(Mis)allocation of Managerial Training Within the Firm*

Achyuta Adhvaryu
Snaebjorn Gunnsteinsson
Emir Murathanoglu
Anant Nyshadham

October 6, 2021

PRELIMINARY AND INCOMPLETE

Abstract

We study the allocation and productivity consequences of managerial training within the firm via a randomized controlled trial among production line supervisors in an Indian ready-made garment firm. We designed a training program using best practices identified in previous work in a similar setting (Adhvaryu et al. 2021). We asked factory and floor managers (FFMs), who are directly above supervisors in the hierarchy, to identify which supervisors should be prioritized for training, and then randomized access to the program within these rankings. Productivity on lines managed by treated supervisors increased by 7.3% during training relative to control, and remained 5.8% higher relative to control six months after program completion. These gains exhibit substantial heterogeneity across FFM rankings: lines managed by highly recommended supervisors showed nearly zero impact on productivity, while lines managed by supervisors with low recommendation rankings showed impacts of more than 11%. We show that this was not due to a lack of information about baseline skills, or about who would likely gain the most. It also does not appear to be due to discrimination or favoritism along observable dimensions. Instead, consistent with the fact that

*Adhvaryu: University of Michigan, BREAD, NBER, William Davidson Institute & Good Business Lab; adhvaryu@umich.edu; Gunnsteinsson: Independent Researcher; Murathanoglu: University of Michigan; emirmur@umich.edu; Nyshadham: University of Michigan, BREAD, NBER & Good Business Lab; nyshadha@umich.edu. We are very grateful to Anant Ahuja and Chitra Ramdas for their coordination, enthusiasm, and guidance. We acknowledge funding from the International Growth Centre (IGC). Many thanks to Sofia Calderon, Bopanna Changappa, Sadish Dhakal, Priota Paul, and Varun Jagannath for excellent field coordination and research assistance. The views expressed herein do not represent IGC, or Shahi Exports. All errors are our own.

attrition has large personal costs for FFMs in terms of recruitment and onboarding, FFMs prioritized the impacts of training on supervisor retention. Indeed, we find that treated supervisors were 15% less likely to quit than controls during the study period, and that this gain was most pronounced for highly recommended supervisors. Marginal treatment effect analysis demonstrates that much of the determinants of FFM rankings is unobserved i.e., that FFMs leveraged private information in making their recommendations and that these unobserved factors negatively predict productivity effects while positively predicting retention effects. We use this analysis to evaluate the effects of several counterfactual allocation rules. In sum, we show that heterogeneous returns and the inability to effectively target costly training help explain the persistently low managerial quality and small impacts of training observed within many firms in low-income countries.

Keywords: managerial training, productivity, decentralization, resource allocation, India
JEL Codes: J24, M53, O15

1 Introduction

Variation in managerial quality contributes to the documented vast dispersion in productivity and performance both across countries and firms (Bandiera et al., 2020; Bloom and Reenen, 2011; Bloom et al., 2016; Bloom and Van Reenen, 2007) and even within firms across teams and workers (Bertrand and Schoar, 2003; Lazear et al., 2015). A wide array of skills and practices of managers make up the productive value of managerial quality (Adhvaryu et al., 2019b), including people management skills like the subjective evaluation of subordinate workers and teams and the allocation of workers to tasks (Adhvaryu et al., 2019a; Frederiksen et al., 2020; Hoffman and Tadelis, 2021). Given their importance why then do managerial skills and practices (and as a result productivity) vary so substantially within large, sophisticated, and competitive firms when other elements of the production technology (e.g., machinery, raw inputs) appear more optimized and standardized?

We study whether training line supervisors in Indian garment factories on managerial skills and best practices can indeed improve productivity of supervised teams. Furthermore, if the productivity gains are indeed large but heterogeneous, as might be expected given the documented variation in baseline deficiencies and scope for improvement, would production floor and factory managers (FFMs, who lie between line supervisors and top executives in the firm's managerial hierarchy) allocate the training in a way that maximizes gains to the firm if the allocation decision were delegated to them? Or does the targeting of this heterogeneously valuable investment and scope for its misallocation potentially contribute to the persistent variation in managerial quality within firms?

Having multiple layers of management in a firm can be valuable (Caliendo et al., 2020, 2015; Caliendo and Rossi-Hansberg, 2012), particularly if it is possible to decentralize some responsibilities and decisions to lower levels of the hierarchy (Aghion et al., 2021; Bloom et al., 2014; Bloom and Reenen, 2011). The argument is that these middle managers may have some private information and/or specialized understanding that makes them better equipped for making optimal resource allocation decisions, for example. The classic tradeoff, however, is that this decentralization creates a principal-agent structure in which the middle manager may act according to private incentives which do not align perfectly with those of the organization and that limited information at the top of the organization may make enforcing organizational incentives difficult (Acemoglu et al., 2007; Aghion et al., 2014).

To study this exact tradeoff as it relates to the allocation of managerial training within a firm, we solicited from FFMs rankings of which line supervisors should be prioritized for training. We randomized access to training within these rankings to both recover average treatment effects of training on line productivity, supervisor pay and retention as well as to investigate whether FFMs indeed recommended supervisors who ultimately gained most. We find that line supervisors gained substantial knowledge from the training, with test scores of treatment supervisors increasing by 40 to 100% as compared to control supervisors who exhibited no significant gains as expected. As a result, productivity of teams managed by trained supervisors increased substantially and persistently on average. Specifically, lines for which all managers were trained exhibited on average 4.1 percentage points higher productivity during the training (a 7.3% increase over the control mean) and a persistent 3.3 percentage points (5.8% over the control mean) over the 6 months after training completion when the observation period ended.

These productivity gains, however, were quite heterogeneous and line supervisors recommended highly by FFMs to receive the training actually gained little to nothing from the training. Analysis of treatment impacts on supervisor retention provide insight into a competing incentive which might help to explain this misalignment between the firm's objective to maximize productivity gains and the recommendations provided by FFMs. Training generated a significant positive treatment effect on retention, with trained supervisors more than 15% less likely to quit than control supervisors, and these impacts appear to be driven entirely by the highly recommended supervisors, who exhibited a 23% reduction in quitting as compared to only a 3% reduction for low recommendation counterparts. In addition high recommendation supervisors in the control group are more likely to quit in the absence of training than are low recommendation control supervisors. Though this retention heterogeneity analysis is not as well-powered as the average treatment effects and the productivity analyses, taken all together, the results suggest that FFMs may know which supervisors are most likely to quit and that allocating a training investment of this sort to them may improve their retention. Accordingly, FFMs appear to tailor their training recommendations to take advantage of this potential improvement in retention.

Anecdotal evidence from discussions with FFMs confirms that attrition of production line supervisors places a disproportionate private burden on them to screen and train replacements and adopt their duties in the interim. We note that the productivity treatment effects reflect gains net of any supervisor turnover on average, and that the return on investment implied by these net productivity gains is several orders of magnitude larger than any monetary costs borne by the firm to screen and train new supervisors. Accordingly, the firm clearly favors allocating the training to maximize gains in productivity (as would workers and supervisors who all earn significantly greater incentive pay as a result of treatment effects on productivity), but the FFMs have competing incentives to improve line supervisor retention in order to minimize the private burden to them of screening and training replacements and covering the supervisor duties in the interim.

Additional LASSO analysis of drivers of FFM recommendations suggests that they partially reward perceived motivation and effort as well as quantity and variety of experience among of supervisors. In order to assess whether these observable drivers make up a substantial component of the variation in recommendations, we leverage a recent approach by Dal Bó et al. (2021) to estimate marginal treatment effects (MTE). This analysis confirms that substantial variation (at least 80%) in FFM recommendations derives from unobserved drivers, and that this unobserved component (perhaps most indicative of the private information to be leveraged via decentralization of the training allocation decision) positively predicts improvements in retention despite negatively predicting productivity gains. This evidence supports the interpretation that FFMs recommend training according to their private incentive to reduce attrition at the expense of the firm's productivity-related priorities.

Finally, we assess whether FFMs in fact have little or no private knowledge upon which to base recommendations and evidence of discrimination or favoritism as alternative interpretations. The MTE analysis allows us to estimate treatment effects resulting from counterfactual allocation rules. From these, we show that FFM ratings of line supervisors' management and industrial engineering skills (reflecting the communication and planning skills most centrally addressed in the training) contain accurate assessments of deficiencies which could have been used to allocate the training effectively. We interpret this as evidence against the alternative interpretation that FFMs did not possess any valuable private knowledge upon which to base allocation decisions. We also note that productivity treatment effect heterogeneity by recommendations is significantly negative rather than insignificantly zero as might be the case if recommendations contained no valuable signal. Though discrimination or favoritism may be partially at play, we find no empirical evidence in support of this interpretation, leaving competing incentives as the primary interpretation for which we have clear empirical evidence.

Our study contributes to the rich empirical literature on management and productivity (Bandiera et al., 2020; Bloom et al., 2016; Bloom and Van Reenen, 2007), particularly building on recent papers studying variation in quality across managers within a firm and the resulting dispersion in team and worker outcomes (Adhvaryu et al., 2019a,b; Frederiksen et al., 2020; Hoffman and Tadelis, 2021; Lazear et al., 2015). Specifically, we build on recent randomized controlled trials studying interventions to improve managerial practices and quality (Bloom et al., 2013; Gosnell et al., 2020) by documenting the substantial heterogeneity in treatment impacts and the scope for misallocation of training investments. That is, we show that though the average productivity gains from a random allocation were large, persistent, and generated tremendous return on investment, some supervisors gained little to nothing. If these supervisors had been targeted (as would have been the case if the allocation decision were decentralized to FFMs) the productivity gains and resulting return on investment would have been much smaller or even negligible. In this sense, our results provide one potential explanation for why managerial quality remains low on average in many firms despite strong evidence of potential gains from investments in management such as the training program we evaluate.

We also contribute to the empirical literature studying decentralized decision-making in firms (e.g., Acemoglu et al. (2007); Bloom et al. (2010)). We add to recent evidence showing how decentralization can positively enable performance particularly during "bad times" (Aghion et al., 2021), by documenting empirical evidence of the hypothesized risk that competing incentives at lower levels of management may lead to decisions which do not optimally serve firm objectives (Aghion et al., 2014). Specifically, we show that managers at lower levels of the organizational hierarchy may indeed have valuable private information which can be used to target managerial training given heterogeneous impacts, but may prioritize impacts which are of higher priority to them personally than to the organization as a whole. Importantly, the retention of line supervisors which FFMs appears to prioritize is, of course, not without value or importance to the firm, but rather the firm would simply prioritize productivity gains (which deliver orders of magnitude larger returns) when the two priorities are at odds, as turns out to be the case in our scenario.

2 Context, Program Details, and Experimental Design

2.1 Context

We partnered with the largest contract manufacturer of readymade garment in India (among the top five largest garment exporters in the world), Shahi Exports, Pvt. Ltd., to implement and evaluate a management training program among production line supervisors. Given the continued labor-intensive production technology in the garment industry despite adoption of modern production concepts such as specialization, assembly lines, and lean production, garment manufacturing provides an excellent setting in which to study the impacts of training in personnel management practices on productivity.

Roughly 60 factories owned and operated by the firm produce orders for hundreds of international brands each year, generating revenue of over a billion USD annually. There are three stages in the production process. First, fabric is cut and organized into bundles of subsegments for different parts of the garment (e.g., sleeve, front placket, collar) by cutting teams. These bundles of materials are then transferred to sewing lines in which machine operators construct each portion of the garment and attach these portions together to make complete garments. Finally, the sewn garments go through finishing (e.g., washing, trimming, final quality checking) and packing for shipment in advance of a contracted delivery date.

Across the cutting, sewing, and finishing departments representing these three stages of production, each factory employs thousands of workers allocated across tens of teams, each with at least 1 supervisor, and often several assistants of various designations (e.g., assistant supervisors, feeder, floater, captain). In smaller factories, a single cutting team and finishing team will service most or all sewing lines, but in larger factories each sewing line may have its own matched cutting and finishing teams. Each sewing line produces a unique order or style until completion, before progressing to the next contracted order.[1]

[1] Orders from brands are allocated across factories by the marketing department of each production division

As discussed below, we randomized access to the training across supervisors from all three departments (as well as occasionally some additional supervisors deemed eligible by the factory from other support departments such as HR), but when studying impacts on productivity we focus only on the supervisors who are mapped to specific production lines for which we measure productivity (i.e., primarily sewing department in most factories, but some cutting and finishing in some factories when those supervisors are linked to specific sewing lines). Supervisors of sewing lines are assigned permanently to their line and are responsible for several key oversight tasks. First, when a new order is assigned to a line, the line supervisor must determine how to organize the production process, taking into account both the machines and workers available as well as the specific operations required and overall complexity of the garment style. This initial line architecture (known as "batch setting") is always set at the start of a new order and is rarely and minimally changed for the life of that order to avoid downtime.

Over the order's production run (lasting usually weeks, but sometimes months), if productivity imbalances or bottlenecks arise (often due to idiosyncratic worker absenteeism), sewing line supervisors will most often switch the task allocations of some set of workers across machines, or add a helper or second machine to some critical operations (often borrowing from other lines), but preserving the line architecture otherwise (Adhvaryu et al., 2019a). This recalibration of the worker-machine match (known as "line balancing"), which depends crucially on effective communication with workers and substantial monitoring effort represents one pathway by which managerial quality contributes to the marked increases in productivity seen over the life of an order in this setting (Adhvaryu et al., 2019b). The initial "batch setting" depends crucially on the planning and organizational skills of the supervisor and contributes to vast dispersion in the productivities achieved under different line supervisors, even after accounting for garment style and worker skill and quality (Adhvaryu et al., 2020).

The managerial hierarchy of the firm involves several layers. Supervisors of teams as discussed represent the frontline of management. They report to production floor managers in larger factories with multiple floors and/or many lines on a floor, or to factory level production managers directly in smaller factories. These are the FFMs from whom we solicited training allocation recommendations. The factory level production manager works

(Knits, Mens, and Ladies) based on capacity and regulatory and/or compliance clearance (i.e., whether a particular factory been approved for production for that brand given its corporate and governmental standards), and within factory, by first availability (i.e., whichever line is closest to finishing its current order when an incoming order is processed will be allocated the new order).

alongside the general manager of the factory who also oversees broader operations at the factory level. As mentioned above, there are roughly 60 factories with this structure, organized into 3 divisions of the firm (Knits, Mens, and Ladies) with roughly 20 factories in each division. The production and general managers of each factory report to the COO and CEO of their division. These 3 division CEOs and 3 division COOs report to the board (on which they also serve), and the board is overseen by a Managing Director (i.e., the head of the organization). Accordingly, we think of the FFMs at the factory level as "middle managers" who report to division and firm level top management and who possess knowledge regarding factory level operations and work closely (daily) with the frontline supervisors among whom the training is being allocated.

2.2 Program Details and Background

Drawing from our prior work in this specific context showing the productive value of both soft skills such as communication and specific managerial skills and practices such as control, autonomy, and attention (Adhvaryu et al., 2018, 2019b), the STITCH program was designed to train line supervisors in the skills and practices most likely to improve productivity. The program consisted of 4 modules, each of which focusing on a different aspects soft-skills and leadership training. Figure 1 presents a diagram with all 4 modules and the topics they cover.

