Managerial Capital and Productivity: Evidence from a Training Program in the Bangladeshi Garment Sector

Rocco Macchiavello, Andreas Menzel, and Christopher Woodruff#

3 October 2014
Preliminary and incomplete draft: Please do not circulate

Abstract: Data from the garment factories in Bangladesh show that four of every five production workers are women, while just over one in 20 supervisors is a woman. Is it likely that 95 percent of the management talent comes from 20 percent of the workforce? We confront this question head on by providing training to four women and one man in a large number of factories. Promoting all five trainees clearly induces behavior on the part of the median factory different from what it has done in the past, and almost certainly what it would have done in the absence of our intervention. Using a randomly selected comparison group, we examine the whether the trainees continue to work at he same factory, whether they are given a trial as a supervisor, and whether they are promoted. We find that the female trainees are as likely as the male trainees to remain at the factory, but less likely to be tried out or to be promoted. In management simulation exercises, the female trainees outperform the male trainees. Using detailed data from production lines on which the trainees work as supervisors, we find some evidence that the male trainees outperform the female trainees, though the differences across gender are not statistically significant. Evidence from survey responses and exercises suggests that the female trainees face some initial resistance as supervisors, which could account for the lower initial performance on the line.

Macchiavello: Associate Professor of Economics, University of Warwick, CV4 7AL, UK; CEPR, BREAD
Menzel: PhD student, Department of Economics, University of Warwick, CV4 7AL, UK
Woodruff: Professor of Economics, University of Warwick, CV4 7AL, UK; CEPR, NBER, BREAD, IZA.
The project was carried out with the financial and logistical support of Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ), the International Growth Center, the IFC, the ESRC-DFID program in growth and the IPA SME initiative. Behind the scenes among many of those supporting the project is the UK Department for International Development (DFID), which also provided initial assistance in making contacts to begin the project. We thank the factories and workers who have participated in this project for their time and cooperation, the IPA-based staff in Dhaka who carried out the project on the ground, and Paula Lopez-Pena and Athena Sharma for research assistance.

Managerial Capital and Productivity: Evidence from a Training Program in the Bangladeshi Garment Sector

I. Motivation

Productivity differences across firms in the same markets are both surprisingly large and persistent (Syverson 2011). Moreover, the dispersion in efficiency appears to be much higher in lower income countries (Hsieh and Klenow 2008). Within countries and industries, efficiency does appear to be higher where firms face more competitive markets (Syverson 2004 Backus 2014). Firms face pressure to increase efficiency from both product and financial markets (see, for example, Syverson (2004) Foster, Haltiwanger and Syverson 2008). Bloom and Van Reenan (2007) and Bloom, Sadun and Van Reenan (2012) show that management practices are better in firms facing more product market competition and in firms with dispersed ownership (rather than family control). The factories in our setting all sell in highly competitive international markets, and hence face substantial pressure at least from the product market side. They are typically owned by a small group of investors, and so could face lower pressure on the financial market side.

The work by Bloom and various co-authors raises an important puzzle: The management practices they study are well known and seemingly simple to implement. Why do firms fail to implement them? One explanation offered by Gibbons and Henderson (2012) is that changing practices is actually quite complex, both because individual practices more be complementary to one another (see also Ichniowski, Shaw and Prennushi 1997) and because management involves both formal rules and informal norms. Managers may know what is wrong, know how to fix what is wrong, but yet be unable to implement the required changes because they unable to shift the equilibrium of the game between managers and workers (or between managers at one level of the hierarchy and managers at another level).

With this set of issues in mind, we examine productivity in factories in the ready-made garment (RMG) sector in Bangladesh. We focus on the choice of production-line managers, and on the role of gender in that choice. The ready-Bangladeshi RMG sector now accounts for around 13 percent of GDP and employs around 4 million workers, 80 percent of whom are female. However, in our data, 10 percent of line supervisors – the lowest and most populous management level – are women.[1] The question we ask is whether the 80/20, 10/90 ratios are efficient from a productivity perspective, or whether factories are promoting men over more able women. To do this, we implement an operator-to-supervisor training program in around 60 factories. [2] The program

[1] Our data likely overstate the participation of women in management because our sample includes two factories which are entirely female, at all levels. These factories are quite unusual in the sector.

[2] The training program was designed by the German bilateral aid agency, Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ), together with local training companies.

induces factories to promote more female SVs that they otherwise would. We then compare the performance of females and males trained in the program, and the response of operators working for them, using both very detailed production data and in-factory surveys.

Given that the firms operate in a very competitive international market, the default assumption should be that the promotion decisions are efficient. But as we discuss in the next section, the circumstances are such that we view it as possible that firms are in inefficient equilibrium where complementarities and collective action make transitioning to the more efficient equilibrium difficult. In short, because only men are promoted, women do not invest in the skills required to become supervisors and men enter factories with the expectation of becoming supervisors. As a result of this and cultural norms, women require more initial training to become supervisors. The training involves general skills, and so individual firms are reluctant to provide it – and for credit constraint and coordination issues, individual workers are unable to acquire it. When male workers see women being promoted, they resist the female supervisors.

If it is in an inefficient equilibrium, how did the industry end up there? Historically, the garment sector has been an entry point for women into wage labor markets, and Bangladesh is no exception. The 1991 census indicates there were just over 500,000 female wage workers (2% of the population 16-65 years of age), while in 2001 there were almost 1.9 million (5.5% of the population 16-65). ILO data indicate that female labor force participation increased from 22 percent in 2000 to 34 percent in 2010. [3] Initially, then, firms would have promoted males because they were a larger share of the workers and because the labor force attachment of women was uncertain.[4]

In spite of its success by many measures, the RMG sector in Bangladesh is seen as less efficient than many of its competitors. Many associated with the industry view the rapid growth as entirely dependent on wage rates that are lower than any other major garment exporter. Internationally comparable data are difficult to obtain, but one study of wage rates places the average rates for machine operators in Bangladesh in 2011 at less than one-third the level in China, 40 percent of the level in Vietnam and half the level in Indonesia (Center for American Progress, 2013). These lower wage rates allow factories to be competitive even if they are less efficient. Again, comparable data are difficult to obtain, but trade and industry statistics suggest that the output per worker in Bangladesh - $5900 in 2013 – are a third lower than output per worker in Vietnam - $7800 in 2013.

[3] The 2011 census data from IPUMS do not include the detailed class of worker data, so we do not have more recent census data.

[4] Women have usually been under-represented in management, though often by less extreme levels than we observe in Bangladesh. In the U.S., for example, by 1910 almost 60 percent of workers and 30 percent of managers in the garment sector were women. By 1950, 80 percent of workers– the same proportion as in Bangladesh now – and 50 percent of managers were women. (Calculations by the authors based on U.S. census data obtain from IPUMS.)

The sewing section of garment production is a particularly good place to measure productivity difference. As we discuss in more detail later, the quantity of output can be measured and compared across factories producing different products. Moreover, we have good measures of within-factory productivity, measured at the production-line level. We are able to measure productivity at the line level, and make very careful and accurate comparisons both within and across factories using "standard minute values," or SMVs. Calculating an SMV requires measuring the time it takes for a fully efficient worker to sew each seam or other stich, and aggregating these steps for each garment. Using comparable SMVs, we are able to compare the output of producers of (many varieties of) shirts with the output of producers of (many varieties of) pants. By comparison, we usually have either physical output in a very homogeneous sector (e.g., Syverson 2004; Foster, Haltiwanger and Syverson, 2008), or are limited to revenue. With regard to the role of line supervisors on productivity, there is one disadvantage: the lowest level of aggregation of data is at the production line, and there are typically two to three line supervisors on each line. The team aspect of line supervision adds some noise to the measurement of a given individual's contribution.

Most industry insiders confirm impressions that Bangladeshi factories are less efficient than those in competing countries. Our early conversations with stakeholders pointed to mid-level management as one place where skill-upgrading was needed, and where training is lacking. Consistent with this, our survey data indicated that only around 15 per cent of line supervisors report having received any formal training for the positions either inside or outside the factory, and the majority of this training was provided inside the factory. Arguably, developing strong managerial capacity in RMG is a precursor to supplying managers to sectors involving more complex production technologies and more capital.

II: Analytical framework

At the heart of the exercise we conduct are the questions: Has the industry found the most efficient equilibrium with regard to selection of supervisors? The most straightforward way to start with the rather obvious statement that the answer to this question could be either yes or no. Firms in our sample have incentives to efficient, and indeed, given the level of competition their survival depends on reaching some level of efficiency. Men may make better managers, or may have – as higher level managers often told us – stronger attachments to the labor force. In that case, they may be willing to make larger investments in skills, given the longer period over which they may earn a return to those investments. Given high rates of promotion among those starting as male operators – likely in excess of 50 percent – there may be selection into the sector by males with supervisory talent.

Data from surveys in the factories offer some support for the selection story, but little support for a stronger attachment of males to the industry. In samples of randomly selected operators, males report having worked in the sector only slightly (and insignificantly) longer – 5.6 years compared with 5.1 years for women. They also have comparable tenure in their current factories, 33 months compared with 31 months for women. When asked how long they expect to remain in the factory, the mean response (median) for men is 71 months (60 months), while for women the mean (median) response is 48 (49) months. Note however, that expected tenure may depend on the prospects for promotion, and indeed we find no differences between men and women among those selected for training. Moreover, the idea that women leave the labor force after having children is not supported in the data. Almost half (46 percent) of randomly selected female operators, and over 40 percent of the women selected as potential trainees, have children.

