory that intended to train six women nominated 10 for the program, and provided us with a list of the 10 ordered from the first woman they wanted to train to the tenth.[4]

Approximately two weeks after receiving this information from the factory, we conducted baseline surveys and diagnostics with all of the nominees.

[3] The variance in this expected demand led to variance in the number of trainees across factories. For example, some factories anticipated opening new production lines, and hence anticipated the need for a larger number of new supervisors.

[4] We asked factories to nominate additional women because the trainees needed to pass a literacy and numeracy diagnostic in order to be eligible for training. Although we requested that they select workers with at least eight years of schooling, we anticipated that some of the nominees would not pass this examination. We also anticipated that some trainees would decide to drop out before the training began, and need to be replaced. The additional nominees provided a pool of replacements

The survey elicited information on education, labor history, and other demographic data from each of the nominees. We also asked questions about relationships in their household, including measures of their participation in household decision-making. For each nominee, we then conducted a diagnostic test with seven components. The diagnostic, described in more detail in Appendix A, included tests of literacy, numeracy, fluid intelligence and technical knowledge related to garments production. In the analysis, we show results for the individual diagnostics and for an Anderson index (Anderson 2008) that combines these four diagnostics into a single score that we refer to as "aptitude." We also had batteries of questions which we group into three scores reflecting softer skills or attitudes: a measure of how much the nominee's family would support her working as a supervisor, a measure of the nominee's own interest in the position, and a measure of the nominee's confidence to perform in the position. In the analysis, we will sometimes group these three diagnostics using an Anderson index into a single score we refer to as "attitude."

Management training

Our first intervention focused on introducing factory managers to the idea of using quantitative diagnostic tools to select the best candidates for supervisory positions. We refer to this as the "management training" intervention. Factories were randomized into two equally-sized groups. The first received a two-hour training session on the diagnostic tests, and the second served as a control group without any further intervention. The intervention provided data from previous projects that supported the case for promoting female supervisors. It also introduced the managers to the seven different diagnostic scores and allowed them to take part of the test themselves. This training was aimed at mid- and high-level managers – those involved in nominating the female operators and in making decisions about whether they would be promoted.

None of the factories reported using similar diagnostics in their promotion decisions prior to the management training. At the end of the training session, managers were shown the scores for the seven diagnostics that we conducted with the nominees at the same time the management training session was run. After the completion of the management training, managers from factories participating in the management training were offered the opportunity to revise their rankings of the operators they had nominated for the supervisor training program.

For the factories that were in the control group for the management training the original ranking of nominees was changed only if nominees failed the literacy and/or numeracy diagnostic. In that case, all of the nominees were moved up in the original order. The highest ranked nominees were selected for the supervisory training, up to the number determined in advance with the factory. The pool of trainees was selected similarly from the re-ordered list of nominees in the management training factories.

Supervisory training

The highest-ranked operators represented the pool for the second intervention, which was aimed at training the selected operators to be line supervisors. The training program was a streamlined version of the program developed initially by GIZ and described in more detail in Macchiavello et al. (2017). For this project, the training was reduced to nine days, divided into four days of training on soft skills and five days on technical skills. The soft skills focused on stress management, assertive communication, and leadership, elements which earlier research (Macchiavello et al., 2017) identified as being particularly important for female operators.[5] The hard skills training focused on the technical skills required for line supervisors, i.e. production processes, sewing machines, quality control, cutting, finishing, printing, embroidery, and the general supervisory role.

[5] This part of the training also included sessions on understanding harassment, developing integrity and fairness, workers' rights and responsibilities, and human resources management, including types of management styles.

In each factory, we conducted a public lottery to randomly allocate trainees into one of three groups. The first received both the soft- and hard-skills training immediately, the second group received only the soft skills training immediately, and the hard skills training around six months later; and the third received both the soft- and hard-skill training around six months later. The training sessions were held on consecutive weeks at a local training center two days per (six-day) work week.[6]

In addition to the classroom-based training, all of the trainees were expected to begin working as assistant supervisors on the line at the start of the program. We asked factories to choose a set of trial lines, and to assign the selected trainees to those lines. These choices were made before anyone was aware of which operators would be randomized into which of the three training groups. Hence, while the assignment of individual trainees to lines is not random, the assignment of training to the lines is random.

The standard practice in the factories is to train supervisors on the job by having them work for a period as assistant supervisors alongside an experienced supervisor. The assistant supervisors gradually take responsibility for an increasing number of sewing machine operators until they are managing a full section of the line. The group receiving both the soft- and hard-training sessions with six months delay thus mimicked the standard practice at factories, with one modification.[7]

For logistical reasons, the 30 participating factories were divided into five sessions of roughly equal size. The first training session began in November 2016

---

[6] At the end of the classroom training, all trainees, supervisors, line chiefs, and industrial engineering officers in charge of the respective trial lines participated in a team-building session at the factory. This was designed to increase collaboration among those working on the same line to ease the incorporation of the trainee in the management team. This was implemented in the same manner for all three treatment groups.

[7] To increase acceptance of the new female supervisors, on the first day each trainee began working as an assistant supervisor, the project trainers conducted an "onboarding training" that involved line chiefs and supervisors from the lines where the trainees were assigned to trial as assistant supervisors. Lower level managers were provided with a short training on the effectiveness of female line supervisors and the best way to support female line supervisors in succeeding in their new role. The onboarding exercise also included a session in which higher-level management introduced the trainee to the workers on the production line. The onboarding was implemented in the same manner for all three treatment groups.

and the fifth began in March 2017. Three factories dropped out of the project after we conducted the baseline survey, but before the supervisory training began, leaving 27 factories that completed the program.[8] The initial training for the Hard and Soft and Soft Only groups finished in May 2017. Soft and hard skills training began for the control group, and hard skills training for the soft-skill only group, in August 2017; all training was completed in October 2017. We conducted a follow-up survey in each factory just before the control group training started in each factory. The follow-up survey included samples of operators working on lines on which trainees were assigned to work as assistant supervisors, and operators working on lines where the trainees were actually working at the time of the follow-up survey, since they had sometimes moved to another line. We use these data to conduct intention-to-treat[9] regressions on the effectiveness of trainees as supervisors.