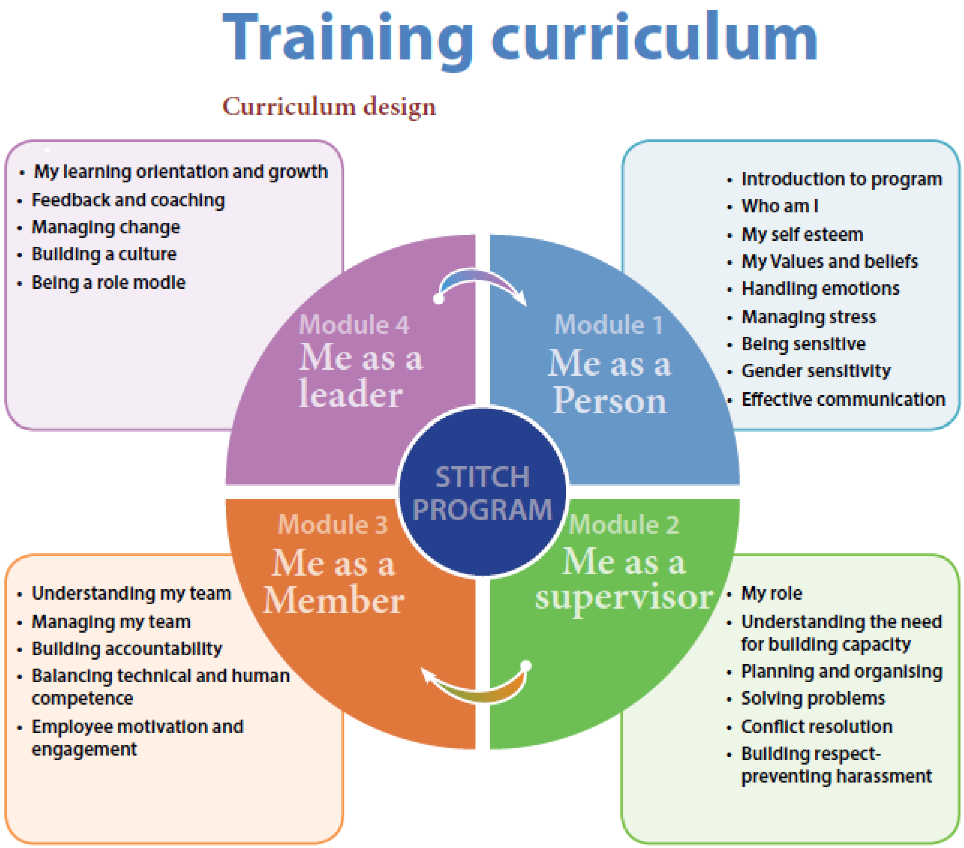


Figure 1: STITCH Curriculum

2.3 Experimental Design

Training participants were chosen from a pool of supervisors indicated by management to be eligible for training. All eligible supervisors were administered the baseline survey and were randomized into treatment and control. This gives us a baseline sample of 1849 supervisors. Employees that oversaw supervisory roles yet were not officially designated as supervisors (such as assistant supervisors or floaters) could also be indicated by management as being eligible for training. We do not make a distinction based on official designation and refer to each eligible employee as supervisors for the rest of the text. The FFMs, again as indicated by management, have also been administered a baseline. In the FFM baseline, among other items, FFMs were asked to rank the supervisors they managed (from 1 to 5) according to how much they believed the supervisor would gain from training (referred to as FFM recommendation for the rest of the text). We collected FFM recommendations for

1175 supervisors included in the analysis.

Randomization was stratified in multiple dimensions. First, each factory floor were randomized into high or low saturation where 75% of the supervisors in high saturation and 25% of the supervisors in low saturation floors were randomized into treatment.[2] Second, supervisors were grouped by whether they were highly, moderately, or low recommended for training by their FFM. Finally, supervisors were put in similarity clusters based on personal and line characteristics.

While randomization was at supervisor level, our key outcome of productivity is at line level. Of the 1849 supervisors administered a baseline, a subset of 954 supervisors who (1) undertook duties directly related to production[3] and (2) could be linked to specific production lines, are included in our productivity analysis. These 954 supervisors were linked to 561 production lines. The line level treatment is defined as the proportion of supervisors in a line who were treated. This leads to a continuous line level treatment between 0 and 1. Treatment is evenly centered around 0.5 (see line level treatment distribution in Appendix Figure A.2)

Figure 2 presents a schematic diagram of our experimental design. Of the baseline sample 1849 supervisors, 921 were randomized into control and 928 were in the treatment group. Of the 561 production lines, 164 had no supervisors treated, while 397 had at least one treated supervisor. Summary statistics and balance checks are presented in section 3.4.

[2]Supervisors who could not be linked to a production floor had 50% probability of being treated.

[3]This excludes supervisors who, for example, are in HR and Admin departments or are data entry operators in accessory stores.

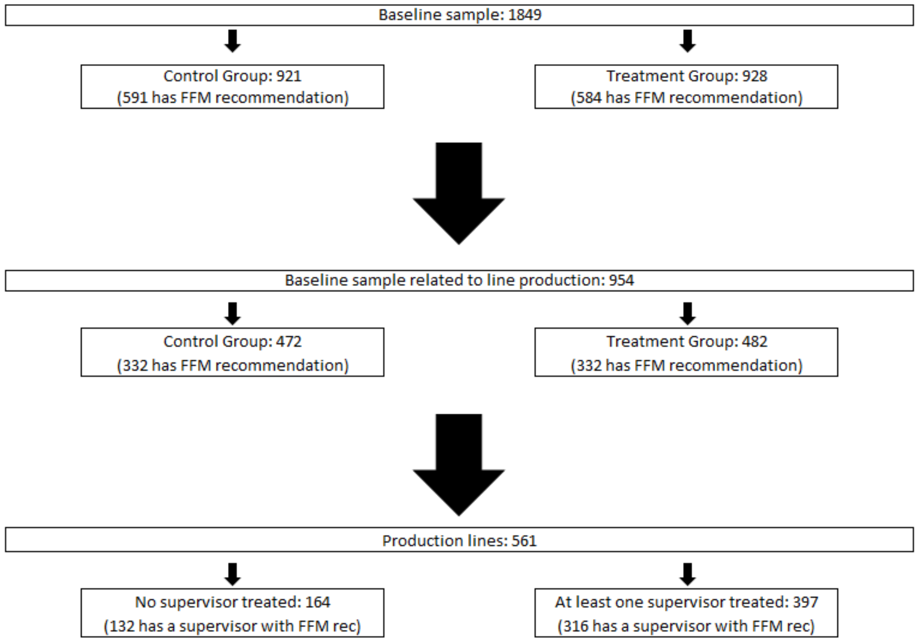


Figure 2: STITCH Study Experimental Design

Appendix Figure A.1 presents a timeline of our intervention. FFMs and supervisors were administered the baseline surveys from December 2016 to March 2017. Training start and end dates were different across different factories.[4] The earliest training started in April 2017, and the latest completion was in March 2018. We discuss the various survey instruments and other data sources in Section 4.

3 Simple Model of Training Allocation

In this section, we posit a simple model of the effects and the allocation of supervisor training. The model consists of a firm, an FFM, and the supervisors whom training will be allocated to. Training affects both the individual productivity and the retention probabilities of supervisors. We study how the firm (the principle) and the FFM (the agent) would choose to allocate the training to supervisors who gain differentially from it. The assumptions of the models are set to mirror our empirical context.

[4]While the training start dates for factories were not randomized and were partly driven by logistical reasons, within each factory the lines are randomized into treatment and control.

3.1 Setup

Supervisors There are two periods. A population of line supervisors are present in period 1. They have identical per-period productivity p and have quitting probability $(1-\delta)$ from period 1 to period 2. Training affects both the productivity and the quitting probability of supervisors. Supervisors are heterogeneous with regards to their responsiveness to training, where supervisor i has productivity responsiveness τ_p^i and retention responsiveness τ_δ^i. Supervisor training takes place in the first period and its productivity effects are realized in period 1. So, a trained supervisor produces $p+\tau_p^i$ both periods, conditional on not quitting. The quitting probability of trained supervisors are $(1-\delta-\tau_\delta^i)$. Line supervisors who quit after period 1 are replaced. The replacement supervisors have productivity zp where $z \in (0,1]$. This term is meant to capture, for example, that replacement supervisors can initially be less productive or that during the process of replacement the line may be idle for a period. Note that, because the replacement supervisors are not trained, their productivity is not shifted by τ_p.

Payoff for the Firm and the FFMs Both the firm and the FFMs are risk neutral. The firm's objective function is to maximize line productivity. The FFMs also aim to maximize line productivity, but they also pay an additional personal cost c in period 2 if the line supervisor quits after period 1. This term corresponds to both the personal replacement and training costs the FFM incurs in addition to the productivity costs, and also to the fact that, in our context, part of the FFMs job is to ensure retention of supervisors. Therefore, there can be a professional cost to losing supervisors. The economic consequence is that the personal replacement cost c misaligns the principal's (the firm) and the agent's (the FFMs) objectives.

We write the value an FFM gains from a trained and untrained supervisor i as:

$$\text{Not Trained:} \quad \underbrace{p}_{\text{Period 1 payoffs}} + \underbrace{\delta p}_{\text{Period 2 payoffs if supervisor stays}} + \underbrace{(1-\delta)(zp-c)}_{\text{Period 2 payoffs if supervisor replaced}}$$

$$\text{Trained:} \quad \underbrace{(p+\tau_p^i)}_{\text{Period 1 payoffs}} + \underbrace{(\delta+\tau_\delta^i)(p+\tau_p^i)}_{\text{Period 2 payoffs if supervisor stays}} + \underbrace{(1-\delta-\tau_\delta^i)(zp-c)}_{\text{Period 2 payoffs if supervisor replaced}}$$

The difference between the trained and the untrained value yields the FFM payoff from training i, denoted Δ^i:

$$\Delta^i(\tau_p^i, \tau_\delta^i) = \underbrace{\tau_p^i(1+\delta) + \tau_\delta^i((1-z)p) + \tau_\delta^i\tau_p^i}_{\text{Productivity Gains} \equiv \Delta_p^i} + \underbrace{\tau_\delta^i c}_{\text{FFM Personal Costs} \equiv \Delta_c^i}$$

where Δ_p^i is the fraction of the gains due to line productivity effects and Δ_c^i is the fraction due to avoiding personal replacement costs through retention effects. Given the firm's objective would be to maximize Δ_p^i when allocating the training, Δ_c^i represents the wedge in payoffs induced by FFMs personal cost to losing a supervisor. Overall line-level productivity gains from training Δ_p^i is not equal to the per-period productivity gain of the supervisor τ_p^i, as effects on retention do indirectly influence productivity through changing the probability of holding on to incumbent (and possibly more productive) supervisors and avoiding supervisor replacement costs that effect productivity (all of which are captured by the term z).

FFM Information and the Distribution of Types Supervisor types are indexed by their training gains in terms of productivity and retention. The marginal distribution of the productivity gains is $\tau_p^i \sim G(.)$. We assume the FFMs have perfect knowledge about both the τ_p^i of each supervisor and the conditional average of retention gains $\mathbb{E}[\tau_\delta^i|\tau_p^i] = f(\tau_p^i)$. Critically, we assume that τ_p^i and τ_δ^i are negatively correlated. This induces a trade off between gains in terms of retention and productivity, which is consistent with what we observe in the empirical analysis. Specifically, we impose $f'(\tau_p^i) < 0$ and $f''(\tau_p^i) < 0$. Finally, the ultimate retention probability $\delta + \tau_\delta$ must lie in the unit interval.

3.2 Ideal Supervisor Type to Train

First, we consider the question of what supervisor type has the highest training payoffs from the perspective of the FFM, which we call the *ideal type* from the perspective of the FFM.[5] The FFM aims to maximize the expected payoff from training a type with τ_p^i:

$$\max_{\tau_p^i} \; \mathbb{E}[\Delta^i(\tau_p^i, \tau_\delta^i)|\tau_p^i] = \max_{\tau_p^i} \; \tau_p^i(1+\delta) + f(\tau_p^i)((1-z)p) + f(\tau_p^i)\tau_p^i + f(\tau_p^i)c$$

[5]Note that this is a distinct exercise from choosing to allocate the training to all the members of a type, as this would also depend on the relative density of the type in the distribution of supervisors and availability of training. Here we are simply interested in training which type of supervisor individually provides the highest value to the FFM.

which leads to the ideal type:

$$\tau_p^* = \frac{1+\delta+f(\tau_p^*)}{-f'(\tau_p^*)} - p(1-z) - c. \tag{1}$$

and the expected productivity gain for the ideal type:

$$\mathbb{E}[\Delta_p^*] = \tau_p^*(1+\delta) + f(\tau_p^*)((1-z)p) + f(\tau_p^*)\tau_p^* \tag{2}$$

Remark The expected productivity gain from the ideal type decreases as the FFM personal cost increases, i.e. $\frac{d\mathbb{E}[\Delta_p^*]}{dc} < 0$.

We show the derivation of the result in Appendix B.1. The intuition is straightforward. As the personal cost of losing a supervisor increases, the FFM put more and more emphasis on targeting supervisors who show large retention effects, giving up productivity gains in the process. This implies that the relative ordering of supervisors (and the allocation of scarce training) increasingly differs between the firm and the FFM as c increases, due to the tradeoff between τ_p^i and τ_c^i.

3.3 Treatment Effects in the Model

The Average Treatment Effect If the firm randomizes allocation (or trains every supervisor), the average treatment effect in terms of productivity would be $\mathbb{E}[\Delta_p^i] = \mathbb{E}[\tau_p^i](1+\delta) + \mathbb{E}[f(\tau_p^i)]((1-z)p) + \mathbb{E}[f(\tau_p^i)\tau_p^i]$. This expression is the theoretical counterpart of the productivity ATE estimates we get in our empirical work as we do our analysis at the level of the production line.[6] Again, the overall productivity gains $\mathbb{E}[\Delta_p^i]$ are distinct from the supervisor level per-period productivity gains τ_p^i. By focusing on the line productivity over time, our estimates take into account any possible productivity gains/losses induced by changes in supervisor retention, captured in the model by retention effects and the parameter z.

Our model does not assume that the per-period treatment effects τ_p^i and τ_δ^i are distributed

[6]In the stylized model, there are no dynamics to the treatment effect and the training is instantaneous. In reality, the training is spread over many months and we take simple dynamics into account by estimating treatment effects for while the training is ongoing and after the training is complete.

such that the ATE is positive. $\mathbb{E}[\Delta_p^i] > 0$ only if $G(.)$ and the $f(\tau_p)$ are such that enough weight is put on supervisors that provide an overall productivity gain to the firm.

Heterogeneity with FFM Recommendation If the firm knew $(\tau_p^i, \tau_\delta^i)$ for every supervisor, it could allocate the scarce training by ordering the supervisors by Δ_p^i and allocate the resource accordingly.[7] This would guarantee a larger treatment effect than random allocation. However, if the firm relies on FFMs to allocate the training due to information frictions, whether the treatment effects are higher or lower than random allocation depends on the relative size of the personal replacement cost c. If c is negligible, the FFM allocation would approximately be the same as the firm allocation as the FFM also wants to maximize line productivity. If, on the other hand, c is relatively large, FFMs could heavily target retention, producing productivity gains below randomization in the process.