On the other hand, we do find that males entering the factory have more schooling than females. 10 percent of males and only 3.5 percent of females have secondary schooling completion certificate. And 84 percent of randomly selected male operators say they would like to be a supervisor in the future, compare with only 49 percent of females. While there is a sense of stronger selection among males, at least in measured characteristics the females outnumber the males in any education category, because of their four to one numerical advantage.

These data suggest a different equilibrium, in which women are not promoted and therefore do not invest in skills, leave factories earlier and therefore are less qualified than men to be supervisors. A way out of this equilibrium is to provide training to enough women to convince other women that their career prospects have changed. Indeed, this is underlying goal of GIZ in developing the training program we implement in this project. But the training will be useful in any factory in the industry, and so individual firms may be unwilling to pay for it. And there may be further costs in the transition from one equilibrium to another, also borne by the firm. For example, if most males enter the factory with the hope of becoming a supervisor, they may be expected not to react well to a shift toward promoting women as supervisors.

This story is one of simple path dependence. If owners initially promoted men for any of the reasons above, then men entering the factory have an incentive to acquire the technical skills necessary to become supervisors, and women do not. And if the characteristics that make good supervisors are different for men and women, then managers may learn how to select the right men, but not how to select the right women.

There is also the possibility of discrimination in favor of men as supervisors. But while there is evidence of discrimination in other settings (see, for example, Bertrand and Mullanaithan 2004), we abstract from that possibility here because at this point we have little ability to test it directly. Note, however, that the outcome of a pure discrimination model and the path-dependent story we sketch above will look essentially the same.

III. Project Design and Data

The project was designed to provide training to 5 operators to become line supervisors. We began contacting potential factories, with a letter of introduction from a large mid-tier foreign buyer, in August 2012. The first training session began in November 2011. After six rounds of training, we stepped back in January 2013 to assess the design. The analysis here is based on the data from these first six training sessions. We discuss the second phase of the project, launched in February 2014, in the concluding section. Data from that part of the project are not yet available.

II.A. The training program

The training program on which the project is based was developed by GIZ in 2009. GIZ trained six Bangladeshi training centers to offer the program. For this project, they selected three of those training centers. The program is designed for operators on the sewing lines in woven / light knit factories. Trainees are expected to be experienced sewing machine operators, but are not expected to have experience working as a line supervisor. Training is intensive – six days per week over six weeks, 288 hours in all, covering three components: production planning, quality control, and leadership / social compliance. GIZ developed the program with the goal of increasing the number of women in management positions. This likely affected the mix of material included in the course, in particular the emphasis on leadership skills. However, the all parties agreed that the topics were appropriate for male trainees as well.

The focus of the GIZ program on women precluded training existing supervisors, because there are few women working as supervisors. That focus also meant that GIZ were not interested in training males. In the end, the selection of four females and one male from each factory represented a compromise between the research team – for whom more gender balance would have provided more statistical power in comparing female and male trainees – and GIZ – who preferred to train only females.

On joining the program, factories agreed to give all of the trainees a trial as a line supervisor after the completion of training. Factories were not asked to commit to promote all of the trainees, because we recognized that some trainees might prove to be inadequate as supervisors, and some might decide they did not want to be supervisors after training. As we discuss below, compliance with the agreement to give trainees a trail was good but not perfect.

II. B. Selection of factories

Our aim was to select a sample of factories capable of selling directly to large international buyers. We selected an initial target list using transaction-level trade data obtained from the Bangladesh National Bureau of Revenue. The data covered the period 2005 – 2010, and included information at the shipment level, identifying the seller, the buyer, the HS codes of the products shipped, and the weight and value of the shipment. We aggregated the trade data at the seller level. We then selected a sample of factories selling woven products directly to large buyers of a mid-range quality level. A screen of the data for the year 2010 – the latest available when the project started – yielded 665 factories meeting these criteria.

We began the process of enrolling factories in the program in August 2011. After contacting around 200 factories, we reached the target sample of 96 factories in mid-November 2011. The sample selection process and the comparison of the sample with the full spectrum of exporters in Bangladesh is detailed in Appendix A. In short, the sample is reasonably representative of large factories in the 3rd and 4th quintile of manufacturers ranked by unit values. Participating factories are slightly larger than non-participants in the selected pool, with faster growth between 2009 and 2010. They sell to slightly lower-end buyers than non-participants, measured by average unit values of the buyers calculated across all sellers.[5]

Participating factories were randomly placed into one of eight treatment rounds of 12 factories each. In practice we allowed factories to defer participation to a later round once, and in the end, several factories decided not to participate. By December 2012, when began training round 6, we had exhausted the initial list of 96 factories. Note that all of the comparisons we will make with trainees control for factory fixed effects, so we view the factory-level attrition issue as mainly one of external, but not internal, validity.

Table 1 shows characteristics of the factories participating in rounds 1-6. The factories are large – averaging 19 production lines and 2100 workers. Somewhat more than half of the employees in a typical factory work in the sewing section. The distributions are slightly right-skewed, with the median factory having 15 production lines, with 2000 workers in total, of which 59 percent are in the sewing section. A typical factory had been operating for 12 years. Given the rapid growth of the sector, this is very likely older than the industry average.

Around one in six factories reported have no female supervisors at all at baseline. In nearly half (45 percent) of the factories fewer than 5 percent of the supervisors were female.[6] Just over 10 percent of all supervisors are female. However, there are two outliers - one factory where all of the supervisors are female, and another where almost all are. Leaving aside these two factories, just under 8 percent of the existing supervisors were women. Given that the median factory has 32 supervisors, adding four female supervisors would, for all but a handful of factories, represent a radical change in the gender make-up of the supervisory workforce.

II.C. *Selection of trainees*

Our aim was to select from each factory four female and one male operators for training, and a valid comparison group against which to measure the trainees. The details of selecting workers evolved a bit across training

---

[5] For each buyer in the administrative data, we sum the total net mass and total FOB shipment value for all purchases. Then, for each factory (seller), we construct a distribution of buyers from the one with the lowest unit value (measured across all sellers) to the one with the highest unit value. The "90th percentile buyer" is the unit value of the buyer at the 90th percentile in this distribution.

[6] Most of these data come from the first follow-up survey, after training and promotion of some female trainees. We explicitly asked for the supervisor numbers by gender excluding any promoted trainees. To the extent that factories included the trainees in the numbers, the percentage of female supervisors at baseline would be even lower.

rounds, and we describe those in more detail in Appendix B. Generally, either we or the factories selected two women directly for training – according to the scores on a diagnostic exam in the case of the project team selections and by whatever criteria they wanted for the factory sections, subject only to the selected candidates passing a basic literacy test. Factories were asked to provide an additional six women and four men as nominees for training. We administered a diagnostic exam to these ten nominees, and ranked them according to their score. The diagnostic was intended to test for literacy – a requirement for the training – and numeracy. We also gave the potential trainees a short non-verbal reasoning test and asked them questions about aspirations to work as a line supervisor. Because women were sometimes forbidden to participate in the training by their families, we also asked the potential trainees if their families would allow / support them to attend training. Potential trainees were excluded if they did not pass the literacy test or said their families would not allow them to participate in the training.

After first removing the two females selected directly for training, we selected at random two of the top four female nominees and one of the top two male nominees. The random selection of trainees provides us with a control group against which we can measure certain outcomes. In initial discussions, management often reported that their reluctance to train / promote women was based on higher turnover rates for women. Since training and the prospect of promotion may affect turnover, the comparison of turnover rates for those selected and not selected or training is one particular comparison of interest. Note that with regard to the effect of training on productivity, comparison of trainees and controls is not likely to be viable, because those not selected are unlikely to be promoted.

There was a non-trivial amount of noncompliance. Over the six rounds, 50 workers assigned to training did not attend at all, and an additional eight attended for less than one full week. Factories most often reported that these workers either had decided they did not want to attend, or their families had said they could not attend. However, the family was most likely to intervene in the case of female trainees, while we note that the percentage of non-complying males assigned to training (21.2 percent) was higher than the percentage of noncomplying females assigned to training (15.2 percent).[7] These non-compliers were replaced by 40 workers receiving training even though they were not assigned to training – including 19 workers assigned as controls. Thus, non-compliance is a concern when we compare the outcomes of those assigned to treatment against the controls. We will use assignment to treatment (ITT) in these comparisons. However, for the analysis related to productivity, we will compare the females and males who actually receiving training.

Over the first six training rounds, 271 operators – 213 females and 58 males – received training. We exclude from this total eight workers who

[7] We interpret this as suggesting that factories cared more about which males received training than they did about which females received training, either because they did not plan to promote all of the females.

attended for five days or fewer. Conditional on attending at least one week, attendance was very high. Out of the 36 training days, males attended 34.4 days on average and female 34.5 days. All but two of the men attended at least four of the six training weeks, as did 96 percent of the women.

IV. Data

We use two types of data in the analysis. First, we conducted surveys at the factories at three points in time: A baseline survey shortly after the workers had been selected and before training began, a first follow-up survey around 4 months after training was completed, and a second follow-up survey around 10 months after training was completed. In each of the surveys, we aimed to interview each of those assigned to receive training, those assigned as controls, and non-compliers sent by the factory to receive training. Where any of these groups of workers were not in the factory at the time of the follow-up survey, we attempted to conduct short interviews by telephone.