3. Results

Trainee selection

We begin by evaluating the effect of the management training on the characteristics of trainees. As noted, after the management training, we offered managers the chance to re-rank their nominees. Managers in 11 of the 13 factories randomized into management training revised their lists.[10] We begin by asking which diagnostic scores are associated with stronger movements up the list of nominees. This is: What information presented in the diagnostics did the managers respond to? We calculate a difference in ranking by subtracting the ranking after management training from the ranking made by the factory prior to

[8] We use the sample of 27 factories for the main results, but show in the appendix the management training results for the 30 factories that continued through the completion of that intervention.

[9] Future versions of the paper will include an analysis using actual assignment of trainees and more formal matching techniques.

[10] See Appendix Tables A.3 and A.4 for results of the re-ranking exercise for the full sample of 15 factories completing that intervention. Those results are very similar to what we report here for the sample of 13 factories that continued through the full program.

the management training. We run the following regression for individual i in factory f:

$$diff\_rank_{i,f} = \alpha + \beta' S_{i,f} + \theta_f + \epsilon_{i,f}$$

where *diff_rank* is the difference between the original rank of a nominee minus the final rank of a nominee[11] and ***S*** is a vector of the seven diagnostics scores or, alternatively, of the indices for aptitude and attitude. We also include factory fixed effects θ and cluster standard errors on the factory level. A positive (negative) value for *diff_rank* means that the nominee moved up (down) in the ranking after the management training. Thus, a positive coefficient implies that nominees scoring highly on that diagnostic were more likely to be moved up in the ranking. Note that the net movement as a result of re-ranking is zero – every move up is associated with an equivalent move down. Hence, a positive coefficient implies that managers placed a higher weight on the characteristic represented by that diagnostic. For this analysis, we include the 138 nominees in 13 factories that completed the management training.

We begin by testing whether the factories were implicitly accounting for the information provided by the diagnostics by regressing the rank of the operators before any potential re-ranking against the seven diagnostic scores for all factories. A higher rank indicates a better position in the ranking. The result of this regression for the sample of taking part in the management training is shown in the first two columns of Table 1.[12] The first column reports the seven scores individually and the second the two indices. Since our intention here is to understand how the baseline diagnostic scores correlate with the rank, we include all participating factories in the regressions reported in the first two columns. Somewhat surprisingly, we find little correlation between the diagnostics and the

[11] Because the number of nominees varies across factories, both the initial cardinal ranking and movements in the ranking may be disproportionately affected by a few factories with a large number of nominees. We report it this way because that is what we defined in the pre-analysis plan. However, regressions using percentile rankings, which reduces the influence of factories with more nominees (though arguably leads to excess influence of factories with a very small number of nominees), produce similar results (available on request from the authors).

[12] The sample on Table 1 is limited to factories randomized into the management training. However, the results are very similar if we use the full sample of factories.

initial ranking. Indeed, a negative coefficient on numeracy is the only significant correlation we find, indicating that factories rank nominees with higher numeracy skills lower. Since this could partly be driven by high correlations among the diagnostic scores, we investigate the relation of the initial rank with the aptitude and attitude indices in column 2. While it appears that the indices are positively related to a better rank, these coefficients are not significant. This indicates that, at baseline there was no correlation between the diagnostic scores and the preference for training the nominees. We can think of several explanations for this. First, factories may have lacked good measures of these skills. Alternatively, they may not have thought these skills were important. Finally, this may simply reflect the highly-selected nature of the pool of nominees to begin with: a typical factory selected one woman for every two or three production lines, that is, around one in 40 female operators.

Columns 3 and 4 show the results using the difference in ranks as defined above as the dependent variable, where we only include factories that were assigned to the management training. Here we see that scores on the numeracy, family support and confidence tests are significantly associated with movements up the list. Processing speed has a large, though statistically insignificant, negative coefficient. With the concern for correlation among the scores in mind, column 4 reports the results using the two indices. We find that managers react to both higher aptitude and higher attitude scores, with movement on attitude being insignificantly larger than movement on aptitude.[13]

Does the re-ranking leave the selection factories with better trainees, that is, those having higher diagnostic scores? And if so, in which dimensions? We address this in Table 2, reporting the results of a treatment effects regression of the form:

$$S_{i,f} = \alpha + \beta \; HR_f + \gamma^{'} X_f + \epsilon_{i,f}$$

[13] The regressions in columns 3 and 4 do not control for the initial ranking of the nominee. When we add the initial ranking, we find that it is insignificant and does not change the magnitude or significance of the diagnostic measures or indices.

where S is a diagnostic skill, HR indicates that factory f participated in the management training, and X is a vector of variables used in the randomization of the management training. We again cluster on the factory level. The coefficient on the treatment variable tells us the difference between the average diagnostic scores of the final trainees – those nominated at a sufficiently high enough level to be included in the training, after any re-ranking – in factories randomized to receive the management training compared with trainees in factories randomized out of the management training.

The top half of Table 2 reports the results for the diagnostics measuring aptitude – literacy, numeracy, fluid intelligence and technical knowledge related to garment production. We see that factories participating in the management training have marginally significantly higher fluid processing and technical knowledge, though in magnitude, the effect is fairly small. However, the aggregate aptitude index is larger and statistically more significant, indicating that the trainees in the selection factories have higher scores in the aggregate aptitude dimension. The bottom half of Table 2 shows the differences in average scores for the three attitude diagnostics and for an Anderson Index of all these three diagnostic scores. The differences for the soft skill diagnostics are much larger than those for the aptitude diagnostics, and much more significant.[14] When testing the cross-equation restriction whether the coefficient is the same for the attitude and aptitude index, we can reject equality with a p-value of 0.015.

Table 2 indicates that the factories completing the management training have trainees with significantly higher attitude diagnostic scores. The baseline balance table shown in Appendix Table A1 indicates that the family support score is imbalanced at baseline (p=0.09), so some portion of the results on Table 2 may be driven by this imbalance. However, the combined results from Tables 1 and 2 do suggest that managers moved toward nominees with higher diagnostic scores following the management training.