Our research design allows us to compare the ATE of supervisors who were and were not recommended by their FFMs by decoupling the selection from the treatment allocation. Relative magnitudes $\mathbb{E}[\Delta_p^i|\textit{Recommended}]$ and $\mathbb{E}[\Delta_p^i|\textit{Not Recommended}]$ is informative about the relative size of c, as we would only see a smaller ATE for the recommended supervisors if the incentive of the FFMs to overemphasize retention gains are high.

Decomposing Selection and Marginal Treatment Effects Here, we extend our model to decompose the FFM selection into observable and unobservable components.[8] To do this, we follow the setup above that supervisor are perfectly knowledgeable about the per-period productivity gains of the supervisors, and they recommend supervisor both based on the overall line productivity gains (Δ_p^i) and for the retention gains above-and-beyond its effect on productivity (Δ_c^i).[9] We introduce a reduced form version of the stylized model terms with tilde notation $\tilde{.}$. Denote the value of recommending supervisor *i* as $\tilde{\Delta}^i$:

[7] This decision can include not to train supervisors who would not gain from training (or would gain less than the cost of the training if there is a cost to tranining an individual)

[8] This framework closely follows the methodology of Dal Bó et al. (2021). Our setup differs from their in one key dimension. In their model what leads to a null relationship between the agent's selection and productivity gains is information frictions, where the agent only has imperfect information about productivity gains. If the agent's signal is very weak, agent's selection may not be related to treatment gains. We do not focus on information frictions, but instead allow for productivity related and unrelated unobservables to be negatively correlated.

[9] Technically, any other non-productivity related preference towards recommending a supervisor can be subsumed under Δ_c^i as well.

$$\tilde{\Delta}^i = \underbrace{\beta' X_i + \eta_i}_{\equiv \tilde{\Delta}^i_p} + \underbrace{\psi' X_i + \theta_i}_{\equiv \tilde{\Delta}^i_c}$$

where X_i is the observable characteristics of supervisor i, $\tilde{\Delta}^i_p = \beta' X_i + \eta_i$ is the productivity gains from training for supervisor i, and $\tilde{\Delta}^i_c = \psi' X_i + \theta_i$ is the idiosyncratic supervisor preferences for recommending *i*. Both the gains from training and the manager preferences have a component that can be explained by observable characteristics (β and ψ) and a component that is unobservable to the analyst (η_i and θ_i). We pool the observable and the unobservable terms together as $\Gamma \equiv \beta + \psi$ and $u_i \equiv \eta_i + \theta_i$. We then model the decision to recommend a supervisor as recommending the supervisors above a threshold (normalized to 0): $Rec_i = 1[\Gamma_i + u_i > 0]$. We impose further structure to the model by assuming that (η_i, θ_i) are jointly normally distributed with mean 0. This structure yields the following expected productivity gain (derived in Appendix Section B.2):

$$\mathbb{E}[\tilde{\Delta}^i_p | X_i, Rec_i] = \beta' X_i + \rho_{u\eta}\sigma_\eta \times \lambda(X_i, Rec_i) \tag{3}$$

where $\lambda(X_i, Rec_i) \equiv \frac{\phi\left(\frac{-\Gamma_i' X_i}{\sigma_u}\right)}{Rec_i - \Phi\left(\frac{-\Gamma_i' X_i}{\sigma_u}\right)}$ is the Inverse Mill Ratio (IMR) and the $\rho_{u\eta}$ is the correlation coefficient between u_i and η_i. The IMR can be estimated using a probit regression[10] and be plugged in as a covariate to the estimating equations. If the unobservable component of productivity gains from treatment is negatively correlated with the entire unobserved component of the FFM selection decision (i.e. $\rho_{u\eta} < 0$), the coefficient on the IMR interaction will be negative. This would imply the unobserved component of the FFM selection is negatively related to productivity gains. β captures the effects of all the observable components. Once we estimate the model, we can estimate marginal treatment effects and compare the performance of FFM allocation with randomization.

[10]Specifically, parameters Γ can be estimated using a probit regression and be plugged into the IMR equation $\lambda(X_i, Rec_i)$.

4 Data

We use a combination of administrative data from the factories and survey data to evaluate the program and study its allocation. We discuss our different data sources below.

4.1 Production Data

Each production line on the sewing floors record hourly productivity data. We aggregate the hourly data to daily-level for each line. The key productivity measure in our analysis is efficiency. Efficiency is defined as daily garment quantity produced over the target quantity for the day.. Efficiency accounts for the complexity of the operations performed as the target quantities are calculated by the firm using a global garment industry standard measure called Standard Allowable Minutes for each garment type. Finally, measurement of these measures have been undertaken by the firm independent of STITCH training; therefore we have access to productivity date before, during, and after the training. For baseline values, we use averages of three months before training start (January - March 2017). Our analysis period spans the 6 months before and 6 months after the training start and end for each production line.

4.2 Human Resources Attendance, Salary, and Personnel Data

Human resources collects daily attendance data reporting whether an employee has attended work on a given day. We use this data to analyze supervisor attendance. More importantly for our purposes, we use the attendance rosters to ascertain whether a worker is retained by the firm on a given day to investigate retention results.[11] Using human resources personnel rosters, we further match around 55000 workers to the production lines with randomized supervisors. We use the attendance rosters data for these workers to analyze both baseline values and treatment effects regarding attendance and retention for workers (as opposed to supervisors). We also have have access to monthly salary data. We use this data to see whether trained supervisors experience differential salary growth. Finally, the firm has an incentive scheme where bonus payments are made to employees based on performance. Daily data on incentive payments are collected by the firm and, importantly for our purposes, indicates the designation and the line of the individual who has recieved the incentive

[11] While employees who quit are eventually dropped from the roster, this can happen with delay. We can use the trailing absences before a worker is dropped to pin down the effective date an employee has quit.

payments on a given day.

4.3 Survey Data

We complement the administrative data with three surveys. First, we administered a baseline survey of supervisors eligible for training in January to March 2017. The survey covers demographics, experience and tenure, various aspects of managerial quality and style, personality characteristics, and self assessment of skills. We use these characteristics to investigate determinants of FFM recommendation (discussed in the next paragraph) and the heterogeneity of treatment gains. B.4 provides a list of survey indeces we use.

We also conducted a survey of FFMs, who are above the supervisors in the firm hierarchy. We primarily use this survey to elicit information about who the FFMs think would gain the most from training. Specifically, we communicate that the firm is unrolling a supervisor training program focused on leadership and soft-skills yet does not have the capacity to train all supervisors. Therefore, the firm wants to prioritize supervisors who would gain the most. We then ask the FFMs to recommend (from 1 to 5) supervisors based on who would benefit the most. We also elicit information from the FFMs about the managerial skills, technical skills, industrial engineering skill, and the motivation of the supervisors they manage.

4.4 Pre and Post Module Test Scores

Before and after each training module, a randomly selected group of treated and control supervisors are given a short test covering the material of the module. We use the percentage point scores of these tests to assess whether treated supervisors learn to content of training (at least theoretically).

4.5 Summary Statistics and Baseline Balance

Table 4.1 presents summary statistics and balance checks across many characteristics of interest at the supervisor level. Given our key outcome of productivity is at the line level, we show further summary stats for production lines and check for balance at the line level in table 4.2. Overall, we do not see any balance issues in our full sample between the treated and control supervisors and lines. In appendix Table C.1 , we present further summary statistics and balance checks for several analysis subsets. Specifically, as we further discuss below, we drop lines with above a certain cutoff of zero productivity days from our analysis in our preferred specification, as in our context this is likely a data entry error as opposed to actual zero productivity. Further, in our heterogeneity analysis with regards to FFM recommendation, we limit our sample to lines for which we have FFM recommendations (i.e. lines who are mapped to supervisors with FFM recommendations). Some imbalance is introduced for the subsets.[12] However, as we discuss later, we are using a difference-in-differences specification for line level outcomes, for which level differences do not pose an identification concern in the presence of parallel trends. In Appendix Figure C.1 we show that there is no evidence of pre-trends for our main analysis lines. Regardless we present results using the full set of lines available to us alongside our preferred subset to show that the coefficients are stable across samples.

Table 4.1: Supervisor Level Descriptive Statistics and Balance

	Num Supervisors	Mean	SD	Coefficient/SE
Supervisor Age	1849	31.27	6.21	-0.216
				(0.289)
Supervisor 1(Male)	1849	0.75	0.43	-0.012
				(0.020)
Supervisor Finished Highschool	1849	0.12	0.32	-0.023
				(0.015)
Supervisor Worked Different Line Before	1849	0.41	0.49	-0.035
				(0.023)
Supervisor Ever Oeprator	1849	0.73	0.44	-0.022
				(0.021)
Supervisor Worked Different Factory	1849	0.24	0.43	-0.026
				(0.020)
Months as Supervisor	1849	61.45	45.14	-2.495
				(2.100)
Months Supervising Current Line	1848	29.18	29.83	0.629
				(1.388)
Years in Shahi	1849	6.72	4.73	-0.109
				(0.220)
FFM Recommendation	1175	3.05	1.61	-0.066
				(0.094)
Technical Skill (scored by FFM)	1175	3.95	0.94	-0.023
				(0.055)
Industrial Engineering Skill (scored by FFM)	1175	3.96	0.94	-0.004
				(0.055)
Management Skills (scored by FFM)	1171	4.01	0.93	-0.010
				(0.054)
Supervisor Motivation (scored by FFM)	1171	4.17	0.86	0.011
				(0.050)

Summary statistics are for all supervisors. The coefficient(SE) is from regressing the outcome on the binary treatment indicator. Robust standard errors are reported (* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.).

[12]For the analysis subsample baseline productivity is 6% lower for lines with all treated supervisors (significant at 5%). For the FFM subsample, baseline attendance is 11 % lower for fully treated lines (significant at 10%).

Table 4.2: Line Level Descriptive Statistics and Balance

	Num Lines	Mean	SD	Coefficient/SE
Line Efficiency (Baseline)	540	54.96	9.31	0.940
				(0.978)
Line Attendance (Baseline)	541	0.89	0.05	-0.007
				(0.006)
Line Retention (Baseline)	528	0.83	0.14	0.004
				(0.016)
Line Budgeted Efficiency (Baseline)	541	60.60	7.58	-0.217
				(0.836)

Summary statistics are included for all lines. The coefficient(SE) is from regressing the outcome on the continuous treatment indicator. Robust standard errors are reported (* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$). All baseline values are from 3 months preceding training start (January - March 2017). Baseline (budgeted) efficiency is an average of daily (budgeted) efficiency values for this period. Baseline attendance and retention are the attendance and retention outcomes for the workers matched to these lines using the personnel rosters.

5 Treatment Effects

Below, we test the average impact of the program on several key outcomes. We begin by showing the treated supervisors perform better on tests administered after each training module. Treating this as a first-stage, we then analyze the effects of the STITCH training on productivity and retention. We follow this section by an examination of whether supervisors selected by the FFM indeed gained more in terms of productivity and attendance, analyzing the observable and unobservable components of FFM selection. Treatment effects on other outcomes that are not central to our research design are presented in Section 7

5.1 Pre and Post Module Assessment

We first investigate the whether treated supervisors have performed better on tests on the material that were covered in the four training modules.

Figure 3 presents the average difference between the pre- and post-training module test score differences for the control and treatment supervisors. While the increase in score is statistically indistinguishable from 0 across control supervisors, treated supervisors increase their test scores significantly. Appendix Table D.1 presents estimates of the treatment effect for each module using an ANCOVA specification. Consistent with the figure with raw differences, treated individuals outperform the control supervisors significantly for each module. At the low-end, treatment increases the performance of supervisors on module 2 tests by around 22 percentage points, which corresponds to 40% from a baseline mean of 55. The treatment effectively doubles the post-training test scores for module 4 by inducing an increase of 39 percentage points. We conclude that training leads to learning of the material covered by the modules.

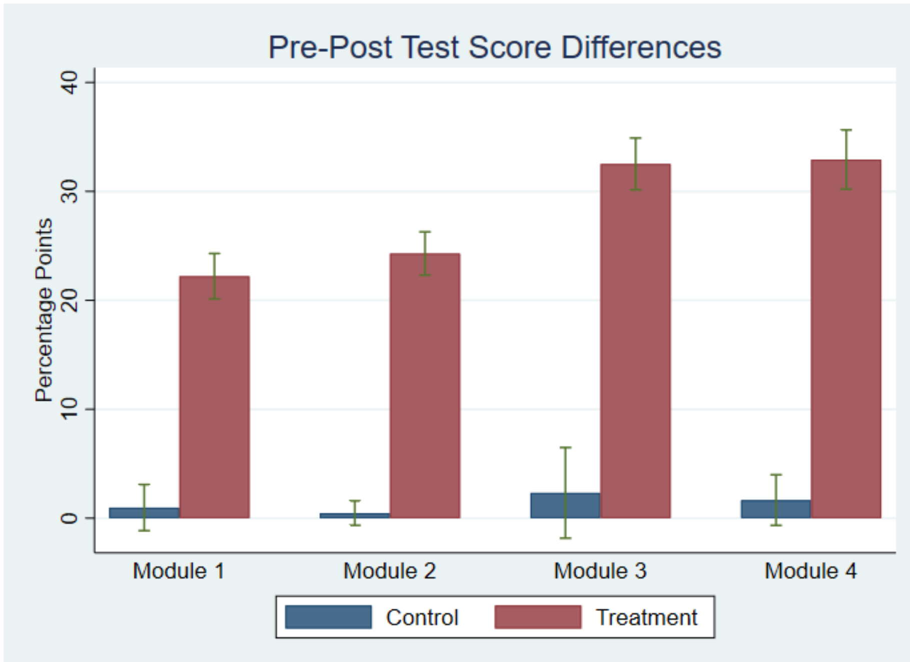


Figure 3: Percentage point difference between the pre- and post- module test scores for the treatment and control supervisors. 95 % CI shown.

5.2 Productivity

Next, we investigate the impact of training on line productivity. Our outcome of interest is efficiency, which is the industry standard measure of productivity defined as quantity produced over target quantity. We use the following ITT difference-in-differences specification to assess the productivity effects on a line-day level:

$$y_{ltr} = \alpha + \beta_1 T_l \times 1[During_t] + \beta_2 T_l \times 1[Post_t] + \delta_l + \mu_t + \gamma_r + \epsilon_{ltr} \tag{4}$$

where y_{ltr} is productive efficiency of line l on date t and the relative time indicator r, T_l is treatment as defined by fraction of supervisors treated, and $1[During_t]$ and $1[Post_t]$ are indicators for whether training is ongoing or over in the factory of the line. The training-relative date r is set to 0 on the first day of the month of training start and it captures how many days have elapsed since the beginning of training on the factory of the line (with negative dates for before training).[13] In our analysis sample, we exclude line-day observations with 0 efficiency as these likely reflect data errors as opposed to days where lines actually produced no output. Further, certain lines have many days reported as 0 productivity. We exclude any line that has over 20% of recorded days with 0 productivity in any period (pre-, during- or post- training) from the analysis.[14] Our coefficients of interest are β_1 and β_2 which estimate the causal effect of fraction of supervisors treated on line level productive efficiency.