In addition to those directly involved as trainees, non-compliers assigned to training, and controls, in each survey we surveyed an additional group of operators and supervisors. In the baseline, the additional respondents serve as random samples against which to compare those selected to receive training. In the later surveys, the additional respondents come mostly from lines where trainees work, and provide opinions on how the trainees as promoted supervisors compare to other supervisors.

In the baseline, we selected five additional operators and five existing line supervisors in pairs from five production lines. One of the additional operators and one existing supervisor was selected randomly from a lines where one of the trainees was working, and one operator-supervisor pair was selected a line where one of the three controls was working. The remaining three pairs were selected at random from those line without any treatment or control worker.[8] We also surveyed one of the factory's top managers – generally the HR Manager or the Production Manager – to obtain factory-level information on size, HR practices (e.g. bonuses and promotion) and manager attitudes.

Around four months after the end of the training session, we returned to the factory to conduct another survey. We surveyed each of the trainees and controls who were present in the factory at the time of the survey, along with eight randomly selected operators. We also surveyed five additional supervisors. These included the supervisor who was most recently promoted, but not one of our trainees, up to two supervisors from lines where the trainees were working as supervisors ("matched" supervisors), and other randomly selected supervisors. We also conducted a management simulation exercise, described in more detail below.

Key outcomes at follow-up are whether the trainees and controls are still working in the factory, whether they have been tried out as supervisors and

[8] Beginning in round five, we selected one of the random operator – supervisor pairs from the line with the newest appointed supervisor who was not a trainee.

whether they have been promoted to supervisor. We obtain this information from the trainees and controls themselves. But we also asked the HR Manager about each of the trainees and controls. Sometimes the information we obtain from the worker is different from the information we obtain from the manager, and we discuss these differences below.

A third and final survey was conducted around 10 months after training. Again we surveyed the trainees and controls, along with two operators from lines where any trainee was working as a supervisor. We also surveyed one of the other supervisors working on the line where trainees worked, and the supervisor identified at the most recently promoted non-trainee in the first follow-up survey. As in the first follow-up, we obtained from the HR Manager information on whether each of the trainees and controls still worked in the factory and her/his position.

In addition to the in-factory surveys, we conducted short follow-up surveys by phone at three points in time, focusing on those trainees and controls we were unable to reach in the factories. These surveys provide information on whether the worker remains in the factory (and was just absent on the day of the survey) and if not, what she / he is doing at the time.

The surveys, reports from the manager, and phone call surveys together result in very low levels of attrition for the most basic outcomes – whether the worker remains in the factory, whether she/he was tried out as a supervisor, and whether she/he was promoted as a supervisor.

Characteristics of trainees

The baseline data allow us to compare the trainees to three groups of workers – the pool from which the trainees were selected – those assigned to training and the controls – typical operators in the factory and typical supervisors in the factory. Table 1 shows these comparisons, first for basic demographic characteristic, ad then for attitudes expressed by the workers.

Comparing first the trainee pool with typical operators, we see that the trainees have 1.7 years more schooling, three-quarters of year more experience working in the garment sector and a half year longer tenure in the factory. Interestingly, they are also almost twice as likely to report having received training than typical operators, though only a small minority (14 percent) say they have received training. Compared with the typical existing supervisors, however, they have 1.3 years less schooling, 2.3 years less experience in the sector and year's less tenure in the factory. Hence, they are clearly positively selected with regard to the pool of operators, but the comparisons with existing supervisors are not so flattering. [9]

[9] The gap in education between existing supervisors and the males in the potential trainee pool is smaller, around 0.7 years, but still there. We can think of two reasons for differences between existing supervisors and the trainees. First, the factories may have selected individuals they thought needed the training in order to be good enough to be supervisors, or at least some factories may have selected individuals they thought they could spare for six weeks, rather than the best supervisor candidates.

We also asked respondents a series of questions about their attitudes and expectations. The attitudes were asked as agreement or disagreement with a statement (e.g., I find work stressful), allowing for four responses – strongly agree, agree, disagree and strongly disagree. Turning to the responses, we find, perhaps not surprisingly, that trainees are much more likely than the typical operator to say they want to be promoted to supervisor – though 14 percent of the trainee pool says at the time of the baseline survey that they are not interested in promotion. They also expect to work at the factory slightly longer than the random operator, but not as long as the existing supervisors. With regard to attitudes, there are two areas of difference. Existing supervisors are more likely than trainees, and random operators less likely, to say that their job requires they learn continuously. And existing supervisors report being the least stressed, while random operators are the most stressed, with the trainees intermediate between the two.

Panel B of the table compares male and female trainees. Here we see that females selected by the factory as potential trainees are younger by 1.5 years, have 0.7 years less schooling and almost 1.2 years less experience in the sector. They are also less likely to say they wanted to be promoted to supervisors (82 percent vs. 96 percent) and are less likely to strongly agree with the statement that they are treated with respect in the factory. When we exclude the two best females and compare the scores of male and female trainees on the diagnostic exam, we find that the males score about a third of a standard deviation higher than females, scoring higher on both the literacy and numeracy parts of the exam.[10]

B. Production data

A key innovation of this project is the gathering of very detailed production data for a large sample of factories. In the first training round, we asked factories for data in two week intervals every other month. After the second round we began collecting data for full calendar months, every other month. In a few cases, we have daily data over the entire study period.[11]

There are four main outcomes of interest: productivity, quality defects, hours the line operates, and absenteeism. By focusing on sewing, we are able to capture a measure of output which is very close the a pure quantity measure. A trained industrial engineer can take any garment and estimate the number of minutes a fully-efficient worker will take to produce the garment. These calculation come from summing the time required for each stich. The times come from a combination of international databases and in-factory time-and-motion studies. By multiplying these 'standard minute values' – SMVs (or standard allowable minutes – SAMs) by the number of units of a give garment which are

[10] We administered the diagnostic to the random operators and existing supervisors as well, but as the stakes were quite different for these two groups, we do not compare them with the trainees. We have conducted a much more extensive diagnostic comparing male and female trainees in the second phase of the project, and later drafts will incorporate these data.

[11] In a second project, we are collecting daily data over the entire study period for all factories.

produced, we obtain a measure of output – output minutes – which is highly comparable across products. For example, if a line producing 1000 shirts with an SMV of 15 minutes has production of 15,000 output minutes. For productivity, we divide the output minutes by input minutes – the sum of minutes worked by operators and helpers on the line over the same time period[12] – to obtain the industry standard measure of efficiency. This is s essentially a measure of Q/L:

Output * SMV / [(# Operators + # Helpers) * runtime in minutes]

The average efficiency in the sample we are currently using is 47 per cent, which is slightly higher than the 35-40 per cent that those in the industry typically quote. [13]

A second measure of interest is the number of quality defects. Factories typically report both the number or percentage of garments that require some re-work and the number or percentage that must be rejected. Reject rates are typically very low, around 1 per cent. Rework rates are much higher, averaging around 10 per cent (with a median of almost 8 per cent). Because the re-work time is included in the measure of "input minutes", the efficiency measure incorporates improvements in quality.

Finally, we measure hours worked and absenteeism for operators and helpers working on the line. For some factories, we receive reports only for operators, and for some the two combined. Since operators typically represent more than three-quarters of the line workers, this difference is not overly concerning.

V. Initial outcomes: migration and promotion

The first outcomes of interest are whether the workers remain in the factory four and 10 months after training, and whether they have received trials as supervisors and been promoted at either of these points in time. The retention rates are interesting because one reason often give for the reluctance to promote females to supervisory positions is that worker exit rates are higher for females.

Table 3 shows the percentage of trainees remaining in the factory at four and 10 months. We report results first for the females and males assigned to training, compared with those assigned to control. The raw outcomes are shown in the rows labeled 1 (females) and 2 (males). We see that those assigned to training and those assigned to control are almost equally likely to remain in their

---

[12] We could improve this measure by a step if we had the wage bill for the line. However, the industry typically uses three grades for operators, and we most often know only the total number of operators, not the number by grade.

[13] The higher efficiency in our sample may come from having a more efficient sample of factories. However, the data across factories are not always comparable because the international SMV values are often adjusted upward by factories to account for some expected level of inefficiency. We are currently working to ensure the data are comparable across factories, but we include factory fixed effects in all of the regressions using production data, which will absorb systematic measurement differences across factories.

original factory at both four and 10 months. However, these averages mask larger differences related to non-compliance. We show these differences in the rows marked 1A/B and 2A/B. Fewer than half (45.5 percent) of the females assigned to but not attending training remain in the factory after 10 months, compared with 72 percent of those assigned to and attending training. A similar pattern holds for men, though the gap is somewhat smaller (76.9 percent vs. 63.6 percent). On the other hand, compared to complying treatment group (row 1A, assigned / attending training), women who attended training as non-compliers (row 2A) are more likely to remain in the factory; the similar comparison for men shows no difference in retention rates.

Factories agreed ex ante to try out all of the trainees as supervisors. Compliance on this was reasonably high, with 80 percent of trainees reporting having been tried out as a supervisors by 10 months following training.[14] There is some difference by gender, with males being more than 10 percentage points more likely to report receiving a trial. As with retention in the factory, we find some difference among the complying and non-complying groups. Those not assigned to training but actually attending (row 2A) are most likely to be tried out, an outcome that suggests they may have been some maneuvering by factories to train certain workers. Those assigned to training and not attending (row 1B) are least likely to have been trialed. Among the complying controls (row 2B), more than half of the males were given a trial by the second follow-up survey. Overall, among those attending training, 93 percent of males and 76 percent of women report having been tried out as a supervisor by ten months after training. The difference between these two is highly statistically significant.