[14] Fluid intelligence and technical knowledge have lower variances than the other five diagnostic scores, with standard deviations around 0.10 rather than 0.18-0.23 for the other five scores, and 0.54 for the Anderson index of all seven scores.

In sum, the re-ranking of nominees following the management training led managers to put additional weight on both aptitude and attitude diagnostic measures, especially confidence of the nominees in their ability to perform as supervisors. The resulting pool of trainees has, on average, higher diagnostic scores in both aptitude and attitude dimensions, and the effect is significantly larger for the attitude diagnostics. A question then is whether the shift in the characteristics of the trainee pool produces better outcomes. We will examine this using both promotion rates and performance.

Training outcomes

Recall that the re-ranking exercise and the removal of nominees who failed the literacy and /or numeracy diagnostic resulted in the final list of nominees selected for training. The selected operators were next randomized into one of three groups, one receiving both the soft- and hard-skills training immediately, one receiving only the soft-skills training immediately, and one receiving neither training immediately.. We use this randomization to examine the effects of both the selection and supervisory training on three downstream outcomes. First, we ask whether the supervisory training increased the skills of those receiving it. Second, we ask whether either the supervisory training or improved selection resulting from the management training resulted in higher levels of continuation on the path to promotion. For example, if the supervisory training leaves workers better prepared to be supervisor, they may be more interested in continuing to work as an assistant supervisor, or factory managers may judge them more capable as supervisors. Finally, we ask whether either better selection or the supervisory training affected the performance of the trainees.

To measure these outcomes, we use a combination of survey and administrative data. Survey data collected just prior to the start of the delayed training sessions allows us to assess the short-term effect of soft-skill and combined soft- and hard-skill training on skills and on the number of days working as a supervisor during the trial period. Regular telephone follow-up surveys during the initial training period allow us to measure the percentage of days the trainees worked as assistant supervisors rather than operators, and whether they were promoted to a full supervisory position. Finally, surveys of operators working under the direction of the trainees and administrative data provide measures of trainee performance. In addition to the effects of the management training and supervisory training interventions, we are interested in how the baseline diagnostics themselves relate to both completion of the training program and performance in a supervisory role.

Trainee skills

We begin by asking whether the soft-skills or soft and hard skills training increased the skills of the trainees. Table 3 shows the effect of supervisory training on five skills measures. The reported regressions are ANCOVAs, with baseline measures of the same skills taken as controls. The first measure is the score on a 86-item test of knowledge about garment processes and production. This is the garments knowledge diagnostic that is described in Appendix 1. The second column is a self assessment of the trainee's expected performance as a supervisor. We asked trainees to rate both the typical supervisor and their expected performance on a scale of one to ten. The self assessment measure takes the difference between the two ratings as reported by the trainee. A second self assessment is the self-efficacy score, which aggregates responses from the Generalized Self Efficacy scale (Scholz et al., 2001).[15] The fourth measure is internal locus of control, which we measure by asking respondents to one of two statements in each of seven different pairs of statements (based on Rotter 1965). The final measure is the score on the Generalized Anxiety Disorder 7-item scale. This measures stress and anxiety over the preceding two weeks. Among these

[15] Scholz et al. (2002) show that the questions produce reliable responses in 25 countries. Recent work by Laajaj et al (2019) suggests that psychological measures may be measured with more noise in lower-income countries, and that there may also be concerns with response bias. The responses to the Generalized Self Efficacy scale have reasonable internal consistency, with a Cronbach alpha of 0.75 in our data.

five measures, the supervisory training has a significant effect only on the trainee's self-assessment of their expected performance as a supervisor. The regressions also control for participation by factory management in the management training. The management training has a positive measured effect on all five follow-up skills conditional on baseline skills, but the effect is significant only for the general self-efficacy measure. [16]

Promotions

The next outcome of interest is whether the trainees complete the training program and are promoted to supervisor. At one extreme, some operators dropped out of the program shortly after being selected as a participant; at the other extreme, some were officially promoted to a supervisory position by the time of the follow-up survey. Just over four-fifths (81 percent) of the trainees reported working as an assistant supervisor at least once in weekly phone surveys. Table 4 reports the effects of selection and supervisory training on two measures of program participation. The first two columns of Table 4 regress the percentage of days during the period between the start of the program and the follow-up survey that each trainee worked as an assistant supervisor. This data was collected in high-frequency phone surveys during the trial period of the operators. Management often returned workers to machines during peak production periods, or when there was particular production pressure on a given day. Still, compliance in this regard was reasonably high: among those working at least one day as an assistant supervisor, 75 percent worked more than half of the available days, and one-third worked every day in a supervisory capacity. We find that the percentage of work days the trainee worked in a supervisory role is significantly higher among trainees receiving at least one of the formal supervisory training sessions when they began their trial as an assistant supervisor. This is particularly relevant given that the normal procedure in most

[16] We have not yet adjusted these results for multiple hypothesis testing, but will in the future draft.

factories is to start the trial period without any formal training. We do not find that trainees in factories receiving the management training were more likely to work in a supervisory role.

The second column of Table 4 adds controls for the Anderson indices of aptitude and attitude. We find no evidence that trainees with higher aptitude scores worked more often as an assistant supervisor, but the attitude index is very strongly associated with work as a trainee. The other columns in Table 4 repeat the same two regressions, with promotion to full supervisor the dependent variable. Columns 3 and 4 use data from the follow-up survey at the end of the trial, while columns 5 and 6 use data collected in household surveys, described in more detail in Uckat (2019). The household survey was conducted between January and March of 2018, six to ten months beyond the follow-up survey, and after all trainees had received the full training in soft and hard skills. The results in columns 3 and 4 Table 4 indicate that trainees receiving earlier training were more likely to have been offered promotion and to be promoted, though the difference with the control group is significant only for those in the soft and hard skills training group. Recall that at this point, the control group had not yet received any training. By the time the household survey was conducted in early 2018, trainees in all groups had completed training. The data show that while around two-thirds of the trainees in each of the three randomization groups were offered a promotion, those in the group that received early soft training were more likely to have accepted the promotion and hence be working as a supervisor (column 6). Though the effects of the HR *management training* are not significant even in the longer run, the measured effect on both offers and promotion rates is substantial.