As shown in Table 5.1, treatment has a statistically and economically significant effect on efficiency. Column 3 reports our preferred specification, which includes line, date, and relative date fixed effects (columns 1 and 2 show robustness to including a less stringent set of fixed effects). Lines with all treated supervisors are, on average, are 4.1 percentage points (7.3% of control mean) more efficient during training. Given the training were administered over a considerable duration (the average length of training was 267 days across lines), this is an economically significant effect. For the 6 months following training end, lines with all supervisors treated still have 3.3 percentage points (5.8% of control mean) higher efficiency then lines with no supervisors treated. This implies that while the productivity impact of the training is stronger during the lengthy training period, the effects persist after training completion. The decreasing line-level productivity effects over time makes sense given we train individual supervisors, yet analyze productivity at the level of the supervisors' line during randomization. As supervisors can leave the firm or get assigned to new lines (or successful practices can spillover to control lines) over time, the line-level treatment effect from treating supervisors would dampen over time as well.

A main outcome of interest in this paper is the heterogeneity of productivity gains with regards to FFM recommendation (which we investigate in the next section). Given this analysis can only be done on the subset of lines with FFM information, column 4 of Table 5.1 shows that our results also hold in this subset. The effect size after treatment is effectively unchanged, and during treatment coefficients are only about 5% smaller. Further, column 5

[13] We can include this control because within each factory we have lines that are randomly treated, so even after training start (i.e. $r > 0$) we have lines that remain untreated. Our setup is parallel to staggered adoption designs about which there have been an active literature. This literature raises concerns with causal interpretation of two-way fixed effects models. A key feature of our design, however, is that we have randomized variation within each treatment cohort (different factories with different training start dates). Inclusion of the relative time indicator solves the contamination issue as discussed by Abraham and Sun (2020).

[14] We have experimented with using other cutoffs and results are stable under reasonable cutoff values. We further show robustness of our estimates to not dropping these lines.

shows robustness to including both the 0 productivity line-day observations and the lines with many 0 productivity days in our analysis.

Table 5.1: Effects of Training on Line Productivity

	Dependent Variable = Efficiency (Produced/Target)				
	Analysis Lines			Lines w/ FFM Match	All Lines
	(1)	(2)	(3)	(4)	(5)
During Training X Treatment	3.986***	3.994***	4.089***	3.873***	4.208***
	(1.113)	(1.115)	(1.116)	(1.249)	(1.286)
After Training X Treatment	3.079**	3.075**	3.267**	3.258**	3.296**
	(1.364)	(1.366)	(1.338)	(1.564)	(1.632)
Observations	228167	228166	228166	189380	254138
Number of Lines	480	480	480	395	553
Cont. Mean of Dep. Var.	55.865	55.865	55.865	55.865	55.865
Line FE	X	X	X	X	X
Month FE	X				
Day FE		X	X	X	X
Relative Date FE			X	X	X

Standard errors are clustered at line level (*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$). The analysis covers six months prior to training start month and the six months post the training end month for each factory. For columns (1) - (4) days with 0 efficiency are dropped from the analysis as these are reporting errors. Lines for which more than 20% of the days have zero efficiency for any of the three periods are dropped from analysis. Column (5) includes both the dropped lines and the production days with 0 efficiency.

The Appendix Figure D.1 presents the same results in a monthly event study design. First, as discussed before, the event study shows no clear pre-trend that should cause concern for identification. Second, it provides hints about the dynamics of the treatment effect. The treatment effect rises the first 4 months after training start and peaks at around month 4 on average. After that, the treatment effects start getting smaller, but coefficients stay positive for the rest of the analysis period.

Finally, productivity spillovers from training may be present in our context. Adhvaryu et al. (2021) and Adhvaryu et al. (2020) demonstrates worker mobility and sharing across production lines is common in Shahi factories. As our unit of analysis is the production line, if training leads to differential worker mobility patterns across lines due to, for example, changes in efficiency freeing workers to be transferred to lines with high absenteeism, the effects of training may reverberate across neighboring production lines. Similarly, supervisor mobility across lines or the diffusion of successful managerial practices across the production floor can further lead to spillover effects. Our experimental design allows us to check for spillover effects by randomly varying the saturation of training across a subset of the production floors. 237 lines in our analysis sample are in floors with 75% of supervisors treated (high saturation) and 185 lines are in floors with 25% of supervisors treated (low saturation). To check for spillovers, we modify our estimating equation 4 with the triple interaction between treatment, the training periods, and an indicator for whether the line is in a high saturation floor. Appendix Figure D.2 presents results from the regression. Specifically, the figure presents the during/after treatment productivity of (1) treated lines in low saturation floors, (2) control lines in high saturation floors, and (3) the treated lines in high saturation floors relative to the control lines in low saturation floors. The results show evidence for positive spillovers. First, control lines in high saturation floors are have 3 p.p. higher efficiency than control lines in low saturation floors (significant at 10%). Second, post training treatment effect in low saturation lines is a well estimated 0, indicating that long-run treatment effects are driven by treatment in high saturation lines. Existence of positive spillovers imply that the total training effect is higher than what is implied by our average treatment effect calculations. A simple back of the envelope calculation reveals that the total treatment effect calculated from average treatment effects ignoring the spillovers underestimate total gains from training by 25% during training and 32% after training for the 422 lines with randomized saturation.[15]

5.3 Supervisor Retention

Next, we focus on the impacts of training on supervisor retention. As highlighted by our theoretical model, retention results are important in our context for two reasons. First, retention effects can be one factor that contributes to the large productivity effects we find. Line level productivity gains would subsume any of the gains stemming from the retention

[15]This calculation takes the point estimates as given, without accounting for precision. Running estimating equation 4 on the 422 lines with randomized saturation gives average treatment effects of 4.41 pps for during training and 3.34 pps for after training. We calculate total training gains implied by these coefficients as $(237 * 0.75 * 4.41) + (185 * 0.25 * 4.41) = 988$ for during training and $(237 * 0.75 * 3.34) + (185 * 0.25 * 3.34) = 748$ for after training. Taking the spillovers into account, we calculate the analogous values $(237 * 0.75 * 5.57) + (237 * 0.25 * 3.06) + (185 * 0.25 * 3.375) = 1326$ for during training and $(237 * 0.75 * 5.04) + (237 * 0.25 * 3.26) + (185 * 0.25 * 0.052) = 1092$ for after training. We then calculate the ratios $988/1326 = 0.75$ and $748/1092 = 0.68$. Note that this calculation assumes the spillover effects are 0 for lines in low saturation floors.

effects on line supervisors. Second, specifically in our context, focusing on retention provides important clues as to who the FFMs have recommended for the treatment. We discuss this in section 6 where we explore FFM recommendations.

To assess the retention effects of the treatment on supervisors, we estimate a Cox proportional hazard model, taking the randomization strata into account:

$$q_{ti} = h_{0t} \times exp(\mu_s + \beta T_i)$$

where q_{ti} is the hazard function for quitting, μ_s is the strata fixed effect for individual i and T_i is the treatment indicator. We present results for both the full set of supervisors in our study and for the subset of supervisors for whom we have FFM recommendations, as this is the subset we use to assess heterogeneity below. We limit our sample to supervisors who have been with the firm when the training started in their factories. The analysis spans, for each supervisor, day of raining start to end of May 2018.

Table 5.2: Supervisor Retention

	All Supervisors	Supervisors w/ FFM Rec
	(1)	(2)
Treatment	-0.168**	-0.105
	(0.070)	(0.105)
Observations	1419	889
Relative Hazard of Treatment	0.845	0.901
Strata FE	Yes	Yes

*** p<0.01, ** p<0.05, * p<0.10. The sample is restrictied to supervisors that could be matched to the attendance roster and supervisors who did not quit the firm between the baseline survey and the training start in their factories. Column 2 further restricts the analysis to supervisors for which we have FFM recommendations. Relative Hazard is calculated as the exponent of the coefficient on treatment.

Retention results are presented in Table 5.2. Both in the production sample and the full sample, treatment leads to a decrease in the hazard ratio for quitting. For the smaller sample with FFM recommendations, the treated supervisors are 10% less likely to quit between training start and May 2018. While this is not a small effect, it is imprecisely estimated and it is not significant. In the full sample, treated individuals are 15% less likely to quit, and the estimate is statistically significant.

6 FFM Recommendation and Treatment Heterogeneity

Central to our research question is whether the FFMs would choose to allocate the research to supervisors who would gain the most. If FFMs possess private information about who would gain the most from training, they could allocate training to maximize gains. However, FFMs can also have different objectives than the the firm, which would conceivably lead to allocation rules that do not gains from training.

In this section we explore this question. Specifically, we begin by analyzing whether production lines with highly recommended supervisors gain more from training in terms of production efficiency, the outcome the firm would presumably aim to maximize. Then, finding that FFMs recommend supervisor who gain very little in terms of productivity, we ask what drives this outcome. Using our rich baseline observable characteristics, we can partially characterize which supervisors get recommended more. Yet much of the variation remains unexplained by observables. Further, we document that the unobservable component of the recommendation accounts for a significant portion of the negative relationship between recommendation on productivity gains. We then present suggestive evidence that FFMs have targeted supervisors who were more likely to leave the firm and had larger response to training in terms of *retention*. Finally, we review possible alternative explanations (discrimination/nepotism and lack of useful information) and present additional analysis suggesting these explanations are not the main drivers of the patterns we observe.

6.1 Lines With Recommended Supervisors Gain Less in Productivity

In order to see if there is evidence of such private information, we investigate whether productivity gains are higher for supervisors who have been highly recommended by their FFM. To do so, we modify our DiD specification in equation 4 to include three way interactions between treatment, the treatment periods, and whether the line has highly recommended supervisors. In order to go from supervisor level FFM recommendation to the line level, we average the recommendation of every supervisor on the line. We set $1[Rec_l] = 0$ if the line level average is above the median recommendation rank of 3.

Figure 4 charts the treatment effect estimates for the high recommendation and the low recommendation lines from the interaction specification. It is clear that there is significant heterogeneity in how much different supervisors gain from treatment. Interestingly, the supervisors who were highly recommended by their FFMs for training gain significantly less then their low recommendation counterparts. For example, during training, treated lines with low recommendation supervisors experience a 6.2pp (11% relative to baseline) increase in productivity. For lines with recommended supervisors, the corresponding treatment effect is only 1.5pp (3% relative to baseline) and statistically insignificant. 6 months following treatment, lines with highly recommended supervisors effectively gain 0 from training. Appendix Table D.2 reports the underlying regression results with additional robustness checks.

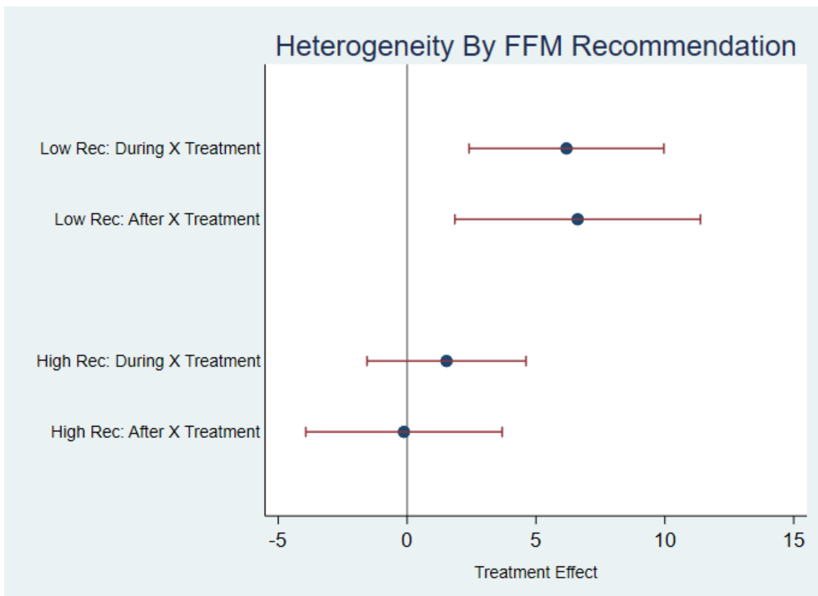


Figure 4: Average treatment effects for high and low recommended supervisors.

6.2 Do the Observable Determinants of FFM Recommendation Explain the Production Heterogeneity?

Why are the highly recommended supervisors gaining less from training? Can the observable characteristics of the recommended supervisors explain either the selection criteria of the FFMs or why they gain less from training.

Given the rich set of baseline information we have about the supervisors, we use a simple lasso procedure to see which variables out of 52 supervisor characteristics and supervisor-FFM joint characteristics (for example, whether they share the same religion or are from the same state) are associated with high recommendations.[16]

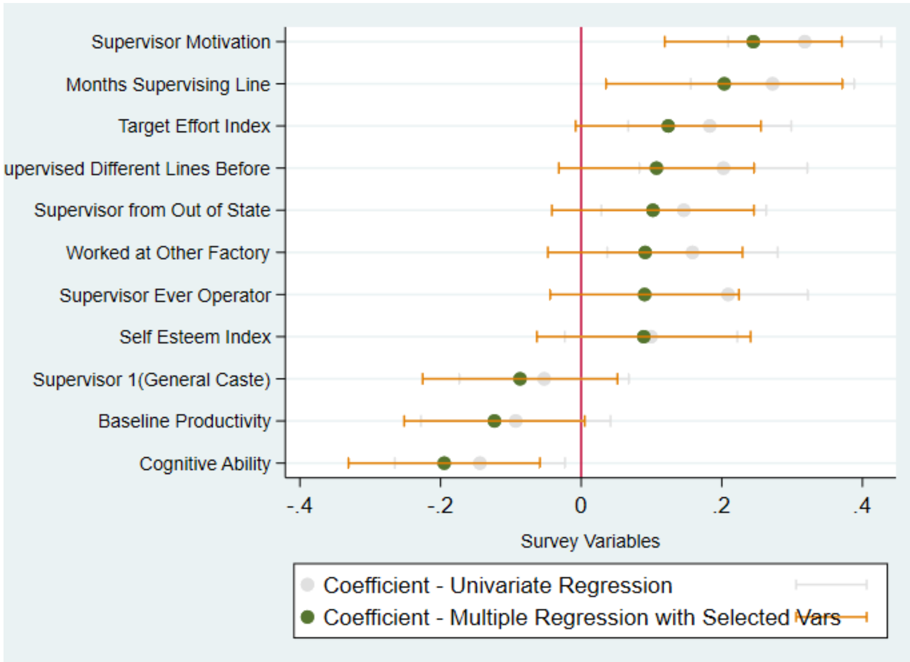


Figure 5: Variables selected from the lasso procedure. The light gray coefficients are from a regression of FFM recommendation on the selected variable of interest. The green coefficients are from a regression of FFM recommendation on all selected variables. 95% CI are shown using robust SEs.

As presented in Figure 5, 11 variables are selected using this procedure. While there is a risk of over interpreting each selected variable, a few patterns emerge. First, variables indicating high tenure and variety of experience predicts higher FFM recommendation. These include months supervising current line, whether the supervisor has worked in a different line or factory before, or whether the supervisor has ever been an operator. Second, FFMs seem to recommend individuals who they view as motivated, as evidenced by not only the high positive coefficient on supervisor motivation, but also the target effort index from the management style survey. Finally, high baseline productivity of the supervisor's line and the supervisor's cognitive ability (measured by arithmetic and digit span recall tests) predicts lower FFM recommendations, implying supervisors may view the training to substitutes to baseline productivity and cognitive ability.