Panel C of Table 3 shows the promotion rates for the some assignment / compliance groups. The differences between female and male trainees become starker. After ten months, 77 percent of the males assigned to training had been promoted to supervisor, compared with 50 percent of females (row 1). Very few of the females not attending training – whether assigned to training or not – were promoted, while around a quarter of the males were promoted without attending the training. The highest promotion rates – 88 percent for men and 53 percent for women – are among the group assigned to and attending training.

In Table 4, we show results of regressions detailing the effect of training on trials and promotions for female and male trainees. Having examined the ITT and non-compliance outcomes in some detail on Table 3, we switch to an assessment of the effect of training on males and females. Regardless of how the trainees were selected, we focus on the comparison of female trainees and male trainees. We run a regression with a specification:

---

[14] At four months, we asked both the trainees themselves and also managers whether the trainees had been tried out. Managers were much more likely to report trials (72 percent of trainees) than were the workers themselves (52 percent). The difference may owe to different views about what constitutes a trial. Or it may be that either side is mis-reporting – workers because they don't want to admit they were tried but not promoted and managers because they want to report compliance with what they agreed to ex ante.

$$Y_{if} = b_0 + b_1 T_i + b_2 S_i + b_3 T_i * S_i + b_4 X_i + h_f + e_{if} \quad (1)$$

where Y is the outcome of interest for individual *i* in factory *f,* T is an indicator that individual i attended training, S indicates the individual is a female (=1, male=0), T*S indicates a female attending training, and X is a vector of individual characteristics which are included in some of the regressions. The regressions all include factory fixed effects. We are interested primarily in the difference between female and male trainees, which is the sum of b_2 and b_3. We show the p-value on the F test for the sum of these two coefficients at the bottom of Table 4. The sample for the regression is all of those assigned to training (ITT=1) plus those attending without being assigned.

Panel A of Table 4 shows results at 4 months. The first two columns explore patterns of attrition from the sample. We define attrition as not being able to determine whether the individual received a trial as a supervisor or not, using the combination of the surveys, manager reports and follow-up phone calls. By this definition, we had attrition of 11 percent at four months and 10 percent at 10 months. The first row of Panel A shows that the attrition of male trainees was equal to the attrition rate of other males it the sample – controls plus non-complying treatments. The sum of rows 2 and 3 (-0.016, not significant) shows that female trainees also had attrition rates comparable to the sample of male non-trainees. In the second column, we add a series of individual characteristics. The effect of training on attrition is not changed, and only tenure in the factory is significantly (and negatively) associated with attrition.

Columns 3 and 4 report results at four months for whether the worker was still in the factory. With or without controls, males trainees are more likely to remain in the factory than the male comparison group, but not significantly more likely to remain than female trainees. The gender gap opens when we examine whether individuals received a trial or not, columns 5 and 6. First note that training has a very large positive effect on receiving a trial for both men and women. However, the effect is significantly larger for male trainees, with an estimated effect of 18.5 percentage points. The gap is somewhat smaller when we control for individual characteristics (column 6), though it remains large (around 15 percentage points) and significant. The slight shrinking of the gap suggests that differences in observable characteristics account for some of the differences in trial rates. In particular, those expressing at baseline that they wanted to be a supervisor are more likely to receive trials, and the female trainees were less likely to report that they wanted to be promoted. (See Table 2.) The last two columns show a similar pattern with regard to promotion at four months, with female trainees about 21 percentage points less likely to be promoted without controls and 19 percentage points less likely with the controls. Both gaps are highly significant.

Panel B repeats the analysis using reports 10 months after training. Again we see small and insignificant differences between male and female trainees with regard to attrition from the sample and exit from the factory (columns 1-4). The gap in the rate of being tried out decrease slightly to around 17 points without controls and 11 points with controls. We see that a higher diagnostic score is now also associated with a greater likelihood of receiving a trial (column 6). However, the gap in promotion rates is somewhat larger that it was at four months – 27 percentage points without controls and 18 percentage points with controls.

If we take the factories' willingness to promote the female trainees as an indicator of their expected effectiveness as supervisors, the results presented on Tables 3 and 4 are clearly a glass half-full / half-empty story. From the 'half-empty' perspective, females are significantly less likely to receive trials and be promoted. But from the 'half-full' perspective, the average factory promotes almost two female trainees to supervisor. Given that the average factory had about five female supervisors (and the median factory just one – see Table 1), this represents a significant increase in the percentage of females working as supervisors in the sample of factories.

V.B The effect of training on exit, and attitudes of workers on the line

With regard to the multiple equilibrium story, we are interested in how training affects attrition rates. We are also interested in whether seeing female supervisors being promoted affects the reported expectations about tenure in the factory. With regard to the former, Table 3 shows that being assigned to training reduces exit rates slightly (but not significantly) for males, and has no effect on females. Table 4 shows that actually attending training does reduce attrition for both males and (to a slightly less extent) females.

More interestingly, we find differences between the expected tenure of random operators working under a female supervisor and random operators working under only male supervisors. The effect is small an positive for female operators – the proportion expecting to stay five years or longer increases from 39 percent to 45 percent, but very large and negative for male operators. More than two-thirds (68 percent) of male operators working for male supervisors say they expect to remain at least five years, while only 38 percent of those working under a female supervisor say the same. These differences are consistent with a perceived shift in probabilities of promotion among males.[15]

VI: Measuring effectiveness of the trainees

The standard way of assessing the effect of a training program is to assume that wages reflect the value of the marginal product of labor, and compare changes in wages before and after training. For many reasons, we don't think that is how the companies would assess the program. Wage rates in the sector depend very heavily on worker grade, so that entry-level supervisors are likely to earn the same amount regardless of their productivity, and to have the same increase in salary regardless of their productivity either as an operator or a supervisor. So we will aim to assess the effectiveness of training by examining

[15] We might expect that news of the promotions would spread to other lines. However, we also ask operators whether they agree or not (scale of 1-4) with the statement: "This factory has promoted able, hard-working women." We find that those working under female supervisors are significantly more likely to strongly agree with this statement.

actual productivity of the workers. Nevertheless, because a comparison of wages in the norm for the literature, we have also carried out that exercise.

We asked both the worker and the HR director at the time of the second follow-up what the worker's salary was. Among controls who did not attend training, the average wage was 5279 BDT per month on average. Among all those attending training, the average was 7322 BDT per month. The latter average is composed of an average of 9025 BDT for those promoted to supervisor and 4941 BDT for those still working as an operator. The direct costs of training were about 40,000 BDT per participant, and the opportunity cost (measured by the worker's wage costs) was about 8000 BDT during the six weeks of training. Thus, the 48,000 BDT cost of training is returned in 24 months on average (7322 minus 5279 as the increase in productivity) or 12 months among those promoted (using 9025 minus 5279 as the increase in productivity). Either of these suggests a reasonable rate of return to the training.

As we noted, we are skeptical that this exercise tells us much about the real effects of the training. In addition to the concerns described above, suspect that factories may be induced to promote the females trainees for reasons other than efficiency, and may fail to promote females even if they are more efficient. We now turn to examining the relative effectiveness of female and male trainees from several perspectives. Before looking at the data for the trainees, we examine generic attitudes about female and male supervisors, as expressed by operators at the time of the first follow-up survey.

We asked the randomly selected operators to compare, generically, the effectiveness of female and male supervisors with regard to a series of traits. For example, we asked whether male or female supervisor generally were "Better at solving technical problems", "Better at resolving conflicts between workers", and other similar characteristics. We find that operators are more likely to say that male supervisors are better for all of the traits. Figure 1 shows the percentage of operators saying that female supervisors are better in each of six areas. We divide the sample of operators into two groups. The left side of the figure are the answers to operators who tell us they have never worked for a female supervisor, while the right side are the responses from operators who report that they have. Note that while a small minority of supervisors are female, operators move across lines with some frequency. Hence, almost half the sample reports having worked for a female supervisor. We see a sharp difference in responses, with operators who report having worked for a female supervisor viewing females more favorably in each of the six areas. All of the differences between the two groups of operators are significant at the .05 level or better except for resolving conflicts with workers. This suggest that attitudes toward female supervisors may be shaped by experience, or more properly, a lack of experience, working with them.

How do the trainees perform on the job? Ideally, we would answer that question both by observing management styles and by measuring performance. We face particular challenges with regard to management style, because there is no reasonable way for us to observe the trainees working as supervisors on the factory floor. While we have quite detailed productivity data, the trainees are typically one of three supervisors on a line. Hence, their effect is likely to be more difficult to measure.[16] Moreover, managers tell us that new supervisors take four to six moths to reach full effectiveness, and given lags in promoting after training, we often do not have data to follow the trainees for that length of time.

VI.A. Management simulation exercises

With these challenges in mind, and to gain some insight into the relative effectiveness of male and female trainees as supervisors, we conducted a management simulation exercise as a part of the first follow-up survey. In each factory, eight randomly selected operators were placed into four teams of two each. The teams each played two "production" games, one involving Legos and one involving buttons. We randomized the order in which the games were played at the factory level. Each team was assigned a leader whose job it was to explain the particular exercise and manage the operators as they performed their tasks. The team leader was either a trainee, an operator from the control group, the most recently promoted supervisor who was not a trainee, another supervisor from the same production line as one of the trainees selected at random (a "matched" supervisor), or another supervisor selected at random (a "random" supervisor). Each pair of operators played the production game twice, once with Legos and once with buttons. Each team leader played only one session – either Legos or button – so there were eight team leaders in each factory, and each pair of workers played with two different team leaders.