Performance as a supervisor

Does the formal training affect the performance of trainees as supervisors? We note at the start that the design of the experiment is better suited to answer the question of continued interest and promotion than performance, given that the experimentally-generated variation lasts only a short time after the normal trial period ends. Only one-fifth of the trainees had been offered a promotion to full supervisor at the time of the follow-up, though an additional 40 percent were still working as assistant supervisors. We use two sources of data to attempt to provide evidence on performance as a supervisor. The first is surveys of 708 operators working on lines where the trainees were assigned to work during the trial period. These surveys were conducted at follow-up, when the early training was completed but the late training not yet started. We ask these operators to rate, on a scale of one to ten, a typical supervisor in the factory, and then to rate the trainee on the same scale.

We report ITT regressions in Table 5, using the line where the trainee was assigned to work.[17] On average, trainees are rated about a half point (one-third of a standard deviation) lower than typical supervisors. In the first two columns of Table 5, we show results from regressing the rating against variables indicating the trainee participated in early soft-only or hard- and soft training. The regressions also control for the responding operator's rating of a typical supervisor in the factory. We find that all three variables related to training – soft-only, hard and soft, and factory-level management training – are positive but far from statistically significant. In the second column, we include the two Anderson indices, and find that both are significantly associated with the supervisees' opinion of the performance of the trainees, the aptitude score marginally so, and the attitude score much more strongly so.

We also ask operators a series of questions related to their wellbeing, including the GAD-7 diagnostic discussed above, questions about verbally and physically abusive behavior on the line, aspirations to be a supervisor and a

[17] There are two types of non-compliance relevant for interpreting the results. First, as we have noted, almost one-fifth of the trainees never work as assistant supervisors. For the survey questions, we therefore begin by asking operators if they recall the trainee working as a supervisor. Those responding "not at all" or they report that the trainee never worked in a supervisory capacity on their line, we ask for a generic comparison between a typical supervisor and a "typical female supervisor." Just over one-third of operators (38 percent) answer the generic question.# Second, some of the trainees work on a line other than the line where they were assigned. Of course, these movements may be endogenous, so we present the intention to treat regressions as the primary results.

question on general happiness over the previous two weeks. These questions are detailed in Appendix 3. In columns 3 and 4, we regress an index that combines these measures into a single "operator well being index" against the training and (in column 4) the diagnostic indices. Again, we find positive coefficients on the supervisory training, though they are both insignificant in the regression without diagnostic controls (column 3). When we add the two diagnostic indices in column 4, we find that the hard and soft training marginally significantly leads to higher operator well-being. Moreover, the attitude diagnostic score itself is highly significantly associated with supervisee well being, and we can reject equality between the effects of the attitude and aptitude indices.

In addition to the survey data, we use administrative records measuring daily productive efficiency, quality defect rates and absenteeism at the production-line level. These measures are also described in Appendix 3. We report results on these performance measures on Table 6. Note that, in addition to the two sources of non-compliance discussed above with regard to the operator opinions, there is a third measurement issue that is relevant in interpreting these results. Our productivity measures are made at the production line level. More complex products are produced on lines that typically have two or even three line supervisors, each responsible for only a part of the line. The line-level measures will therefore reflect the combined effort of more than one supervisor. Moreover, even on lines with only a single supervisor, our trainees were almost always working as an assistant supervisor, and hence responsible for only a part of the line.

With these caveats in mind, we report regressions using three efficiency measures on Table 6. The data give little indication that the training had an effect on the performance of trainees. The factory-level management training is associated with significantly worse outcomes for quality defect rates, but neither soft only nor hard and soft supervisory training have significant effects on any of the three outcomes, and the signs indicate insignificant effects in a detrimental direction in several instances. Similarly, the aptitude and attitude indices are always insignificant, though the attitude index does have signs in the correct direction for all three measures. The lack of an effect may not be surprising given that only a small percentage of the trainees are working as full supervisors at the time of the follow-up, and even those would have accumulated little experience at the time. However, given the design, this is the latest point in time for which we have a pure experimental comparison.

4. Discussion of results

After four decades of rapid growth, the garment sector in Bangladesh now represents around one-eighth of GDP and more than 80 percent of exports. Importantly, the sector remains the primary source of employment for women working full time outside agriculture. The 2017 Bangladesh Labor Force Survey indicates that among women without tertiary education and working full time, 40 percent of those employed outside agriculture are employed in the garment sector. However, women's role in the sector is limited almost exclusive to production-level positions; fewer than one in ten factory managers are women. Facing new pressures from several sources, factories are increasingly interested in promoting women to supervisory positions. The results of our intervention should be viewed in the context of the recent interest in promoting women.

The shift toward increased interest in promoting women represents an important cultural shift in the factories. Viewed from this perspective, we find the results related to the role of the attitudes of potential female supervisors to be particularly important. First, we find no evidence that managers take these attitudes into account in the initial rankings that they provide to us. The rankings are orthogonal to the attitude diagnostics. Second, we find that these trainee attitudes are related to both the intensity of treatment - the percentage of days during the trial period the trainees work in a supervisory role - and to promotion to full supervisor. Third, we find that the attitude measures are related to performance in a supervisory capacity, as judged by subordinates, and to the reported well-being of those subordinates.

The three results combined suggest an underlying reason for the difficulty in making sharp transitions in practices. Where managers have little experience with selecting females for promotion, and where the success of the candidates depends on characteristics not well understood by managers, the selected candidates are likely to have suboptimal performance once promoted. The fact that when the diagnostic results are revealed to managers, they re-rank operators in a manner that emphasizes the attitude diagnostic scores provides evidence for the importance of hard measures of candidate abilities (Hoffman et al. 2018), perhaps particularly in these circumstances.