While this analysis suggests FFM are more likely to recommend supervisors who

[16] The full set of included variables and additional details of the procedure can be found in Appendix Section B.4.

are more motivated and experienced, two important questions remain. First, how much of the variation in recommendation scores are explained by this rich set of observables? Second, do the observables shed light on what drives the negative relationship between FFM recommendation and treatment gains?

In order to decompose the negative relationship between FFM selection and treatment gains to observable and unobservable components, we follow the methodology discussed in section 3.[17] Note that the framework underlying this analysis does not take treatment spillovers into account. We proceed with this simplification for two reasons. First, our research question is about ascertaining whether the FFMs identify supervisors who gain the most from training and what factors drive this selection. It is not about optimal roll out scale of the training. While the inclusion of spillovers in our framework would have strong implications about the optimal scale, it is less pertinant to our main research question. Second, our experimental design asks FFMs (who are generally in charge of one floor) to rank their supervisors based on who would gain the most from training. Therefore, in line with our research question, the FFM decision is not about choosing the optimal training scale within the floor, but instead to identify lines/supervisors who would gain the most from training.

The analysis proceeds in two steps. First, we use a probit model to regress FFM selection indicator on a set of observable characteristics. We use the estimates from the first stage to calculate the inverse mill ratio (IMR), denoted $\lambda(X_l, Rec_l)$. We then plug in the inverse mills ratio in the following modification of the estimating equations 4 and 5:[18]

$$y_{ltr} = \alpha'(1[After_{tl}] \times T_l \times X_l) + \rho_{u\eta}\sigma_\eta \times \lambda(X_l, Rec_l) \times 1[After_{tl}] \times T_l + \delta_l + \mu_t + \gamma_r + \epsilon_{ltr} \tag{5}$$

where X_l is observable baseline characteristics and with all lower-level interactions are included. The coefficient on the IMR identifies $\rho_{u\eta}\sigma_\eta$. This term has the same sign as $\rho_{u\eta}$, the correlation between unobserved component of productivity gains (η) and the total unobserved component of FFM selection (u). In words, if the unobserved component of the FFM selection is negatively related to productivity, the coefficient on the IMR should be

[17]While we describe the set up in terms of individual supervisors, we do the analysis at the line level in this section. Therefore, we change the subscripts to l. Observable supervisor characteristics are averaged at the line level in case more than one supervisor is matched to a line.

[18]For this analysis, we collapse the during-training and post-training treatment effects into a single after-treatment-start treatment effect

negative.

First column of Appendix Table D.3 presents the probit results from the first stage of our MTE analysis. For the probit model, along with the 11 variables selected in the lasso analysis, we include additional demographic variables (age, gender, education, local language proficiency), FFM assessment of management skills (industrial engineering skills, technical skills, and management skills), and baseline management style indices from the baseline survey (initiating structure, consideration, active personnel management, and problem index). Despite the rich set of covariates included in the model, the pseudo-R^2 from our model is around 19.5 %. (If we run a linear probability model with the same covariates on individual supervisors, as opposed to line level aggregates, we get an R^2 of 13%). Therefore, we conclude that much of what drives the FFM recommendations remain unobservable to the analyst.[19]

However, the fact that we cannot explain the majority of the variation in FFM recommendations does not necessarily mean that we cannot explain the negative relationship between the recommendation and the treatment effect of training. Observable components of the recommendation decision could still be driving the heterogeneity in treatment effects. To assess this, we turn to the second stage of the MTE analysis. Second column of Table D.3 shows that even after controlling for a rich set of controls, the coefficient on the inverse mill ratio (interacted with treatment and an indicator for after training start) is -2.3. While this coefficient is not precisely estimated with all the covariates, it is nevertheless large and indicates that the unobservable elements FFM recommendation are partially driving the heterogeneity in treatment effects.

We can use the estimated model to obtain predicted treatment effects for production lines and estimate MTEs for different allocation rules and roll-out fractions. In Figure 6, we present three alternative allocation rules following Dal Bó et al. (2021) to further illustrate that both the observed and the unobserved portion of FFM recommendation is negatively related to treatment effects. The figure present three alternative allocation rules: (1) random assignment (solid blue line), (2) assignment based on FFM preference (solid red line), and (3) assignment based only on observable components of FFM preference (dashed red line).[20] The x-axis represents the fraction of lines treated. The slope of the lines at any given roll-out fraction indicates the MTE of the training allocation at that fraction.Dividing the treatment

[19]Due to the aggregation at the line level, some variables that are predictive of FFM recommendation at the individual level are not highly predictive at the first stage of our MTE analysis.

[20]We set the IMR to 0 for the selection to be only driven by observable components.

effect with fraction treated gives the average treatment effect for the allocation rule and the scale.[21] As we would expect from earlier analysis, random assignment outperforms FFM preference by a wide margin. We can observe this by the gap between the blue and the red solid lines. For example, the ATE is approximately 0 with 50% roll-out if training is allocated based on FFM preference, while the average treatment effect under randomization is 2.8 percentage points. Further, the gap between the dashed and solid red line indicates that both the observed and the unobserved components of FFM recommendation is negatively related to treatment gains.

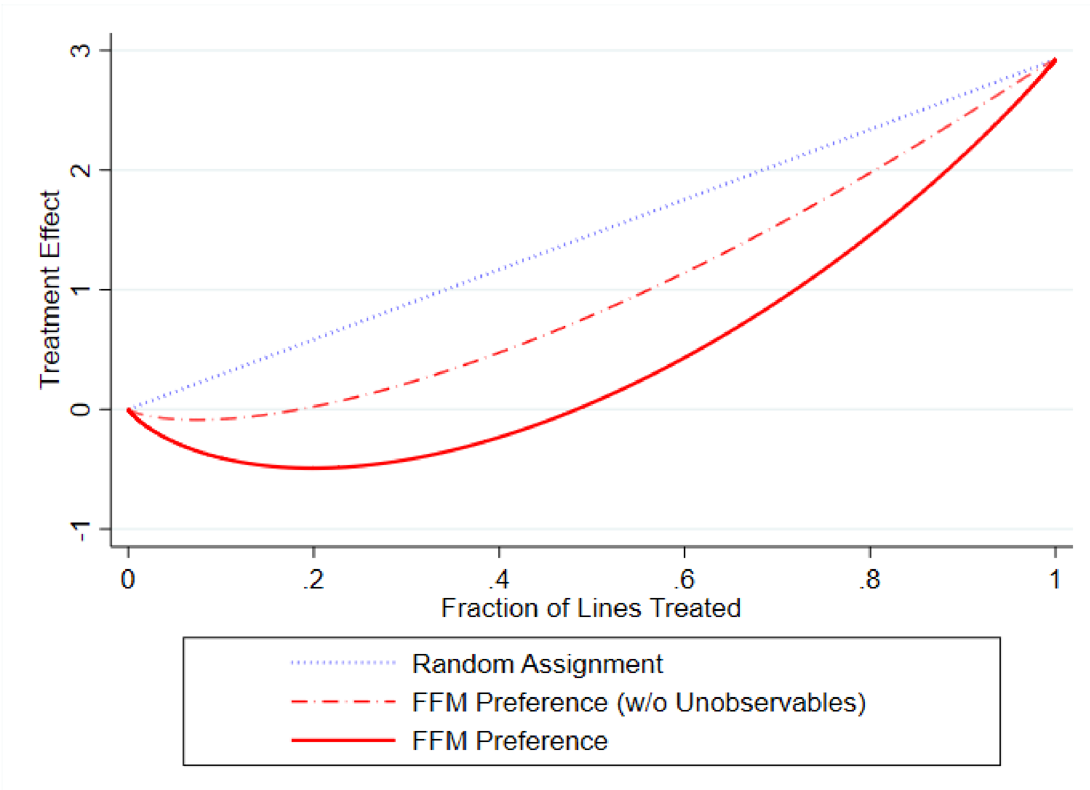


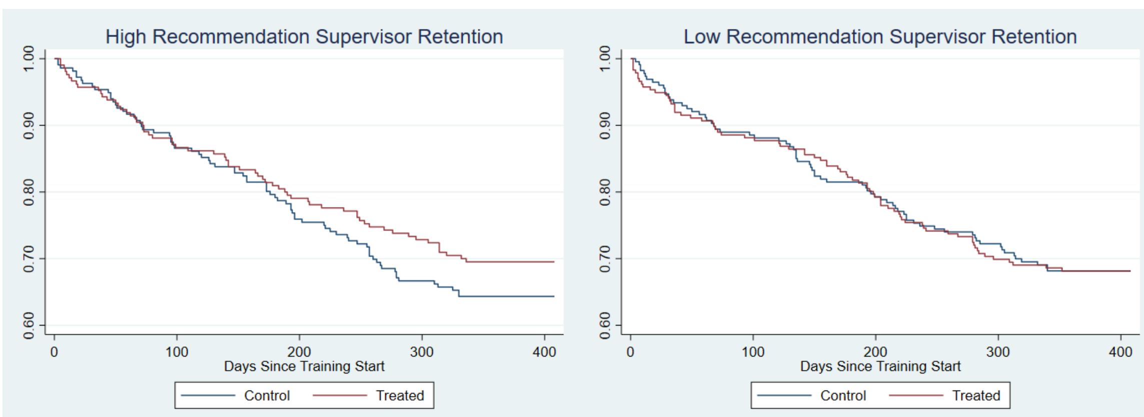
Figure 6: Productivity treatment effects with random allocation and FFM allocation.

[21]For example, given the slope of random assignment line is fixed, the ATE does not change with fraction of lines. This is intuitive as randomization ensures that expected gain of the marginal trainee is constant. Further observe that the ATE is the same for any allocation rule when all of the lines are treated, as the allocation rule does not make a difference once every line is treated. As we noted earlier, we ignore spillovers in this analysis. A framework that takes the positive spillovers in training gains into account would lead to ATE vary with rollout scale.

6.3 FFM Recommendation and Retention

In this section we document suggestive evidence that one dimension that FFMs target with the training is the responsiveness of retention. Where the productivity gains and the retention gains from training are negatively correlated[22] and FFM have incentives to focus on retention, decentralizing the allocation choice can lead to observed productivity gains from training to be low for the recommended supervisors.

To explore this, we plot survival curves (with quitting as the outcome) for treated and control supervisors, separately for high and low recommendation supervisors. Figure 7 presents the results. Figure 7a shows the curves for highly recommended supervisors and Figure 7b include the low recommended supervisors. The retention effect of training is driven by recommended supervisors. The associated treatment hazard ratios for quitting (reported in Appendix Table D.4) are 0.83 for high recommendation supervisors (0.78 if strata are included), and 1 for low recommendation supervisors (0.97 if strata are included). We are underpowered to detect statistically significant changes across the two groups, but the results are suggestive that FFMs target retention effects when allocating the training.



(a) High Recommendation

(b) Low Recommendation

Figure 7: Retention Results for High and Low FFM Recommended Supervisors

We further check if the highly recommended supervisors in the control group have differential quitting patterns. While still statistically insignificant, Figure 8 shows that highly recommended supervisors in the control group are more likely to quit than their

[22]The productivity results explored above do take into account any increase in productivity due to increased *retention*. If the retention effect of training lead to higher productivity, that would by definition would be included in the line level productivity results.

low recommendation counterparts (hazard ratio of quitting is 1.15 [p = 0.29] for the recommended supervisors). We view this, again, as suggestive evidence that FFM recommended supervisors who were more likely to quit and used the training to increase retention of workers who were more likely to respond on the retention margin.

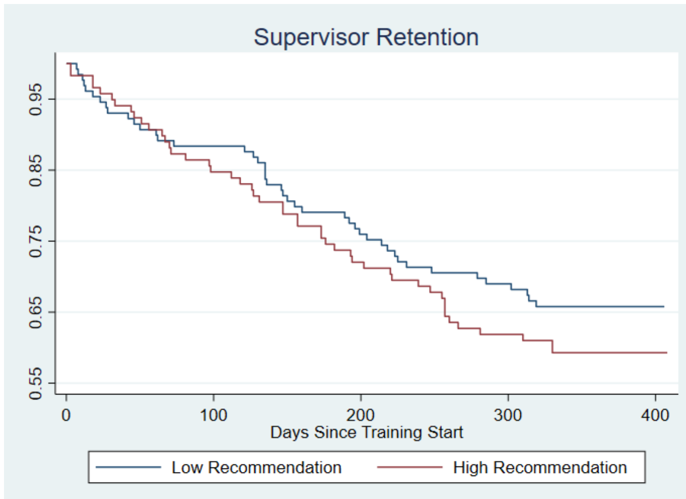


Figure 8: Highly recommended supervisors are more likely to quit the firm.

To further investigate the relationship between the FFM recommendations and retention of supervisors, we apply the MTE analysis discussed above to the retention outcomes. While the production analysis was done using line-level covariates and outcomes in section 6.2, we modify the analysis here to look at supervisor level variables as retention outcomes are at the level of the individual supervisor. Using the estimates of the model (shown in Appendix Table D.5), we look at the MTE of the same alternative allocation rules in Figure 9. The treatment effect in the y-axis of the graph is related to quitting likelihood, so a negative treatment effect means a lower likelihood to quit. Figure 9 makes clear that, as the survival plots above indicated, allocating the training by FFM preferences would lead to higher gains in retention. Using the results in the figure, suppose the firm decided that 50% of the lines are to be treated. The treatment impact on the treated would be a hazard ratio of 0.91 if the firm randomized, and 0.87 if the firm decentralized the decision to the FFMs. Decentralizing the decision outperforms randomization approximately by 35% in terms of quitting likelihood. The figure also shows that the fraction of FFM recommendation that is not explained by observables drive this results (the dashed red line traces the random assignment line fairly closely).