The production games are described in more detail in Appendix C. For each of the Lego and button exercises, the teams played five separate sessions. The first was a simple sorting exercise in each case, sorting either buttons or Legos by color. For Legos, the second, third and fourth sessions involved constructing chains of Legos with a particular color pattern – blue, yellow, green, blue, yellow, green, etc. The three games were differentiated by their payoffs: the first summed the length of the chains produced by the two operators, the second paid based on the length of the longest chain produced by either worker, and the third paid based on the shortest chain produced by either operator. We measured production of the teams according to the payoff function, and will compare the performance of teams led by female trainees with that of teams led by male trainees.

We combine each of the five individual games into a single regression by standardizing the payoffs at the level of the game / round. We then run regressions with the standardized payoffs on the left-hand side and a set of controls for characteristics of the team leader on the right hand side. We focus the discussion here on the subset of games where trainees are team leaders, comparing the performance of female and male team leaders.

[16] They are also not randomly placed on lines. However, we find no significant differences in the measured characteristics of lines where female and male trainees work. While the non-random placement is a concern, we do not believe it seriously undermines the validity of the exercise.

Note that the game (Legos or buttons) was randomly assigned, but the assignment of the team leader to the first or second session was not random. Because factories anticipated that we wanted to talk with trainees, the trainees were more likely to be assigned to the first session, and so the existing supervisors were more likely to be assigned to the second. Logistical complexities working in the factory prevented us from randomizing the session in which any team leader participated. This matters, because even controlling for the team leader and game (Lego vs. buttons) types, operators were significantly more productive during the second session. This is logical because we expect some learning by the operators from the first to the second session – even though they paly different games in each session. We control for the session order effects in regressions.

Table 5 shows results of these regressions. The specification in column 1 includes controls for factory, session (first or second) and game fixed effects. We find that teams led by female trainees have payoffs which are 0.29 standard deviations higher than teams led by male trainees, a difference which is highly significant. In other words, female trainees appear to be more effective as team leaders than male trainees. Column 2 adds team leader demographics – age, education, industry experience and factory tenure – and Column 3 adds operator team fixed effects. Note that the third regression then isolates the cases where a single team was led by both a male and a female trainee. Only 19 teams had this pair of team leaders, so while the table shows a sample size of 600, the effective size is much smaller. Nevertheless, the additional production by female-led teams remains almost unchanged, at just over 0.30.

Columns 4 and 5 explore whether the gender composition of the operator team interacts with the gender of the team leader. Since only 20 percent of operators are male, all-male teams are very rare. Instead, we compare the performance of mixed teams with those of all-female teams. The benefit of female leadership is significant only for the all-female teams (column 4). For mixed teams, female leaders have measured higher but statistically insignificant effects on production.

Finally, in the last two columns of the table, we examine whether those tried out as supervisors and those promoted to supervisor perform better than the converse groups. We find that those promoted to supervisor perform significantly better than those not promoted. Since promotion is not random, we are unable to say whether this is due entirely to selection – more able trainees are promoted, while less able trainees are not – or whether the experience as a supervisor also makes the individual more effective as a leader. To the extent that experience matters, the regressions may understate the true advantage of female trainees, since they have much lower promotion rates than male trainees.

In sum, then, the female trainees were significantly more effective in generating payoffs than were the male trainees. The females perform best when they are matched with a pair of female operators, and trainees who were promoted before the time of the first follow-up survey also perform significantly better than those not promoted.

After the second session, the operators on the production team were asked to compare the management style of the two team leaders they worked with. They were asked whether the first or second team leader they worked with was better at explaining the game, better at answering questions, better at motivating them, always pressuring them, and so forth. Looking at the responses of the 19 teams led by both a female and a male trainee, we find that operators are more likely to say that the *male* trainees were better at answering question, at motivating, et encouraging. Female trainees were selected more often only as "always pressuring." We find these responses surprising given the superior performance of female trainees in the exercise.

VI. B. Line-level productivity

The management simulation exercises provide some evidence of effectiveness as a leader in a controlled environment and where the supervisory effort can be mapped directly to the individual. Of course, conditions and performance on the production line may differ. To measure performance of the promoted trainees as supervisors, we aimed to collect very detailed line-level production data from each of the participating factories. In practice, both the quality of the data available and the level of cooperation of the production staff in providing the data varied. Hence, while we trained workers in 58 factories, we do not have useable production data from all 58. We examine four outcomes – efficiency of the line, defect / rework rates, absenteeism and work hours. The sample for each of these outcomes varies because we sometimes have, for example, very good absenteeism data but not all of the data needed to calculate the line efficiency.

As above, we focus on the comparison of male and female trainees. The placement of trainees on lines is not random, and we lack a valid instrument for placement. In Appendix D, we show results which suggest there are no clear patterns of placement for either the male or female supervisors. Nevertheless, to address at least some of the placement concerns, we focus first on line fixed-effect regressions. The fixed effect will capture the average productivity of the line when the trainee is not working on the line. A dummy variable indicating the trainee worked as a supervisor on the line on a given day will then pick up the change in productivity, averaged over the days the supervisor was on the line. The non-random assignment will then matter only if there are differential trends in the lines, or other changes that coincide with the placement of the worker.

Table 6 shows outcomes for the female and male trainees. The first four columns use a line fixed effect specification, and columns 5 through 8 use an ANCOVA specification. Several comments are in order. First, in in the line fixed effects regressions we use data from lines even where there is never a promoted trainee on the line. The reason for this is that in addition to line fixed effects, we include day fixed effects to control for seasonality and other shocks. The full set of lines will affect the day fixed effect estimates. Second, we show at the bottom of the table the number of lines where there is ever a female or male trainee. The numbers are modest, particularly for absenteeism. They are also slightly lower in the ANCOVA specification because that specification controls for the Y value on the line during the period to time prior to and during training. We lack these data for some factories. Finally, the table also shows at the bottom the p-value for the test that the female and male coefficients are different from one another.

Turning to the results, using line fixed effects, we find that efficiency decreases slightly and hours worked increases slightly after the female trainees are promoted on the line. Defect rates fall slightly on lines where males are promoted. However, the only significant difference between male and female trainees is in the daily average hours the line operates. For the other outcomes, we are unable to reject the null that the two coefficients are the same at any reasonable level of confidence. Hence, while there is a suggestion in the data that the male trainees outperform the female trainees in some dimensions, statistically we conclude that the two differ only in the hours worked on the line.

VII: Resistance?

The combination of higher production by female trainees and more favorable reporting about male trainees by operators in the management simulation exercises raises the question of how operators perceive both female supervisors generally, and the trainees in particular. In the second follow-up survey, we conducted an exercise which designed to allow respondents to express attitudes toward other workers in an anonymous manner. Respondents were given a cup with a large and equal number of red and green buttons. They were told to reach into the cup, pull out five buttons and – without showing the buttons to the enumerator – report how many green and red buttons they had selected. For each green button, they received 20 BDT themselves. For each red button, the gave 20 BDT to a specifically identified worker type.

We refer to the game as the ‘cheating game’ because we expect – and find – that on average the respondents report a higher number of green buttons than chance would allow. Since the cup contained an equal number of red and green buttons, we should on average find 2.5 green and 2.5 red buttons. But in fact, the respondents reported an average of 3.2 green and 1.8 red buttons. The anonymity comes in because even we cannot tell whether an single draw was honest or not. By chance, some workers will have drawn 4 or 5 red buttons. But we can be very sure there was cheating overall, because given the number of times the game was played, the odds of getting 3.2 green buttons on average is infinitesimally small.

For operators, we played the game three times, with the recipient changing from another operator to one line supervisor and then another line supervisor. For line supervisors, we also played three times, with the recipients being an operator, the other line supervisor and the line chief. We are particularly interested in whether operators responded differently when giving to trainees relative to other supervisors, and when giving to female trainees relative to male trainees. And we are interested in how female and male trainees responded when giving to operators, other line supervisors, and the line chief.

The data on Table 7A show that female trainees gave less to operators than either existing male supervisors or male trainees did. The data on the second panel show that other supervisors gave existing females more, but did not give female trainees more, both compared with existing male supervisors. These results are suggestive of tension between the female trainees on the one hand and operators and other supervisors on the other.

VIII Conclusions

We work with local training companies to provide supervisor training to four female and one male sewing machine operators in almost 60 ready-made garment factories in Bangladesh. We find:

- retention rates are slightly lower for females assigned to training than males assigned to training
- promotion rates are higher for male trainees than for female trainees. Nevertheless, the training resulted in 2 new female supervisors in each factory, a large number relative to the baseline
- Female trainees significantly outperform male trainees in a post-training management simulation exercise
- There are no significant differences between male and female trainees with regard to line-level efficiency, absenteeism or quality defects. Female trainees perform insignificantly better for efficiency and absenteeism, and insignificantly worse on quality defects.
- We find hints of resistance by operators to the female trainees. In spite of performing better in the management simulation exercises, female trainees fare worse than male trainees in the opinions of the production teams. Female trainees give less to their operators in the 'cheating games'

The data suggest why it is difficult for factories to transition from the practice of promoting only males as supervisors to one of promoting females and males. Attrition rates appear to be affected only modestly by training, but they are high enough that the factory' return on training would be affected. There is an indication of a negative reaction by male operators to the change, perhaps not surprising given that the overwhelming majority of male operators want to be promoted to supervisor, and the new policy reduces the chance of that occurring.