References

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.

Artz, B. M., Goodall, A. H., & Oswald, A. J. (2017). “Boss Competence and Worker Well-Being,” *ILR Review*, *70*(2): 419–450.

Autor, D. H., & Scarborough, D. (2008). “Does job testing harm minority workers? Evidence from retail establishments,” *The Quarterly Journal of Economics*, *123*(1): 219-277.

Beaman, L., Chattopadhyay, R., Duflo, E., Pande, R. & Topalova, P. (2009). “Powerful Women: Does Exposure Reduce Bias?” *The Quarterly Journal of Economics*, 124(4): 1497–1540.

Beaman, L., Keleher, N. & Magruder, J. (2018). “Do Job Networks Disadvantage Women? Evidence from a Recruitment Experiment in Malawi,” *Journal of Labor Economics*, 36(1): 121-157.

Benson, A., Li, D. &Shue, K. (2019). “Promotions and the Peter Principle,” *Quarterly Journal of Economics*, 134(4): 2085-2134.

Duflo, E. (2012). “Women Empowerment and Economic Development.” *Journal of Economic Literature*, 50(4):1051–1079.

Glover D, Pallais A & Pariente W. (2017). “Discrimination as a Self-Fulfilling Prophecy: Evidence from French Grocery Stores,” *Quarterly Journal of Economics*.132 (3):1219-1260.

Heath, R. and M. Mobarak (2015). “Manufacturing growth and the lives of Bangladeshi women,” *Journal of Development Economics*. 115: 1-15.

Heath, R. and S. Jayachandran (2018). “The Causes and Consequences of Increased Female Education and Labor Force Participation in Developing Countries.” In Averett, S. L., Argys, L. M., and Hoffman, S. D. (eds), *The Oxford Handbook of Women and the Economy*. Oxford University Press, New York, NY

Hoffman, M., Kahn, L. B., & Li, D. (2017). “Discretion in hiring,” *The Quarterly Journal of Economics*, *133*(2):765-800.

Kahneman D, & Klein G. 2009. “Conditions for intuitive expertise: a failure to disagree,” *American Psychologist*. 64: 515-526.

Laajaj, R.,Macours, K. and Pinzon Hernandez, D. A., , Arias, O.,Gosling, S. D.,Potter, J. , Rubio-Codina, M. &Vakis, R., (2019), "Challenges to capture the big five personality traits in non-WEIRD populations," *Science Advances*.

Lazear, Edward P., Kathryn L. Shaw, &Christopher T. Stanton. (2013). "The Value of Bosses". *Journal of Labor Economics*. 33(4): 823-861.

Menzel, A. and C. Woodruff, (2019). "Gender Wage Gaps and Worker Mobility: Evidence from the Garment Sector in Bangladesh", working paper, Oxford University.

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.

Rotter, J. B. (1966). Generalized Expectancies for Internal versus External Control of Reinforcement. Psychological Monographs: General and Applied, 80(1):1–27

Scholz, U., Dona, B. G., Sud, S., & Schwarzer, R. (2002). "Is general self-efficacy a universal construct? Psychometric findings from 25 countries," European journal of psychological assessment, 18, 242-251.

Uckat, H. (2019). "Women's Promotions and Intra-household Bargaining: Evidence from Bangladesh," working paper, Oxford University.

Table 1: Operator rankings and diagnostic scores

	Nomination rank		Movements in rank (after HR training)	
	(1)	(2)	(3)	(4)
Literacy	1.084		1.941	
	(1.423)		(2.798)	
Numeracy	-3.161**		5.169**	
	(1.372)		(2.138)	
Processing speed	2.429		-5.264	
	(3.853)		(4.112)	
Garment knowledge	4.826		4.757	
	(2.997)		(4.652)	
Family support	1.192		3.560*	
	(1.842)		(1.856)	
Interest	0.0887		0.637	
	(2.195)		(1.485)	
Confidence	0.0531		5.702**	
	(1.058)		(1.937)	
Attitude		0.161		1.852***
		(0.355)		(0.509)
Aptitude		0.247		1.105**
		(0.293)		(0.501)
P-value (Attitude=Aptitude)		0.87		0.30
Factory FE	Yes	Yes	Yes	Yes
Observations	243	243	138	138
Number rankings	27	27	13	13

Standard errors clustered at the factory level in parentheses. * p<0.1, ** p<0.05, *** p<0.01

Table 2: Characteristics of trainees

Panel A: Aptitude

	(1) Literacy	(2) Numeracy	(3) Processing speed	(4) Garment knowledge	(5) Aptitude
Management training	0.0565	0.0378	0.0144*	0.0245*	0.290**
	(0.0337)	(0.0239)	(0.00793)	(0.0131)	(0.116)
Control mean	0.55	0.45	0.32	0.53	0.10
Randomization controls	Yes	Yes	Yes	Yes	Yes
Observations	199	199	199	199	199
Number factories	27	27	27	27	27

Panel B: Attitude

	(1) Family support	(2) Interest	(3) Confidence	(4) Attitude
Management training	0.107***	0.146***	0.0882***	0.618***
	(0.0296)	(0.0477)	(0.0226)	(0.136)
Control mean	0.68	0.68	0.68	-0.10
Randomization controls	Yes	Yes	Yes	Yes
Observations	199	199	199	199
Number factories	27	27	27	27

Standard errors clustered at the factory level in parentheses. * p<0.1, ** p<0.05, *** p<0.01

Table 3: Effect on trainee skills

	(1) Garment knowledge	(2) Self-assessment	(3) Self-efficacy	(4) Internal LOC	(5) Stress
Soft only training	-1.164	0.883**	0.0832	0.335	0.933
	(1.543)	(0.377)	(0.0749)	(0.265)	(1.015)
Hard & Soft training	0.548	0.486*	0.0550	0.218	1.066
	(1.780)	(0.276)	(0.0778)	(0.238)	(0.954)
Management training	0.236	0.397	0.144**	0.188	0.246
	(1.976)	(0.447)	(0.0634)	(0.194)	(0.597)
Control mean	52.57	-1.68	3.27	4.18	11.30
P-value(soft=hardsoft)	0.29	0.27	0.69	0.68	0.90
Imbalance controls	Yes	Yes	Yes	Yes	Yes
randomization controls	Yes	Yes	Yes	Yes	Yes
Factory FE	No	No	No	No	No
Enumerator FE	Yes	Yes	Yes	Yes	Yes
Observations	188	188	188	188	154
Number factories	27	27	27	27	25