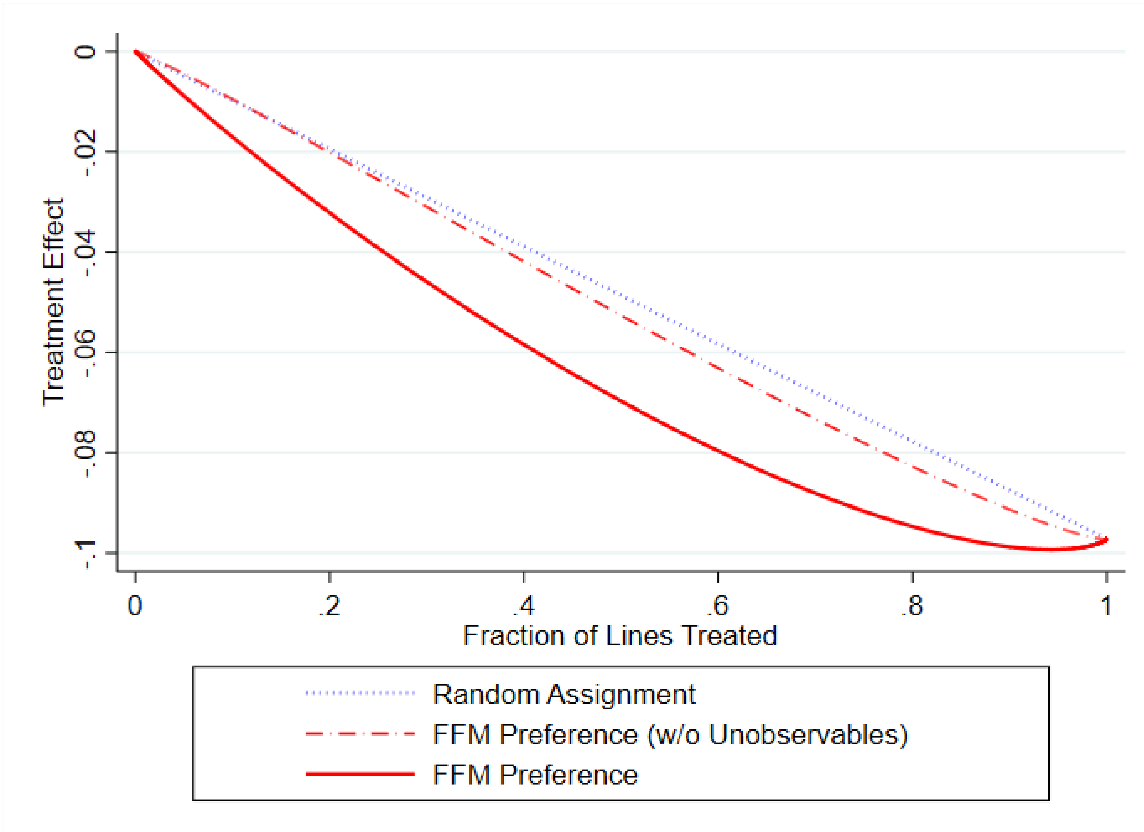


Figure 9: Retention treatment effects with random allocation and FFM allocation

In order to understand the roles of the FFMs in relation to supervisor turnover, we administered a short survey in September 2021 to 50 FFMs and upper managers in 5 factories.[23] While the sample is not representative of our study population, it provides further suggestive evidence that FFMs do bear personal costs when line supervisors turnover. Specifically, FFMs are personally involved in many facets of replacing and onboarding line supervisors. 70% of respondents indicate that FFMs fill in for a departed supervisor before a new supervisor is assigned. 88% of respondents indicate FFMs are involved in the replacement of the line supervisors, where involvement is broadly defined as finding, interviewing, or screening candidates. 88% of respondents also indicate that the FFMs are involed in the training of new supervisors. New supervisors do not learn the necessary skills immidiately. 72% of respondents indicate this process takes one week, while the remaining responded ”two or three weeks” (24%) or ”a month or more” (4%). Finally, all

[23]Of the 50 respondents, 34 are desginated as FFMs. The remaining 16 are Assistant Production Managers or Production Managers, who are above FFMs in the organizational hierarchy.

respondents reported that, from a menu of options, they would provide in kind benefits (such as training or development) programs to retain talented supervisors. To be clear, we do not argue that FFMs are the only employees involved in these processes. The survey results suggest, for example, that HR and upper management are also involved replacing and training supervisors.[24] Similary, 34% indicates floaters or assistant supervisors can be involved in filling in for a supervisor. However, collectively, above patterns indicates that FFMs tend to be involved in each step of filling in for, replacing, and training a supervisor. This is consistent with the idea that FFMs have incentives to target retention above and beyond its effects on productivity.

6.4 Alternative Explanations

Below we explore two alternative explanations for why FFM recommendation is negatively related to productivity gains from training.

Discrimination and Favoritism in FFM Recommendations One explanation for the negative relationship can be that FFMs engage in discrimination based on demographics or favoritism. If the characteristics FFMs discriminate on are negatively related to treatment gains, we could observe the negative relationship we see in the data. While subtle forms of discrimination or favoritism would be indeed hard to capture, we nevertheless don't see strong evidence of discrimination/nepotism in our data. Many demographic characteristics (gender, age, cast etc.) and measures that may relate to favoritism (coincident tenure, whether the supervisor and the FFM started the firm in the same year) are included in the lasso exercise. The only variables related to demographics that are selected in this analysis are whether the supervisor is of general caste (which is associated with lower FFM recommendation) and whether the supervisor is from out of state (which is associated with higher FFM recommendation). In Appendix Table D.6, we directly look for evidence by regressing the FFM recommendation on many demographic and coincident tenure related characteristics. Across the 15 regressors, only whether the supervisor's native language is Kannada (native laguage of the region) is significantly and negatively related to FFM recommendation. The R^2 of the model is only 1.6% and the joint F-statistic is 1.36 ($p = 0.17$). While we show evidence in the paper that recommendation reflects retention concerns, we do not argue that this is the only competing private interest or ulterior motive

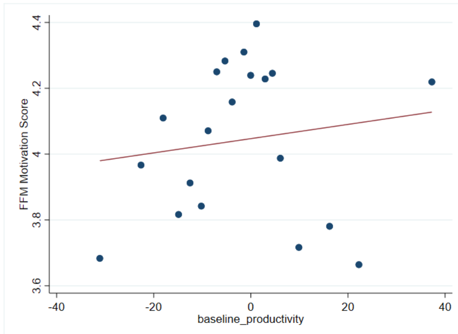
[24] We also ask the respondents to rank the relative involvement of the titles they indicate as involved in a process. HR tends to be ranked lower than FFMs with regards to repalcement and training, while upper management tends to be ranked higher.

for the FFMs when allocating training. Yet, the available measures in our data do not indicate discrimination or favoritism as clear ulterior motives driving the recommendations.

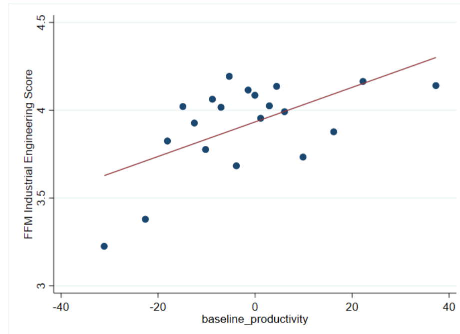
Do FFMs Have Useful Information About Supervisors? An alternative explanation can be that FFMs do not have much useful information on the supervisors which would allow them to allocate the training correctly. One interpretation of this argument would be that FFMs are recommending effectively at random, which is hard to square with the strong negative relationship between recommendation and the productivity gains from training. There is also evidence indicating that FFMs have useful private information about the supervisors. In the FFM baseline survey, we have asked the FFMs to score (from 1 to 5) all the supervisors they list as under them in four dimension: motivation, management skills, industrial engineering skills, and technical skills. In Figure 10, we show the correlation between all four of these scores and the baseline productivity of the line(s) they supervise.[25] The results indicate that these scores are positively correlated with baseline productivity, with the relationship more pronounced for industrial engineering and technical skill scores. Appendix Table D.7 further shows that the association is positive for all four skills and statistically significant for industrial engineering and technical skills. This implies that the FFMs indeed have useful information about the skill sets of their supervisors.

To further investigate the information content of the FFM skill assessments, we use our heterogeneous treatment effect model to see how the productivity treatment effects would evolve if we instead used these skill assessments to allocate the treatment. In a world where FFM assessments do not contain any information, we would expect these alternative allocation rules to not be much different than the random allocation. Our goal here is not to come up with a complicated optimal decision rule using all available information to us, as it would be hard to argue that such a rule would likely be ex-ante obvious to a decision maker. Instead, we look at four simple allocation rules: allocate first to supervisors who lack baseline skills in each four of the skill dimensions FFMs were asked about (motivation, management skill, industrial engineering skill, and technical skill). Each of these decision rules only require the decision maker to have one piece of information about each supervisor: the FFMs skill score for the supervisor. Figure 11 shows the results. We see that allocation based on technical skill has the same MTE as randomizing treatment across rollout scales. Similarly, allocation based on motivation does not have a clean relationship with relative gains. However, allocating to supervisor with low baseline industrial engineering and

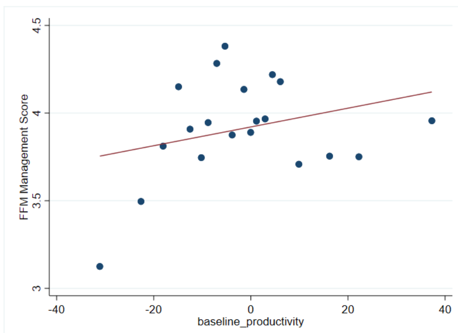
[25]The productivity of the line is calculated following an AKM-style two-way fixed effect model, following Adhvaryu et al. (2019b). Further description of the procedure can be found in Appendix Section B.3



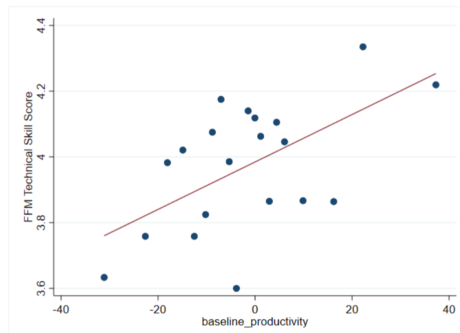
(a) Supervisor Motivation



(b) Industrial Engineering Skills



(c) Management Skills



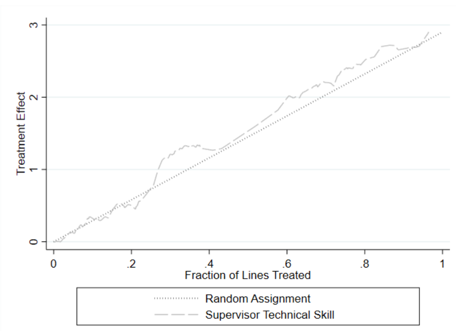
(d) Technical Skill Skill

Figure 10: Binned scatter plots between the FFM assessment of supervisor skill and the line productivity at baseline. Line productivity is calculated from a two-way fixed effect model matching production lines and order styles.

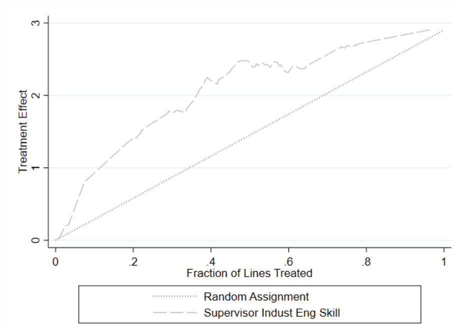
management skills substantially outperforms random assignment in terms of productivity gains. For example, assuming 50% of the lines are to be treated, the treatment affect of allocating based on management skills is about 75% greater than randomizing. This is consistent with the fact that the training is meant to focus on management skills and soft skills. This analysis leads to two conclusions. First, the treatment is a substitute for baseline stock of relevant skills, as opposed to a complement.[26] Second, FFMs do have information that, if used appropriately, can be used to more efficiently allocate scarce training. While we conclude that the FFM indeed have useful private information about the supervisors, this analysis does not rule out the possibility that the FFMs may not have information about what predicts treatment gains.[27]

[26]This is consistent with Adhvaryu et al. (2018)'s conclusions about a similar soft-skills training aimed at sewing workers.

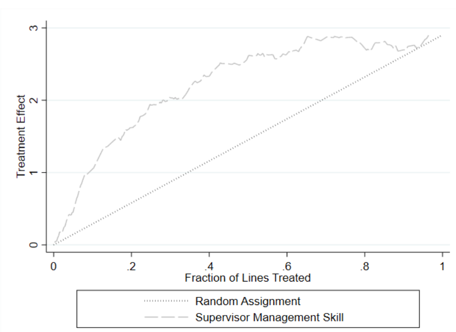
[27]The question that asks the FFM to score their supervisors based on who would gain more from training do specify that the training covers ”communication, leadership, time management, problem-solving and decision-making”.



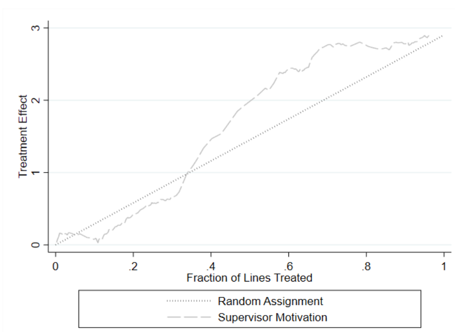
(a) Allocate Based on Technical Skill Score



(b) Allocate Based on Industrial Engineering Skill Score



(c) Allocate Based on Management Skill Score



(d) Allocate Based on Motivation Score

Figure 11: Allocation rules based on FFM assessment of supervisors. Training is allocated first to lines with lowest average supervisor skill scores.

7 Treatment Effects on Other Outcomes

7.1 Supervisor Attendance

Using the attendance roster, we can also analyze the impact of training on daily attendance of supervisors. We use a specification parallel to the difference-in-difference specification in equation 4, except with supervisor-day level observations with daily attendance as the outcome. The results are presented in Appendix Table D.8. We don't find evidence that training increases supervisor retention.

7.2 Supervisor Salary

Appendix Table D.9 presents the results of treatment on salary of supervisors. We regress the percent change in salary from January 2017 to May 2018 (or the latest salary month available if supervisors quit before May 2018) on treatment. We find that treated supervisors experience 0.9 percentage points higher salary growth (7% on a baseline of 12.6 percentage points). In column 2, we additionally control for the number of months that elapses between January 2017 and the latest salary month available (number of months can at most be 18 if the supervisor is still with the firm until May 2018). As the change in salary would be increasing with time before quitting, we control for the number of months in order to control for the retention effects of training. Treatment increases the percent change in salary by 8 percentage points after controlling for retention effects (6 % of baseline)

7.3 Incentive Bonuses

Given the documented productivity effects of the training, we further investigate whether the STITCH training has an impact on the incentive payments which the firm pays out to employees on the basis of performance. To do this, we first aggragate daily data on incentive payments to individual employees to line-day level by summing up the individual payment amounts. We then employ a specification parallel to the difference-in-difference specification in equation 4, with incentive payments as the outcome variable. Specifically, we have two outcomes of interest. First, we ask whether the training has an effect on the extensive margin of bonus payments by looking at an indicator for whether incentive payments have been made on the floor on a given day. Second, we use the inverse hyperbolic sine (IHS) transformation of the payment amount to explore the effects on magnitude of incentive payments.[28] Appendix Table D.10 shows the results, with columns 2 and 4 using our preferred specification. On the extensive margin, we find that the training increases the probability of having any bonus payments on the line by 3 p.p. during and 4 p.p. six months after the training (significant at 10%). These are large magnitudes as they represent a 38% and 51% increase from the control mean. On the intensive margin, we find that lines with all treated supervisors have 27% increase in incentive payments during the training period (not statistically significant), and 39% increase six months after training (significant at 10 %).[29]. In columns 5-8, we replicate the same analysis, but only focus on incentive payments to employees who are not supervisors or managers to assess the impact of training on workers. [30] The results are very similar to the results using the full sample, indicating that the effects accrue to the workers, not just the supervisors who have been trained or to managers. Overall, we conclude that there is suggestive evidence that training has positive impact on incentive payments.

7.4 Worker Attendance and Retention

Finally, we investigate whether workers who were subjected to the treatment through treated supervisors have differential attendance and retention outcomes. Methods used and the results are presented in Appendix D.8. Overall, we do not find evidence for either retention or attendance effects for workers. We caution that this may partly reflect noisy worker-line matches.