In February 2014, we began a second phase of the project. Reflecting what we learned in the first phase, we made several adjustments to the design. First, we are training a more equal number of males and females. Second, we allowed the factories to select all of the trainees. We also allowed the factories to nominate as few as four and as many as a dozen trainees, and trained them in two waves separated by several months. The last changes were all designed to increase promotion rates by better matching supply with demand for new supervisors and by allowing the factories to train those they wanted to train. We also designed a much more extensive skills diagnostic, which was administered before training, after training and after the six to eight week trial period in the factory. Finally, we are collecting from each factory daily production data. All of these changes should give us significantly more statistical power to detect changes in line level productivity, once the data are available to incorporate in the analysis.

References

Acemoglu, Daron and J.S. Pischke, 1998. "Why do firms train? Theory and evidence," *Quarterly Journal of Economics* Vol. 113, pp. 79–119

Backus, Matthew, 2014, "Why is Productivity Correlated with Competition?", working paper, Cornell University.

Bertrand, Marianne, and Sendhil Mullainathan, 2004, "Are Emily and Greg more Employable than Nikisha and Jamal? A Field Experiment on Labor Market Discrimination," *American Economic Review*, Vol. 94(4), pp. 991-1013.

Bloom, Nicholas and John Van Reenen. 2007. "Measuring and Explaining Management Practices Across Firms and Countries," *Quarterly Journal of Economics*, Vol. 122(4), pp. 1351- 1408.

Bloom, Nicholas, Raffaela Sadun and John Van Reenen, 2012. "Management Practices Across Firms and Countries," NBER working paper 17850.

Center for American Progress, Global Wage Trends for Apparel Workers, 2001-2011, Washington, D.C.: Center for American Progress.

Foster, Lucia, John Haltiwanger, and Chad Syverson. 2008. "Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability?" *American Economic Review*, Vol 98(1), pp. 394-425.

Gibbons, Robert and Rebecca Henderson, 2012. "Relational Contracts and Organizational Capabilities," *Organization Science*, Vol 23(5), pp. 1350-1364.

Hsieh, Chang-tai and Peter Klenow, 2009. "Misallocation and Manufacturing TFP in China and India," *Quarterly Journal of Economics*, Vol 124(4), pp. 1403-48.

Ichniowski, Casey, Kathryn Shaw and Giovanna Prennushi, 1997, "The Effects of Human Resource Management Practices on Productivity," *The American Economic Review* Vol. 87, No. 3 (Jun., 1997), pp. 291-313.

Syverson, Chad, 2004. "Market Structure and Productivity," *Journal of Political Economy*, Vol. 112(6), pp. 1181-1222.

Syverson, Chad, 2011. "What Determines Productivity," *Journal of Economic Literature*, Vol 49(2), pp. 326-365.

Appendix A: Sample Selection and representativeness

We obtained an initial sample using data from transaction-level trade data obtained from the Bangladeshi National Bureau of Revenue. These data provide volume (net weight) and value of exports at the shipment level. The data have identifiers which allow data from individual exporters to be aggregated. We aggregated data by exporter and calculated the unit value (USD per kilogram) for each export / product / year. We also summed total exports by exporter. Using these two measures, we selected a sample of firms with annual shipment volumes large enough to sell directly to large foreign buyers, with unit values in the range of mid-level buyers. This selection process yielded an initial sample of 665 exporters. We then selected the group of (20) suppliers to one particular mid-range buyer based in the UK. We created a score which indicated how similar each of the 645 exporters not supplying this buyer were to the 20 exporters supply this exporter. We selected around 400 exporters, and searched local directories and the internet for contact information. This yielded a sample of 230 factories, which we began to contact in August 2011.

By November 2011, we had received an initial commitment to participate in the project from 96 units of 85 distinct factories. Table A-1 shows how the characteristics of the 85 factories differ from the initial list. The table shows both a comparison characteristic by characteristic, and the p-values from a probit regression including several of the characteristics. We find that that those factories agreeing to participate sell to more buyers, and sell to higher-end buyers. The quality of buyers is measured by the average unit price paid by each buyer. For each seller, we then order the buyers by unit price, and measure the unit value paid by the buyer at the 90[th] percentile in the ranking. We also find some evidence that the participating factories had higher rates of recent growth and export products to a larger number of countries.

Table A.1: Take-Up of the Program

	Signed-Up *N = 85*	Not Signed-Up *N = 145*	p-value	p-value probit
Size (Export, 1000 Kgs)	830.4	683.8	0.11	0.44
Average Unit Value (per Kg)	925.9	883.8	0.15	**0.01**
*Growth (Sales 2010 to 2009)**	1.89	1.46	**0.08**	--
Number of Destinations	10.1	8.3	**0.09**	0.18
Number of Buyers	9.75	8.3	**0.06**	**0.02**
Number of Products	3.01	2.91	0.32	0.31
Main Product in Woven	0.59	0.54	0.26	--
Year of first export	2006	2006.2	0.2	--
Median Buyer	560	631	0.18	0.44
90th Pcntile Buyer	183.6	283	**0.03**	**0.01**

* On a sample of 80 and 135 exporters respectively

During the second round of the program, discussions with the local office of the International Finance Corporation led to inclusion of seven factories located in the Dhaka EPZ in the project. These factories were added in training rounds 4 and 5.

We allocated factories randomly to one of eight training rounds containing 12 factories each. Training for the firs round began in November 2011. We contacted each factory about six weeks before its selected training to arrange for selection of trainees and scheduling of the baseline survey. It was not unusual for factories to tell us that they were not able to participate at the selected time because of production pressures. We initially drew a hard line on this, and told factories that the selected round was the only one they would be eligible to participate in. However, we quickly realized the need for some flexibility, and decided to allow factories to defer once before dropping them from the program.

By the sixth training round, we had re-contacted all 96 units on the initial sample list, as well as the seven EPZ factories. Given the addition of the EPZ factories, this implies an additional attrition (before starting training) of around 40 per cent.[17] After the sixth training round, we decided to suspend the training temporarily. Having already gathered a substantial amount of data and information, we felt we would gain by analyzing those data and perhaps tweaking the design for the remaining factories. We resumed the training in February 2014 and will complete the follow-up surveys for the additional eight training sessions in November 2014. Data from the training rounds conducted in 2014 will be included in a future version of the paper.

Appendix B: Selecting samples of workers

The process of selecting workers to participate in training evolved across training rounds as we learned more about the realities of working with large, very busy factories. In all of the training rounds, we conducted a diagnostic exam to measure basic literacy, numerousy, and knowledge of machines and processes. Potential trainees not passing the literacy exam were excluded from the project, because the training involved written materials. We ranked the remaining pool of potential trainees according to their overall score on the diagnostic, and used the ranking to select trainees in the manner described below.

In the first four training rounds, we asked each participating factory to provide 16 females and four males as potential trainees. The two females with the highest diagnostic scores were included in the training pool. From among those with the third to sixth highest scores (the next four on the list), we then selected two at random to participate in the training. From among the two highest scoring males, we selected one at random to participate in training. This process generates a sample of two females and one male who were randomized out of training, and who serve as a comparison group for certain key outcomes.

[17] Around 15 of the 43 dropouts were first offered the chance to participate in round 6, and decline because December is in the middle of the peak production period in the industry.

Beginning in round five, we modified the process somewhat. First, we allowed the factory to select two females they wanted to train directly, subject only to their passing the literacy test. We then selected two additional females from among the four females with the highest diagnostic score once the two directly selected into training were excluded. We also reduced the number of female candidates we asked the factory to provide from 16 to eight. Given the four of the eight would be trained (instead of four of 16), we expected the factories to chose potential candidates more carefully.

Appendix C: Description of the management simulation exercises

During the first follow-up survey we conducted a management simulation exercise. Each exercise involved four teams of two operators who played two "production games", one involving exercises with Legos and the other with buttons. The order of the games was randomized at the factory level. Each session of the games was led by a team leader, whose job it was to explain the game and supervise production. The identity of the team leader changed from the first session to the second. That is, each pair of operators played the Lego game with one team leader and the buttons game with a different team leader. We analyze how the outcomes of the games depend on the identity of the team leader, and in particular whether the teams are more efficient when they are supervised by female trainees than when they are supervised by male trainees.

In each Lego or buttons session, five rounds of games were played. Team leaders were incentivized by being promised payouts based on the actual outcome of one of the rounds, selected at random at the end of the session. The formula for the payoffs in each round varied in ways we describe below.

The first four rounds of the Lego game involved building chains of Legos with alternating colors. The fifth round involved constructing an object. Division of labor was efficient in constructing the object in round 5, and we noted whether the leader organized the two workers each to construct a different part of the structure, or whether each worker constructed entire structures. In the first round, the payoff was based on the sum of the number of correct chains produced by the two operators. Mistakes were ignored. The second round payoff was similar, but the payoff was reduced by twice the number of mistakes made by the operators. In the third and fourth rounds, the team leader was asked to have the operators produce slightly different chains, one longer than the other. The payoff for the third was based only on the single operator who produced the largest number of chains. We wanted to see if the team leader would assign the more agile operator to the easier task. In the forth round, the payoff was based on the smaller output of the two operators. Again, the question was whether the team leader would assign the faster worker to the harder task given that payoff. Finally, in the fifth round the payoff was based on the number of structures produced, and a key question is whether the operators were organized as a production line or worked independently.