ANCOVA specification. Standard errors are clustered at the factory level. * p<0.1, ** p<0.05, *** p<0.01

Table 4: Promotions

	Share of trial days as SV		Promoted (at end of trial)		Promoted (after control training)	
	(1)	(2)	(3)	(4)	(5)	(6)
Soft only training	0.264***	0.282***	0.0407	0.0467	0.150**	0.165**
	(0.0552)	(0.0492)	(0.0338)	(0.0341)	(0.0703)	(0.0685)
Hard & Soft training	0.184**	0.191***	0.0960*	0.0961*	0.0373	0.0408
	(0.0720)	(0.0651)	(0.0509)	(0.0495)	(0.0733)	(0.0675)
Management training	0.0103	-0.0687	-0.0472	-0.0756*	0.139	0.0447
	(0.0900)	(0.0748)	(0.0406)	(0.0408)	(0.111)	(0.107)
Aptitude		-0.00222		0.0148		0.0261
		(0.0347)		(0.0228)		(0.0603)
Attitude		0.142***		0.0361*		0.135***
		(0.0263)		(0.0178)		(0.0423)
P-value(soft=hardsoft)	0.14	0.08	0.25	0.30	0.15	0.13
P-value(Attitude=Aptitude)		0.00		0.42		0.21
Imbalance controls	Yes	Yes	Yes	Yes	Yes	Yes
Randomization controls	Yes	Yes	Yes	Yes	Yes	Yes
Skill scores at BL	No	No	No	No	No	No
Factory FE	No	No	No	No	No	No
Enumerator FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	175	175	188	188	178	178
Number factories	27	27	27	27	27	27

Standard errors are clustered at the factory level. * p<0.1, ** p<0.05, *** p<0.01

Table 5: Effect on operator outcomes

	Operator ratings		Operator wellbeing	
	(1)	(2)	(3)	(4)
Soft only training	0.154	0.184	0.0698	0.0584
	(0.162)	(0.156)	(0.0870)	(0.0848)
Hard & Soft training	0.0460	0.0354	0.152	0.156*
	(0.172)	(0.166)	(0.0915)	(0.0862)
Management training	0.0694	-0.0848	0.0139	-0.0296
	(0.150)	(0.144)	(0.0641)	(0.0655)
Aptitude		0.119*		-0.0640
		(0.0661)		(0.0378)
Attitude		0.185***		0.105***
		(0.0513)		(0.0366)
P-value (soft=hardsoft)	0.41	0.26	0.50	0.39
P-value (Attitude=Aptitude)		0.45		0.01
Control mean	7.83	7.83	-0.01	-0.01
Imbalance controls	Yes	Yes	Yes	Yes
Typical SV assessment	Yes	Yes		
Factory FE	No	No	No	No
Enumerator FE	Yes	Yes	Yes	Yes
Observations	707	707	708	708
Number factories	26	26	26	26

Standard errors are clustered at the factory level. * p<0.1, ** p<0.05, *** p<0.01

Table 6: Effect on production outcomes

	Efficiency		Alteration rates		Absenteeism	
	(1)	(2)	(3)	(4)	(5)	(6)
Soft only training	0.0908	0.157	0.00225	0.00285	0.00192	0.00186
	(1.286)	(1.361)	(0.00375)	(0.00407)	(0.00247)	(0.00259)
Hard & Soft training	0.166	0.168	-0.00356	-0.00336	0.00199	0.00179
	(1.032)	(1.045)	(0.00281)	(0.00279)	(0.00281)	(0.00282)
Management training	-0.570	-0.779	0.0106**	0.0108**	-0.00562	-0.00546
	(1.353)	(1.309)	(0.00434)	(0.00384)	(0.00503)	(0.00533)
Aptitude		0.391		0.00251		0.00138
		(0.575)		(0.00223)		(0.00115)
Attitude		0.0599		-0.00168		-0.00172
		(0.529)		(0.00174)		(0.00216)
P-value (soft=hardsoft)	0.95	0.99	0.14	0.11	0.96	0.96
P-value (Attitude=Aptitude)		0.73		0.21		0.30
Control mean	52.21	52.21	0.07	0.07	0.05	0.05
Imbalance controls	Yes	Yes	Yes	Yes	Yes	Yes
Randomization controls	Yes	Yes	Yes	Yes	Yes	Yes
Factory FE	No	No	No	No	No	No
Observations	5769	5769	5911	5911	5680	5680
Number factories	17	17	23	23	17	17

ITT comparison within the 8 weeks trial period. Date fixed effects always included. Standard errors are clustered at the factory level. * p<0.1, ** p<0.05, *** p<0.01

Appendix

Description of diagnostic tests

A number of different tools were used to assess the abilities of female operators who had been nominated for the Work-Progression & Productivity Toolkit. After the factory has nominated workers, but before the on-boarding session happened, the IPA team visited the factory to assess the abilities of the nominees. Nominees were only disqualified based on numeracy and literacy scores, but we collected much more data, so that we could examine which nominees succeeded towards the end of the program. The areas that were tested and their respective tests are as follows:

1. Literacy: Multiple choice questions testing reading comprehension, vocabulary, basic grammar, paragraph/letter structure, and writing. The maximum score is 20 points.

2. Numeracy: Multiple choice questions testing calculations, fractions, percentages, number patterns, angles, and visual patterns. The maximum score is 20 points.