8 Conclusion

A recent empirical literature has documented the value of having multiple layers of management (Caliendo et al., 2020, 2015; Caliendo and Rossi-Hansberg, 2012) and decentralizing responsibilities and decisions to lower levels of the hierarchy (Aghion et al., 2021; Bloom et al., 2014; Bloom and Reenen, 2011). They hypothesize that these middle managers may

[28] We use the IHS transformation as there are many line-day observations with 0 incentive payments, which the IHS transformation can handle unlike the log transformation. The results are virtually unchanged if we use $log(1 + payment)$ instead.

[29] Bellemare and Wichman (2019) notes that IHS-linear specifications with dummy variables can be interpreted similarly to log-linear specifications under conditions that our setting satisfies. Therefore, we calculate the approximate percentage changes due to treatment using the formula $e^{\beta} - 1$ where β is the coefficient of interest

[30] Specifically, we exclude any employee whose designation includes the words ”supervisors”, ”manager”, ”senior executive”, and ”floor incharge”.

have some private information and/or specialized understanding that makes them better equipped for making decisions; however, the classic tradeoff is that this decentralization creates a principal-agent structure in which the middle manager may act according to private incentives which do not align perfectly with those of the organization and that limited information at the top of the organization may make enforcing organizational incentives difficult (Acemoglu et al., 2007; Aghion et al., 2014).

To study this exact tradeoff as it relates to the allocation of managerial training within a firm, we solicited from FFMS rankings of which line supervisors should be prioritized for training and randomized access to training within these rankings. We find that line supervisors gained substantial knowledge from the training and productivity of teams managed by trained supervisors increased substantially and persistently on average. However, these productivity gains were quite heterogeneous, with line supervisors recommended highly by FFMs to receive the training actually gaining little to nothing from the training.

On the other hand, training generated a significant positive treatment effect on retention, with these impacts driven entirely by the highly recommended supervisors. In addition high recommendation supervisors in the control group are more likely to quit in the absence of training than are low recommendation control supervisors. We then estimate marginal treatment effects using a recent approach by Dal Bó et al. (2021) to estimate marginal treatment effects (MTE). This analysis confirms that substantial variation (at least 80%) in FFM recommendations derives from unobserved drivers, and that this unobserved component (perhaps most indicative of the private information to be leveraged via decentralization of the training allocation decision) positively predicts improvements in retention despite negatively predicting productivity gains.

Taken all together, the results suggest that FFMs may know which supervisors are most likely to quit and that allocating a training investment of this sort to them may improve their retention. Accordingly, FFMs appear to tailor their training recommendations to take advantage of this potential improvement in retention. We note that the return on investment implied by these net productivity gains is several orders of magnitude larger than any monetary costs borne by the firm to screen and train new supervisors. Accordingly, the firm clearly favors allocating the training to maximize gains in productivity (as would workers and supervisors who all earn significantly greater incentive pay as a result of treatment effects on productivity), but the FFMs have competing incentives to improve line supervisor retention in order to minimize the private burden to them of screening and training replacements and covering the supervisor duties in the interim.

Importantly, the retention of line supervisors which FFMs appears to prioritize is, of course, not without value or importance to the firm, but rather the firm would simply prioritize productivity gains (which deliver orders of magnitude larger returns) when the two priorities are at odds, as turns out to be the case in our scenario. Indeed, our results show that though the average productivity gains from a random allocation were large, persistent, and generated tremendous return on investment, if the supervisors who gained little to nothing had been targeted (as would have been the case if the allocation decision were decentralized to FFMs) the gains and return on investment would have been negligible. In this sense, our results provide one potential explanation for why managerial quality remains low on average in many firms despite strong evidence of potential gains from investments in management such as the training program we evaluate.

References

Acemoglu, D., Aghion, P., Lelarge, C., Van Reenen, J., and Zilibotti, F. (2007). Technology, information, and the decentralization of the firm. *The Quarterly Journal of Economics*, 122(4):1759–1799.

Adhvaryu, A., Bassi, V., Nyshadham, A., and Tamayo, J. A. (2020). No line left behind: Assortative matching inside the firm. Technical report, National Bureau of Economic Research.

Adhvaryu, A., Gauthier, J.-F., Nyshadham, A., and Tamayo, J. (2021). Absenteeism, productivity, and relational contracts inside the firm. Technical report.

Adhvaryu, A., Kala, N., and Nyshadham, A. (2018). The skills to pay the bills: Returns to on-the-job soft skills training. Technical report, National Bureau of Economic Research.

Adhvaryu, A., Kala, N., and Nyshadham, A. (2019a). Management and shocks to worker productivity. Technical report, National Bureau of Economic Research.

Adhvaryu, A., Nyshadham, A., and Tamayo, J. (2019b). Managerial quality and productivity dynamics. Technical report, Mimeo, University of Michigan, Boston College and Harvard Business School.

Aghion, P., Bloom, N., Lucking, B., Sadun, R., and Van Reenen, J. (2021). Turbulence, firm decentralization, and growth in bad times. *American Economic Journal: Applied Economics*, 13(1):133–69.

Aghion, P., Bloom, N., and Van Reenen, J. (2014). Incomplete contracts and the internal organization of firms. *The Journal of Law, Economics, & Organization*, 30(suppl_1):i37–i63.

Bandiera, O., Prat, A., Hansen, S., and Sadun, R. (2020). Ceo behavior and firm performance. *Journal of Political Economy*, 128(4):1325–1369.

Bellemare, M. and Wichman, C. (2019). Elasticities and the inverse hyperbolic sine transformation. *Working Paper*.

Bertrand, M. and Schoar, A. (2003). Managing with style: The effect of managers on firm policies. *The Quarterly journal of economics*, 118(4):1169–1208.

Bloom, N., Eifert, B., Mahajan, A., McKenzie, D., and Roberts, J. (2013). Does management matter? evidence from india. *The Quarterly Journal of Economics*, 1(51):51.

Bloom, N., Lemos, R., Sadun, R., Scur, D., and Van Reenen, J. (2014). Jeea-fbbva lecture 2013: The new empirical economics of management. *Journal of the European Economic Association*, 12(4):835–876.

Bloom, N. and Reenen, J. V. (2011). Human resource management and productivity. *Handbook of labor economics*, 4:1697–1767.

Bloom, N., Sadun, R., and Van Reenen, J. (2010). Does product market competition lead firms to decentralize? *American Economic Review*, 100(2):434–38.

Bloom, N., Sadun, R., and Van Reenen, J. (2016). Management as a technology? Technical report, National Bureau of Economic Research.

Bloom, N. and Van Reenen, J. (2007). Measuring and explaining management practices across firms and countries. *The Quarterly Journal of Economics*, pages 1351–1408.

Caliendo, L., Mion, G., Opromolla, L. D., and Rossi-Hansberg, E. (2020). Productivity and organization in portuguese firms. *Journal of Political Economy*, 128(11):4211–4257.

Caliendo, L., Monte, F., and Rossi-Hansberg, E. (2015). The anatomy of french production hierarchies. *Journal of Political Economy*, 123(4):809–852.

Caliendo, L. and Rossi-Hansberg, E. (2012). The impact of trade on organization and productivity. *The quarterly journal of economics*, 127(3):1393–1467.

Dal Bó, E., Finan, F., Li, N. Y., and Schechter, L. (2021). Information technology and government decentralization: Experimental evidence from paraguay. *Econometrica*, 89(2):677–701.

Frederiksen, A., Kahn, L. B., and Lange, F. (2020). Supervisors and performance management systems. *Journal of Political Economy*, 128(6):2123–2187.

Gosnell, G. K., List, J. A., and Metcalfe, R. D. (2020). The impact of management practices on employee productivity: A field experiment with airline captains. *Journal of Political Economy*, 128(4):1195–1233.

Hoffman, M. and Tadelis, S. (2021). People management skills, employee attrition, and manager rewards: An empirical analysis. *Journal of Political Economy*, 129(1):243–285.

Lazear, E. P., Shaw, K. L., and Stanton, C. T. (2015). The value of bosses. *Journal of Labor Economics*, 33(4):823–861.

APPENDIX

Not for publication.

A Experimental Design Tables and Figures

Figure A.1: Intervention timeline.

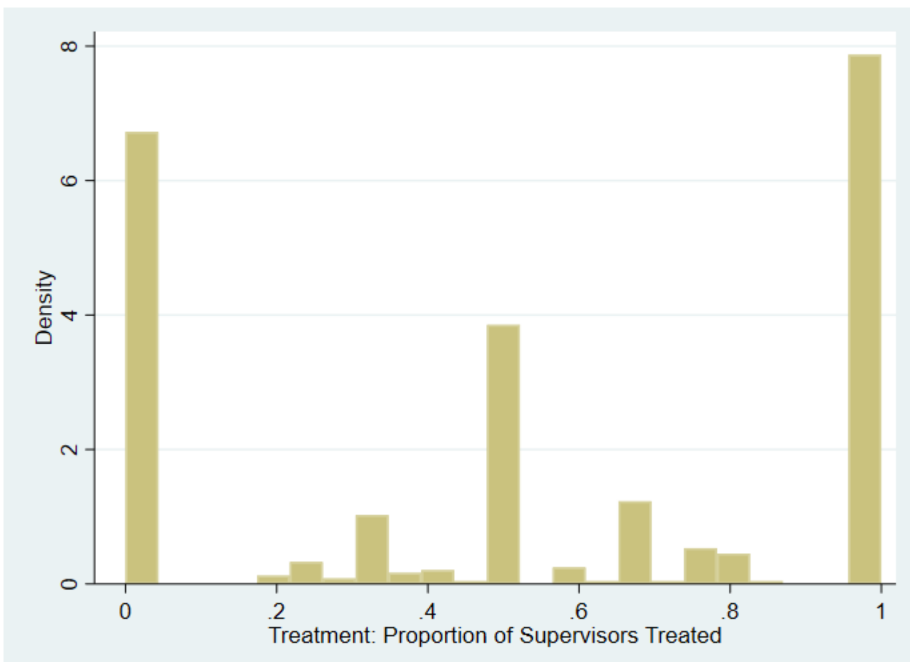


Figure A.2: Distribution of line level treatment, defined as fraction of supervisors treated.

B Model and Analysis Details

B.1 Derivation of Remark

Before we establish remark 1, we show the following lemma holds:

Lemma 1: $\frac{d\tau_p^*}{dc} < 0$, i.e. the productivity gain of the ideal type is decreasing in personal cost c.

Proof: Taking the total derivative of the ideal type equation 2 and reorganizing the terms, we get:

$$\frac{d\tau_p^*}{dc} = -\left(2 - \frac{f''(\tau_p)(1+\delta+\tau_p)}{f'(\tau_p)^2}\right)^{-1}$$

By assumption, $f''(\tau_p) < 0$. We also assume that supervisor gains are distributed such that $\delta + f(\tau_p) \in [0,1]$ since this expression is a probability (probability that a treated supervisor is retained for period 2). As the denominator of the last term is positive, we conclude $\frac{d\tau_p^*}{dc} < 0$. Intuitively, as FFMs incur a higher personal cost from losing supervisors, they shift the training to supervisors with relatively higher retention gains and lower productivity gains.

To show that remark 1 holds, we take the total derivative of the line-level expected productivity gains equation 3 with regards to personal cost c:

$$\frac{d\Delta_p^*}{dc} = \frac{d\tau_p^*}{dc}(f'(\tau_p)(p(1-z) + \tau_p^*) + 1 + \delta + f(\tau_p))$$

Plugging in the ideal type expression from equation 2 for τ_p^*, expression simplifies to:

$$\frac{d\Delta_p^*}{dc} = -\frac{d\tau_p^*}{dc} f'(\tau_p)c < 0$$

The last inequality follows from *lemma 1* and that, by assumption, $f'(\tau_p) < 0$.

B.2 Derivation of Equation 3

The equation follows from standard results on multivariate normal distributions. For recommended supervisors we know the expected productivity gain from training is:

$$\mathbb{E}[\tilde{\Delta}_p^i | X_i, Rec_i = 1] = \beta' X_i + \mathbb{E}[\eta_i | u_i > -\Gamma_i]$$

Using the properties of the normal distribution and that η_i and θ_i (and consequently u_i) are mean 0, we know $\mathbb{E}[\frac{\eta_i}{\sigma_\eta} | u_i = u] = \frac{\rho}{\sigma_u} u$. Combining this with the property $\mathbb{E}[\frac{u_i}{\sigma_u} | \frac{u_i}{\sigma_u} > \frac{-\Gamma_i}{\sigma_u}] = \frac{\phi\left(\frac{-\Gamma_i' X_i}{\sigma_u}\right)}{1-\Phi\left(\frac{-\Gamma_i' X_i}{\sigma_u}\right)}$, we obtain:

$$\mathbb{E}[\eta_i | u_i > -\Gamma_i] = \rho\sigma_\eta \frac{\phi\left(\frac{-\Gamma_i' X_i}{\sigma_u}\right)}{1 - \Phi\left(\frac{-\Gamma_i' X_i}{\sigma_u}\right)}$$

For, non-recommended supervisors, we obtain the parallel result:

$$\mathbb{E}[\eta_i | u_i < -\Gamma_i] = \rho\sigma_\eta \frac{\phi\left(\frac{-\Gamma_i' X_i}{\sigma_u}\right)}{-\Phi\left(\frac{-\Gamma_i' X_i}{\sigma_u}\right)}$$

Combining these two cases yileds the desired result:

$$\mathbb{E}[\tilde{\Delta}_p^i | X_i, Rec_i] = \beta' X_i + \rho\sigma_\eta \frac{\phi\left(\frac{-\Gamma_i' X_i}{\sigma_u}\right)}{Rec_i - \Phi\left(\frac{-\Gamma_i' X_i}{\sigma_u}\right)} \equiv \beta' X_i + \rho\sigma_\eta \lambda(X_i, Rec_i)$$

B.3 Estimation of Baseline Line Productivity

Following the methodology outlined in Adhvaryu et al. (2019b), we estimate the baseline line productivity using a two-way fixed effect model that matches garment styles and production lines. This methodology is parallel to the worker-firm mathching model of Abowd et al. (1999), also known as AKM.

We project line-day level productive efficiency on line, day, and garment style fixed effects.[31] We do this analysis for January 2007 to March 2007, the three months preceding the beginning of training. We use the fixed effect estimates for each line as the baseline productivity of the line.

B.4 Variables Included in the Lasso Procedure

Below, we list the variables included in the lasso procedure. Appendix section B.5 provide further details on the how the personality and management style indices below are created from our surveys.

- **Demographic Variables:** Age (with age squared), Gender, 1[Finished high school], 1[General caste], 1[From out of state], 1[Native language is local language], 1[Hindu]
- **Tenure and Experience:** Tenure in garment industry (months), tenure as supervisor (months), months supervising current line, tenure in Shahi (years), Ever worked as operator, Supervised different line before, Worked at different factory before

[31] Our data includes the garment style a line produces in a given day.