The payoffs in the buttons games were designed to parallel those in the Lego games. The button games involved sorting buttons of five different colors into five cups. As with Legos, the first round payoff was based only on the number of correctly sorted buttons, the second round deducted for mistakes, the third round payoff was based only on the operator who sorted the most, and the fourth round payoff only on the operator who sorted the least. In the fifth round, the operators taped buttons onto sheets, and again the exercise was performed more efficiently when they worked as a production line.

Appendix D: Placement of trainees on lines

The lines on which the trainees were placed when promoted was not selected randomly. We did not ask the factories to do this, as we wanted the trainees to be placed on the lines the factory felt were most appropriate for them. When we asked about how the management selected the lines on which trainees were promoted, the overwhelming response, given in more than 95 percent of the cases, was that the trainee was promoted on a line where a supervisor had left.

In Table D.1 we show the results of an analysis of the characteristics of lines on which trainees were placed. Given the results on Table D.1 and he responses of management, we do not believe that endogeneity of selection – especially conditional on line fixed effects – is a major concern when comparing the outcomes of female trainees with the outcomes of male trainees.

Table D.1: Comparing lines on which will be placed male vs. female trainees, in the 60 days before the placement
Only first placement on a line is considered, lines that get both male and female trainees are excluded

VARIABLES	(1) Efficiency	(2) Sh. Defective Output	(3) Sh. Absent Worker	(4) Daily Hours	(5) SMV	(6) Nhr. Worker
Female trainee	-0.028	0.021	0.017*	-1.049	0.645	-5.981
	(0.02)	(0.02)	(0.01)	(0.67)	(1.22)	(4.26)
Observations	1,502	1,099	458	1,489	1,667	1,481
R-squared	0.542	0.791	0.674	0.824	0.801	0.915
Factory FE	YES	YES	YES	YES	YES	YES
Date FE	YES	YES	YES	YES	YES	YES

Robust standard errors, clustered on the production line level

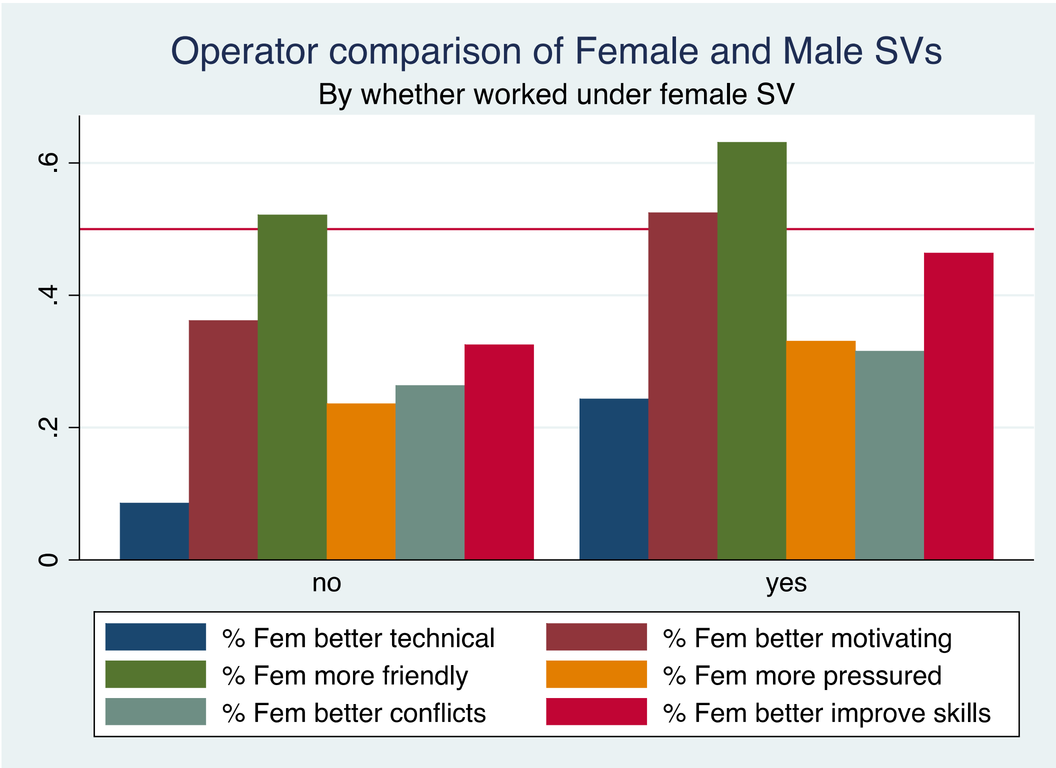
Operator comparison of Female and Male SVs
By whether worked under female SV
0
.2
.4
.6
no
yes
% Fem better technical
% Fem better motivating
% Fem more friendly
% Fem more pressured
% Fem better conflicts
% Fem better improve skills

Figure 1

Table 1: Factory-Level Summary Statistics

	Mean	Median
Number of sewing lines	19	14
Number of employees, total	2116	2000
Number of employees, Sewing	1171	1000
Operators per sewing line	48	47
Number of sewing supervisors	48	36
Percentage female supervisors	10.8%	5.6%
Percent conducting training	68.1%	NA
Percent training outside factory	8.9%	NA
Year factory established	1999	2001

Table 2: Demographic Characteristics

Panel A: Trainees vs. Operators and Supervisors

	Mean			**Comparisons**	
	Trainee Pool N = 576	Operators N = 287	Supervisors N = 292	Trainee vs. Operators	Trainee vs. Supervisors
Gender (female = 1)	0.72	0.75	0.13	-0.03	0.60***
Age	23.7	23.6	28.2	0.06	4.55***
Migrant	0.72	0.75	0.85	0.03	0.13***
Married	0.59	0.64	0.75	-0.05	0.16***
Have Children	0.40	0.49	N/A	0.09**	N/A
Education (= years in school)	8.16	6.46	9.43	1.70***	1.28***
Experience in Garments	5.98	5.21	8.30	0.76***	2.32***
Tenure in Factory	3.11	2.62	4.14	0.49***	1.04***
Ever Received Training	0.14	0.08	0.15	0.06**	-0.01
Would like promotion	0.86	0.58	0.97	0.28***	0.11***
Expected time in factory	5.23	4.80	6.32	0.42	-1.09**
Overall job satisfaction	3.49	3.52	3.61	0.02	-0.11***
A/D: The job requires I work fast	3.10	3.03	3.21	-0.08	-0.10
A/D: Job makes me learn new things	3.05	2.88	3.31	0.17**	-0.26***
A/D: At work I am treated with respect	3.71	3.67	3.78	0.04	0.07*
A/D: I find work stressful	1.50	1.64	1.37	-0.13*	0.13**

Panel B: Female Trainees vs. Male Trainees

	Mean		
	Females N = 416	Males N = 160	Feamles vs. Males
Gender (female = 1)	1.00	0.00	NA
Age	23.3	24.8	1.51***
Migrant	0.71	0.75	-0.04
Married	0.59	0.58	0.01
Have Children	0.43	0.29	0.14***
Education (= years in school)	7.97	8.65	0.68***
Experience in Garments	5.65	6.83	1.17***
Tenure in Factory	3.05	3.27	-0.22
Ever Received Training	0.14	0.15	-0.01
Would like promotion	0.82	0.96	0.14***
Expected time in factory	5.14	5.44	-0.30
Overall job satisfaction	3.49	3.51	-0.03
A/D: The job requires I work fast	3.11	3.08	-0.03
A/D: Job makes me learn new things	3.07	2.99	-0.08
A/D: At work I am treated with respect	3.68	3.81	0.13***
A/D: I find work stressful	1.53	1.41	0.13

Table 3: Initial Outcomes

A: Still Working in Factory

		Females		Males	
		Four months	Ten months	Four months	Ten months
	Assigned to Training				
1	ITT=1	84.5%	67.6%	86.5%	67.4%
1A	ITT=1, T=1	88.5%	71.8%	94.7%	76.9%
1B	ITT=1, T=0	63.9%	45.5%	64.3%	63.6%
	Assigned to Control				
2	ITT=0	85.1%	71.3%	82.9%	68.1%
2A	ITT=0, T=1	95.5%	80.0%	88.2%	75.0%
2B	ITT=0, T=0	83.2%	69.8%	81.4%	66.1%

B: Tried out as Supervisor

		Females		Males	
		Four months	Ten months	Four months	Ten months
	Assigned to Training				
1	ITT=1	61.1%	67.2%	70.5%	79.2%
1A	ITT=1, T=1	69.8%	75.4%	88.6%	92.1%
1B	ITT=1, T=0	0.0%	8.3%	0.0%	30.0%
	Assigned to Control				
2	ITT=0	28.7%	33.3%	54.3%	66.1%
2A	ITT=0, T=1	75.0%	80.0%	93.3%	93.8%
2B	ITT=0, T=0	16.2%	21.1%	35.5%	55.8%

C: Promoted to Supervisor

		Females		Males	
		Four months	Ten months	Four months	Ten months
	Assigned to Training				
1	ITT=1	32.0%	49.6%	56.5%	76.9%
1A	ITT=1, T=1	35.8%	53.4%	62.2%	87.5%
1B	ITT=1, T=0	0.0%	0.0%	33.3%	28.5%
	Assigned to Control				
2	ITT=0	8.3%	16.7%	16.9%	32.1%
2A	ITT=0, T=1	27.2%	66.7%	47.1%	61.5%
2B	ITT=0, T=0	4.1%	7.4%	6.3%	23.2%