3. Processing speed (coding): This test is modeled after the Wechsler Adult Intelligence Scale is called 'Digit Symbol' (WAIS-R), 'Digit-Symbol-Coding' (WAIS-III), 'Coding' (WAIS-IV), also known as the Digit symbol substitution test. It is a neuropsychological test sensitive to brain damage, dementia, age and depression. It consists of digit-symbol pairs followed by a list of digits. Under each digit the subject should write down the corresponding symbol as fast as possible. The number of correct symbols within the allowed time is measured. The maximum score is 133 points.

4. Processing speed (Symbol Search): Symbol Search is a subtest of the Wechsler Adult Intelligence Scale (WAIS). The Symbol Search subtest is designed to assess information processing speed and visual perception. High scores require rapid and accurate processing of nonverbal visual information. During Symbol Search, the examinee is asked to mark either the yes or no checkbox with a pencil in response to as many items as possible within 2 min. The maximum score is 60 points.

5. Garments Knowledge: Consists of q21 - q275 in the survey instrument, and contains multiple choice, and open-ended questions testing what machine to use for what operation, names of processes, cause of mechanical issues, cause of quality issues, identifying quality issues in photographs, identifying working condition issues in photographs, and understanding an operation breakdown. The maximum score is 84 points.

6. Family Support: Consists of q811 - q3121_6 in the survey instrument, and gives 5 statements about family support, and asks the respondent to respond on a four-point scale from agree to disagree. An additional three questions asks about the level of support given to other women in the family who work in garment factories. The answers to all questions are recoded so that higher numbers represent more support, and are summed to give a maximum score of 24 points.

7. Interest: The survey instrument has 2 questions about whether they would want to be promoted to supervisor or line chief. Then, we have four questions that indirectly asks whether they are interested in the supervisor position, and asks the respondent to respond on a four-point scale from agree to disagree. The answers to all questions are recoded so that higher numbers represent

Table A.1: Overview of diagnostic test scores

Testing area	**Maximum score**
Literacy	20
Numeracy	20
Processing speed (coding)	133
Processing speed (symbol search)	60
Garment knowledge	84
Family support	24
Interest	18
Confidence	7

more interest, and are summed to give a maximum score of 18 points.

8. Confidence: Consists of asking how they would rate their performance compared to a typical supervisor on a 5-point scale. Then, we have three questions that indirectly ask whether the respondent is confident. The respondent is asked to choose between two statements. One statement that says "I am confident" using various words, and a dummy statement about the factory. The answers are recoded so that higher numbers represent more confidence and are summed to give a maximum score of 7 points.

Disqualification rule: No nominee should be disqualified on any category other than literacy and/ or numeracy. The rule is:

- If nominee gets 24% on literacy and 24% on numeracy, she fails
- If nominee gets 24% on literacy and 25% on numeracy, she passes
- If nominee gets 0% on literacy and 100% on numeracy, she fails

Table A.2: Factory balance, HR training

	Obs	Control	Obs	Selection	p-value
Participated in prev project"	15	0.33	15	0.40	0.72
Produces knit garments	15	0.53	15	0.47	0.73
Number of workers	15	2262.07	15	3310.93	0.17
Share of female workers	15	0.57	15	0.58	0.75
Date of joining programme	15	20387.13	15	20352.53	0.75
Number of supervisors	15	61.60	14	57.21	0.83
Share of female supervisors	15	0.04	14	0.06	0.44
Number of lines	15	26.80	14	36.57	0.35
(A)PM's age (avg)	16	37.72	14	36.95	0.57
(A)PM's education years (avg)	16	10.58	14	10.59	0.98
(A)PM's spouse works (avg)	16	0.07	14	0.07	1.00
(A)PM exposure (avg)	16	0.98	14	0.98	0.78
(A)PM nr. male SV are better (avg)	16	4.66	14	5.05	0.49
Actual to calculated trainees	15	3.68	12	3.86	0.89
SV/Line Chief IAT (avg)	16	-0.24	14	-0.28	0.57
Factory dropout	16	0.13	15	0.13	0.95
Literacy score	15	0.49	14	0.54	0.19
Numeracy score	15	0.42	14	0.42	0.82
Processing speed score	15	0.31	14	0.32	0.32
Garments knowledge score	15	0.54	14	0.55	0.52
Family support score	15	0.68	14	0.74	0.09
Interest score	15	0.69	14	0.73	0.41
Confidence scor	15	0.69	14	0.73	0.22

Robust standard errors in last column. Scores are the average by factory for all operators nominated by that factory.

Table A.3: Operator training balance

	Control(N=65)	Soft only (N=67)	Hard & Soft (N=67)	p-value
Age	25.26	25.88	25.79	0.28
Married	0.77	0.76	0.72	0.79
Household members	3.18	3.51	3.12	0.22
Household head	0.22	0.25	0.22	0.89
Migrant	0.63	0.69	0.60	0.55
Education years	8.38	8.36	8.42	0.98
Experience in garment sector	5.60	5.84	6.36	0.12
Tenure	3.14	3.40	3.84	0.12
Nr. of factories worked in	1.28	1.34	1.31	0.96
Exposure to female SV	0.68	0.46	0.51	0.04
Nr. male SV are better	3.34	3.37	3.57	0.78
Internal locus of control	4.43	4.57	4.45	0.82
Grit	2.90	2.98	2.92	0.66
Self-efficacy	3.58	3.66	3.60	0.46
Emotional competence	3.05	3.15	3.11	0.17
Multi-factor Leadership	3.67	3.72	3.78	0.19
Life satisfaction	7.72	7.33	7.52	0.45
Numeracy	1.12	0.87	1.19	0.12
Self-assessment	0.12	0.25	-0.15	0.31
Ambition	2.35	2.18	2.12	0.30
Numeracy score	0.49	0.47	0.46	0.57
Literacy score	0.59	0.55	0.60	0.44
Processing speed score	0.34	0.32	0.33	0.37
Garment knowledge score	0.55	0.53	0.56	0.21
Family support score	0.74	0.72	0.74	0.91
Interest score	0.74	0.77	0.73	0.47
Confidence score	0.75	0.71	0.71	0.24

P-value of joint significance test of Soft Only=Hard&Soft=0 in last column, with standard errors clustered at factory level.