Table 4A: Basic Outcome Regressions
Four months following training

VARIABLES	(1) Attrition	(2) Attrition	(3) Still in factory	(4) Still in factory	(5) Tried as SV	(6) Tried as SV	(7) Promoted to SV	(8) Promoted to SV
Attended training	0.016	0.017	0.115*	0.119**	0.636***	0.655***	0.421***	0.431***
	(0.03)	(0.03)	(0.06)	(0.06)	(0.09)	(0.09)	(0.07)	(0.07)
Female	0.069**	0.071**	0.004	0.016	-0.153*	-0.076	-0.099*	-0.05
	(0.03)	(0.03)	(0.06)	(0.06)	(0.08)	(0.09)	(0.06)	(0.06)
Female*Attended training	-0.085**	-0.080**	-0.048	-0.061	-0.032	-0.074	-0.107	-0.138
	(0.04)	(0.04)	(0.07)	(0.07)	(0.10)	(0.10)	(0.08)	(0.08)
Age in years divided by 10		0.031		-0.031		0.081		-0.028
		(0.03)		(0.06)		(0.07)		(0.05)
Married		0.000		0.029		0.050		0.074*
		(0.02)		(0.04)		(0.04)		(0.04)
Schooling in year divided by 10		0.069		0.041		0.062		0.178
		(0.07)		(0.11)		(0.13)		(0.11)
Diagnostic score divided by 10		-0.009		0.009		0.030*		0.005
		(0.01)		(0.01)		(0.02)		(0.02)
Said wants to be SV, baseline survey		-0.003		0.041		0.197***		0.109**
		(0.03)		(0.06)		(0.06)		(0.05)
Tenure, RMG sector, years		-0.002		0.005		0.006		0.007
		(0.00)		(0.01)		(0.01)		(0.004)
Tenure, factory, years		-0.008*		0.003		-0.002		-0.006
		(0.00)		(0.01)		(0.01)		(0.01)
Male non-trained mean	0.086	0.086	0.801	0.801	0.275	0.275	0.105	0.105
F-test: Female vs. Male trainee	0.51	0.74	0.29	0.32	.001***	.007***	.0003***	.002***
Observations	493	493	471	471	374	374	412	412
R-squared	0.663	0.673	0.222	0.233	0.566	0.600	0.479	0.502
Factory FE	YES	YES	YES	YES	YES	YES	YES	YES

Table 4B: Basic Outcome Regressions
Ten months following training

VARIABLES	(1) Attrition	(2) Attrition	(3) Still in factory	(4) Still in factory	(5) Tried as SV	(6) Tried as SV	(7) Working as SV	(8) Working as SV
Attended training	-0.001	-0.007	0.146*	0.157*	0.440***	0.456***	0.475***	0.474***
	(0.04)	(0.04)	(0.09)	(0.09)	(0.08)	(0.08)	(0.09)	(0.08)
Female	0.112***	0.118***	0.033	0.036	-0.330***	-0.231***	-0.283***	-0.216***
	(0.03)	(0.04)	(0.07)	(0.07)	(0.08)	(0.08)	(0.07)	(0.07)
Female*Attended training	-0.086*	-0.080*	-0.100	-0.109	0.168*	0.119	-0.010	-0.035
	(0.05)	(0.05)	(0.10)	(0.10)	(0.10)	(0.10)	(0.09)	(0.09)
Age in years divided by 10		0.004		0.161**		0.085		-0.016
		(0.03)		(0.07)		(0.07)		(0.07)
Married		0.015		-0.040		0.055		0.054
		(0.02)		(0.05)		(0.04)		(0.04)
Schooling in year divided by 10		0.090		-0.048		0.126		0.246*
		(0.08)		(0.13)		(0.12)		(0.14)
Diagnostic score divided by 10		-0.001		-0.021		0.039**		0.010
		(0.01)		(0.02)		(0.02)		(0.02)
Said wants to be SV, baseline survey		0.006		-0.039		0.212***		0.162***
		(0.03)		(0.07)		(0.06)		(0.06)
Tenure, RMG sector, years		-0.002		-0.005		0.007		0.020***
		(0.00)		(0.01)		(0.01)		(0.01)
Tenure, factory, years		-0.011**		0.012		0.002		0.005
		(0.00)		(0.01)		(0.01)		(0.01)
Male non-trained mean	0.086	0.086	0.801	0.801	0.275	0.275	0.105	0.105
F-test: Female vs. Male trainee	0.36	0.24	0.32	0.31	.001***	.028**	.0000***	.0000***
Observations	493	493	451	451	396	396	412	412
R-squared	0.523	0.537	0.268	0.301	0.517	0.559	0.494	0.527
Factory FE	YES	YES	YES	YES	YES	YES	YES	YES

Table 5: Management Simulation Exercises
Trainees Females Vs. Males

Outcome Var.: Pay-Off in Games (Standardized)	I	II	III	IV Female team	V Mixed / male team	VI	VII
Female	0.290***	0.309**	0.302***	0.570***	0.184	0.342**	0.365***
	(0.109)	(0.127)	(0.100)	(0.171)	(0.282)	(0.135)	(0.128)
Tried as Line Supervisor						0.321	
						(0.226)	
Promoted to Line Supervisor							0.538**
							(0.233)
Team Fixed Effects	no	no	yes	no	no	no	no
Game Fixed Effects	yes	yes	yes	yes	yes	yes	yes
Team Leader Demogr.	no	yes	yes	yes	yes	yes	yes
Number of Observations	676	592	600	380	212	592	588

Table 6: Productivity of Trainees on the Line

MODEL VARIABLES	(1) Line FE Efficiency	(2) Line FE Sh. Defective Output	(3) Line FE Sh. Absent Worker	(4) Line FE Daily Hours	(5) ANCOVA Efficiency	(6) ANCOVA Sh. Defective Output	(7) ANCOVA Sh. Absent Worker	(8) ANCOVA Daily Hours
Female Trained Supervisor	-0.019*	-0.007	0.002	0.307**	-0.008	-0.004	0.003	0.166
	(0.01)	(0.00)	(0.01)	(0.15)	(0.01)	(0.01)	(0.01)	(0.13)
Male Trained Supervisor	0.006	-0.011*	0.008	-0.028	0.012	-0.007	0.006	0.159
	(0.02)	(0.01)	(0.01)	(0.13)	(0.02)	(0.01)	(0.01)	(0.12)
Dep.Var. Pre-Placement					0.655***	0.529***	0.746***	0.434***
					(0.10)	(0.10)	(0.17)	(0.08)
Observations	102,348	73,323	44,454	101,689	10,996	8,984	3,749	11,415
R-squared	0.045	0.077	0.163	0.177	0.235	0.576	0.513	0.752
Number of factory_LLI	767	739	577	674				
Line FE	YES	YES	YES	YES				
Factory FE					YES	YES	YES	YES
Date FE	YES	YES	YES	YES	YES	YES	YES	YES
TEST: Male Tr.SV. = Female Tr.SV, p-value	0.296	0.616	0.632	0.084	0.370	0.692	0.751	0.961
Nr. Lines with Male Tr.SV.:	24	21	12	21	24	15	12	21
Nr. Lines with Female Tr.SV.:	50	48	35	47	49	44	34	47

Table 7A: Promoted Trainees vs. Existing Supervisors: Cheating Game, Amount Given to ..

Outcome Variable	Operators		Other Supervisor		Line Chief	
Training	0.41	0.56	0.27	-0.03	0.01	0.19
	[0.29]	[0.39]	[0.29]	[0.40]	[0.27]	[0.40]
Female	0.19	0.06	-0.41*	-0.47**	0.05	-0.03
	[0.21]	[0.23]	[0.21]	[0.22]	[0.28]	[0.33]
Training X Female	-0.97***	-0.83**	-0.21	-0.01	0.31	0.19
	[0.42]	[0.37]	[0.39]	[0.46]	[0.41]	[0.50]
Mean (different from 2.5)	1.88***		2.02***		3.0***	
Factory Fixed Effects	yes	yes	yes	yes	yes	yes
Demographics Control	no	yes	no	yes	no	yes
Number of Observations	348	348	348	348	348	348

Table 7B: Promoted Trainees vs. Existing Supervisors: Cheating Game, Amount Received from ..

Outcome Variable	Operators		Other Supervisor		Line Chief	
Training	-0.15	-0.07	0.23	-0.07	-0.34	-0.28
	[0.22]	[0.31]	[0.34]	[0.40]	[0.28]	[0.43]
Female	0.31**	0.29*	0.92***	-0.91***	-0.21	0.25
	[0.16]	[0.16]	[0.24]	[0.25]	[0.25]	[0.26]
Training X Female	-0.06	-0.10	-0.86*	-0.79*	0.53‡	0.31
	[0.29]	[0.33]	[0.49]	[0.53]	[0.40]	[0.46]
Mean (different from 2.5)	1.86***		2.02***		2.10***	
Factory Fixed Effects	yes	yes	yes	yes	yes	yes
Demo. Controls (Receiver)	no	yes	no	yes	no	yes
Demo. Controls (Giver)	no	yes	no	yes	no	yes
Number of Observations	348	348	348	348	348	348