Table A.4: Operator rankings and diagnostic scores, with dropout factories

	Nomination rank		Movements in rank (after HR training)	
	(1)	(2)	(3)	(4)
Literacy	0.923		1.725	
	(1.356)		(2.734)	
Numeracy	-2.910**		5.211**	
	(1.297)		(2.102)	
Processing speed	2.622		-5.392	
	(3.637)		(4.050)	
Garment knowledge	4.545		4.603	
	(2.894)		(4.586)	
Family support	0.999		3.574*	
	(1.766)		(1.820)	
Interest	0.362		0.446	
	(2.123)		(1.377)	
Confidence	0.0181		5.670**	
	(1.025)		(1.913)	
Attitude		0.165		1.820***
		(0.333)		(0.503)
Aptitude		0.253		1.027*
		(0.277)		(0.482)
P-value (Attitude=Aptitude)		0.86		0.25
Factory FE	Yes	Yes	Yes	Yes
Observations	257	257	143	143
Number rankings	28	28	13	13

Standard errors clustered at the factory level in parentheses. * $p<0.1$, ** $p<0.05$, *** $p<0.01$

Table A.5: Characteristics of trainees with dropout factories

Panel A: Aptitude

	(1) Literacy	(2) Numeracy	(3) Processing speed	(4) Garment knowledge	(5) Aptitude
Management training	0.0564	0.0317	0.0117	0.0246**	0.276**
	(0.0344)	(0.0236)	(0.00865)	(0.0114)	(0.109)
Control mean	0.54	0.45	0.32	0.54	0.13
Randomization controls	Yes	Yes	Yes	Yes	Yes
Observations	233	233	233	233	233
Number factories	29	29	29	29	29

Panel B: Attitude

	(1) Family support	(2) Interest	(3) Confidence	(4) Attitude
Management training	0.111***	0.137**	0.0816***	0.602***
	(0.0300)	(0.0501)	(0.0205)	(0.134)
Control mean	0.67	0.67	0.68	-0.11
Randomization controls	Yes	Yes	Yes	Yes
Observations	233	233	233	233
Number factories	29	29	29	29

Standard errors clustered at the factory level in parentheses. * p<0.1, ** p<0.05, *** p<0.01

Table A.6: Effect on trainee skills

	(1) Garment knowledge	(2) Self-assessment	(3) Self-efficacy	(4) Internal LOC	(5) Stress
Soft only training	-0.799	0.963**	0.113	0.290	0.998
	(1.724)	(0.405)	(0.0783)	(0.270)	(1.193)
Hard & Soft training	1.331	0.513*	0.0587	0.252	1.074
	(1.708)	(0.287)	(0.0774)	(0.249)	(1.027)
Control mean	54.34	-1.17	3.41	4.53	12.54
P-value(soft=hardsoft)	0.19	0.24	0.48	0.90	0.95
Imbalance controls	Yes	Yes	Yes	Yes	Yes
Randomization controls	No	No	No	No	No
Factory FE	Yes	Yes	Yes	Yes	Yes
Enumerator FE	Yes	Yes	Yes	Yes	Yes
Observations	188	188	188	188	154
Number factories	27	27	27	27	25

ANCOVA specification. Standard errors are clustered at the factory level. * p<0.1, ** p<0.05, *** p<0.01

Table A.7: Effect on promotion

	(1) Share of trial days as SV	(2) Promoted (at end of trial)	(3) Promoted (after control training)
Soft only training	0.222***	0.0258	0.167*
	(0.0612)	(0.0362)	(0.0839)
Hard & Soft training	0.177**	0.102*	0.0609
	(0.0797)	(0.0553)	(0.0840)
P-value(soft=hardsoft)	0.44	0.15	0.25
Imbalance controls	Yes	Yes	Yes
Randomization controls	No	No	No
Skill scores at BL	No	No	No
Factory FE	Yes	Yes	Yes
Enumerator FE	Yes	Yes	Yes
Observations	175	188	178
Number factories	27	27	27

Standard errors are clustered at the factory level. * p<0.1, ** p<0.05, *** p<0.01

Table A.8: Effect on operator outcomes

	(1) Operator ratings	(2) Operator wellbeing
Soft only training	0.178	0.0845
	(0.171)	(0.0872)
Hard & Soft training	0.0542	0.167*
	(0.178)	(0.0887)
P-value (soft=hardsoft)	0.36	0.51
Control mean	7.74	-0.07
Imbalance controls	Yes	Yes
Typical SV assessment	Yes	
Factory FE	Yes	Yes
Enumerator FE	Yes	Yes
Observations	707	708
Number factories	26	26

Standard errors are clustered at the factory level. * p<0.1, ** p<0.05, *** p<0.01

Table A.9: Effect on production outcomes

	Efficiency		Alteration rates		Absenteeism	
	(1)	(2)	(3)	(4)	(5)	(6)
Soft only training	0.204	0.226	0.00146	0.00163	0.00302	0.00297
	(1.416)	(1.481)	(0.00306)	(0.00332)	(0.00210)	(0.00216)
Hard & Soft training	0.374	0.374	-0.00288	-0.00287	0.00294	0.00275
	(1.047)	(1.055)	(0.00312)	(0.00309)	(0.00263)	(0.00266)
Aptitude		0.148		0.00115		0.00152
		(0.635)		(0.00209)		(0.000927)
Attitude		0.0654		-0.00161		-0.000309
		(0.502)		(0.00146)		(0.00207)
P-value (soft=hardsoft)	0.89	0.91	0.25	0.23	0.96	0.89
P-value (Attitude=Aptitude)		0.93		0.38		0.50
Control mean	48.80	48.80	0.07	0.07	0.04	0.04
Imbalance controls	Yes	Yes	Yes	Yes	Yes	Yes
Randomization controls	No	No	No	No	No	No
Factory FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5769	5769	5911	5911	5680	5680
Number factories	17	17	23	23	17	17

ITT comparison within the 8 weeks trial period. Date fixed effects always included. Standard errors are clustered at the factory level. * p<0.1, ** p<0.05, *** p<0.01