- **FFM and Supervisor Joint Characteristics**: From same state, Same religion, Same gender, Supervisor hired after FFM, Coincident tenure
- **Personality:** Conscientiousness, Locus of Control, Perseverance, Self Esteem
- **Management Style and Practices:** Consideration, Initiating structure, Conflict Index, Problem Index, Autonomous problem solving, target effort index, Monitoring frequency, Communication index, Active personnel management
- **Self Assessment:** Technical skills, Industrial engineering skills, Managerial skills, Training interest, Expected gain from training, Amount supervisor would allocate to training, Self efficacy index, Instrumentality of training
- **FFM Assessment:** Technical skills, Industrial engineering skills, Managerial skills, Motivation, Months supervising current line
- **Other:** Cognitive ability, Risk preference, Discount index, Baseline line productivity, Suggested hires last month

B.5 Creation of Survey Indices

This appendix outlines what questions are used and how they are combined for the creation of the indices from the baseline supervisor surveys.
[Index Table Here]

Table C.1: Line Level Descriptive Statistics and Balance for Analysis Subsets

	Analysis Subsample				Analysis Subsample w/ FFM			
	Num Lines	Mean	SD	Coefficient/SE	Num Lines	Mean	SD	Coefficient/SE
Baseline Productive Efficiency	476	55.88	13.01	-3.143**	393	55.71	12.80	-1.625
				(1.513)				(1.695)
Baseline Attendance	471	0.90	0.05	-0.009	393	0.90	0.05	-0.011*
				(0.006)				(0.006)
Baseline Retention	465	0.84	0.13	0.010	389	0.84	0.12	-0.013
				(0.016)				(0.016)
Baseline SAM	476	55.79	20.41	-1.638	393	55.71	20.27	2.268
				(2.462)				(2.829)
Baseline Budgeted Efficiency	476	60.97	7.17	-0.377	393	60.27	7.17	-1.468
				(0.841)				(0.966)
Baseline Number of Operators	476	171.32	171.75	-6.450	393	168.37	172.37	-21.155
				(18.931)				(21.741)

The left panel excludes days with 0 efficiency and excludes lines that record over 20% of the days as 0 efficiency before, during or after training. The right panel further reduces the sample to lines we have FFM recommendations for. The coefficient(SE) is from regressing the outcome on the continuous treatment indicator. Robust standard errors are reported (* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.). All baseline values are from 3 months preceding training start (January - March 2017). Baseline (budgeted) efficiency is an average of daily (budgeted) efficiency values for this period. Baseline attendance and retention are the attendance and retention outcomes for the workers we matched to these lines using the personnel rosters.

C Additional Checks and Robustness

C.1 Additional Line Balance

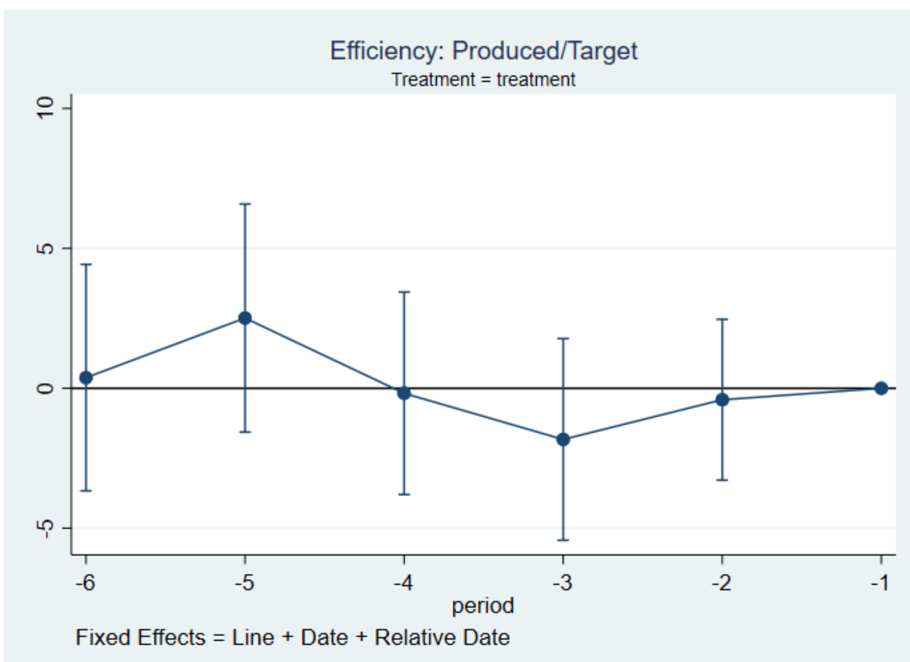


Figure C.1: The event study style model is run for the entire analysis period. The coefficients for months preceding training is shown to focus on pre-trends.

D Additional Results

D.1 Pre-Post Test Scores

To assess the effects of training on learning the module content, we present the results of the following ANCOVA specification:

$$s_{i2} = \beta_0 + \beta_1 T_i + \beta_2 s_{is1} + \mu_s \tag{6}$$

where s_{i2} is the post-module test score in percentage points of supervisor i, s_{i1} is the pre-module test score, T_i is whether the supervisor is randomized into treatment, and μ_s is strata fixed effects. Results for all 4 modules are presented in Table D.1. Across all modules, treatment leads to a significant gain in the post-module tests.

Table D.1: Treatment Effect on Post-Module Exam Scores

	Dependent Variable = Post-Module Test Score			
	Module 1	Module 2	Module 3	Module 4
	(1)	(2)	(3)	(4)
Treatment	22.130***	22.048***	32.248***	38.799***
	(1.440)	(0.846)	(3.603)	(2.216)
Observations	623	574	553	541
Control Mean of Dependentent Variable	48.246	54.605	31.579	35.714
Strata FE	Yes	Yes	Yes	Yes

Robust standard errors in parentheses (*** p<0.01, ** p<0.05, * p<0.1). Specification includes controls for the pre-module test scores. Test scores are in percentage points.

D.2 Productivity Treatment Effect, Heterogeneity, and MTE Analysis

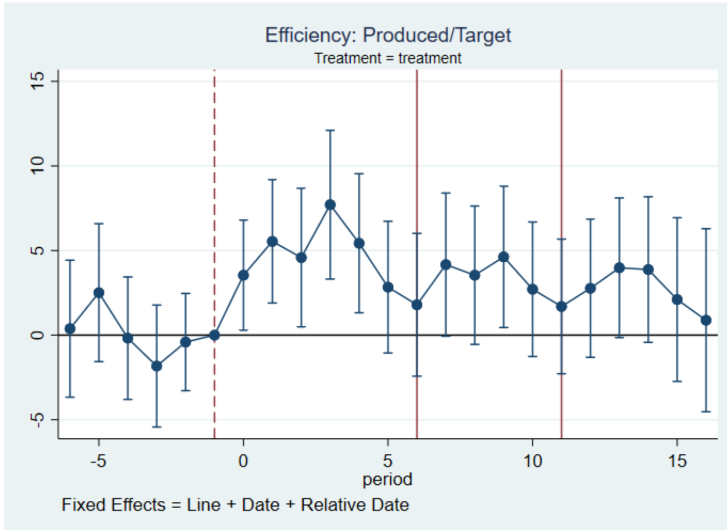


Figure D.1: Month 0 signifies treatment start. Shortest training duration is 6-months (first solid red line). Longest training duration is 11 months (second solid red line). Coefficient on month 17 dropped for readability (coefficient[SE] = 6.87[5.51]).

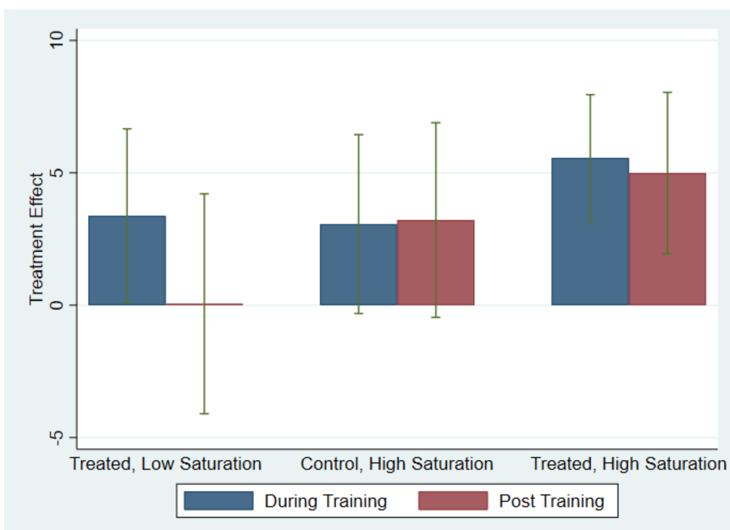


Figure D.2: Treatment effects relative to control lines in low saturation production floors. 95% confidence intervals shown. Standard errors are clustered at line level.

Table D.2: Productivity Effect Heterogeneity by FFM Recommendation

	Dependent Variable = Efficiency (Produced/Target)			
	Analysis Lines			All Lines
	(1)	(2)	(3)	(4)
During Training X Treatment	6.023***	6.015***	6.183***	6.832***
	(1.951)	(1.955)	(1.928)	(2.056)
After Training X Treatment	6.414***	6.383**	6.617***	6.350**
	(2.466)	(2.470)	(2.431)	(2.556)
During Training X High Rec X Treatment	-4.456*	-4.428*	-4.654*	-5.946**
	(2.502)	(2.506)	(2.489)	(2.814)
After Training X High Rec X Treatment	-6.458**	-6.408**	-6.739**	-5.993*
	(3.168)	(3.172)	(3.118)	(3.523)
Observations	189381	189380	189380	208691
Number of Lines	395	395	395	444
Control Mean of Dependentent Variable	55.279	55.279	55.279	55.279
Line FE	X	X	X	X
Month FE	X			
Day FE		X	X	X
Relative Date FE			X	X

Standard errors are clustered at line level (*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$). The analysis covers six months prior to training start month and the six months post the training end month for each factory. For columns (1) - (3) days with 0 efficiency are dropped from the analysis as these are reporting errors. Lines for which more than 20% of the days have zero efficiency for any of the three periods are dropped from analysis. Column (4) includes both the dropped lines and the line-days with 0 efficiency. ”High rec” is an indicator for whether the line has average supervisor recommendation above the median.

Table D.3: Selection and Production Effect Heterogeneity

	First Stage 1[High Rec] (1)	Second Stage Efficiency (2)
Treatment (ATE)		2.769**
		(1.298)
Inverse Mills Ratio		-2.309
		(1.479)
Age	-0.039*	0.716**
	(0.023)	(0.294)
1(Male)	-0.271	2.944
	(0.255)	(3.406)
1(Finished Highschool)	0.182	-11.933*
	(0.439)	(6.514)
Local Language Proficiency	-0.242	-4.062
	(0.221)	(3.017)
Supervised Dif Line Before	0.626***	-4.405
	(0.224)	(2.763)
Ever Worked as Operator	0.493	-2.592
	(0.317)	(4.220)
Ever Worked at Another Factory	0.025	3.048
	(0.236)	(2.979)
Months as Supervisor	-0.001	0.028
	(0.003)	(0.043)
Months Supervising Current Line	0.001	-0.053
	(0.005)	(0.070)
Years in Shahi	0.026	-0.334
	(0.024)	(0.340)
Motivation (Scored by FFM)	0.243*	0.574
	(0.129)	(1.921)
Months as Supervisor (Answered by FFM)	-0.001	-0.108*
	(0.004)	(0.057)
Target Effort Index	0.379***	-5.551***
	(0.124)	(1.609)
Cognitive Ability	-1.212**	-7.059
	(0.519)	(6.955)
Technical Skills (Scored by FFM)	-0.041	3.136
	(0.129)	(2.188)
Industrial Engineering Skills (Scored by FFM)	-0.180	-1.521
	(0.142)	(2.065)
Management Skills (Scored by FFM)	0.010	-2.685
	(0.138)	(2.166)
Self Esteem	0.653***	4.602*
	(0.206)	(2.775)
Initiating Structure	0.016	-0.394
	(0.023)	(0.403)
Consideration	-0.039	0.843**
	(0.032)	(0.399)
Active Personnel Management	0.018	2.236
	(0.118)	(1.554)
Problem Index	-0.165	0.789
	(0.114)	(1.372)
Baseline Productivity of Line	-0.005	0.120
	(0.005)	(0.080)
1(From Different State)	0.520*	-4.722
	(0.308)	(4.684)
1(General Caste)	-0.297	1.958
	(0.221)	(3.095)
Observations	379	
Pseudo R-sq	0.197	

Note: *** p<0.01, ** p<0.05, * p<0.10. The first column shows the results from the first stage probit model for line level FFM selection. Second column presents results from the heterogeneous treatment effect regression. For column 2, all shown coefficients are for the triple interaction of variable of interest with treatment and 1[After Training Start].

D.3 Retention Treatment Effect, Heterogeneity, and MTE Tables

Table D.4: Retention Effects by FFM Recommendation

	Dependant Variable = Quitting	
	High Recommendation	Low Recommendation
	(1)	(2)
Treatment	-0.252	-0.034
	(0.222)	(0.117)
Observations	426	463
Relative Hazard of Treatment	0.778	0.966
Strata FE	X	X

*** $p < 0.01$, **$p < 0.05$, * $p < 0.10$. The sample is restrictied to supervisors that could be matched to the attendance roster and supervisors who did not quit the firm between the baseline survey and the training start in their factories. First column limits the sample to supervisors with high FFM recommendation. Second column limits the sample to supervisors with low FFM recommendation.

Table D.5: Selection and Retention Effect Heterogeneity

	First Stage FFM Selection (1)	Second Stage 1[Quit] (2)
Treatment		-0.103
		(0.098)
Inverse Mills Ratio		-0.047
		(0.146)
Age	-0.001	0.018
	(0.008)	(0.020)
1(Male)	-0.120	0.317
	(0.108)	(0.357)
1(Finished Highschool)	0.186	0.082
	(0.172)	(0.477)
Local Language Proficiency	0.085	0.440**
	(0.087)	(0.215)
Supervised Dif Line Before	0.199**	-0.490
	(0.094)	(0.486)
Ever Worked as Operator	0.507***	-0.438**
	(0.112)	(0.220)
Ever Worked at Another Factory	0.260**	0.445
	(0.107)	(0.366)
Months as Supervisor	-0.000	0.004
	(0.001)	(0.003)
Months Supervising Current Line	-0.001	-0.012***
	(0.001)	(0.004)
Years in Shahi	-0.003	-0.001
	(0.011)	(0.040)
Motivation (Scored by FFM)	0.273***	-0.155
	(0.065)	(0.153)
Months as Supervisor (Answered by FFM)	0.004***	0.008
	(0.002)	(0.005)
Target Effort Index	0.005	-0.316***
	(0.051)	(0.093)
Cognitive Ability	-0.374	-0.586
	(0.239)	(0.704)
Technical Skills (Scored by FFM)	-0.159**	0.390
	(0.065)	(0.246)
Industrial Engineering Skills (Scored by FFM)	-0.151**	-0.153
	(0.067)	(0.167)
Management Skills (Scored by FFM)	0.084	0.356**
	(0.066)	(0.167)
Self Esteem	0.122	-0.372
	(0.086)	(0.408)
Initiating Structure	0.024**	0.051**
	(0.011)	(0.024)
Consideration	-0.028**	0.004
	(0.013)	(0.032)
Active Personnel Management	-0.008	0.082
	(0.048)	(0.100)
Problem Index	-0.018	0.132
	(0.046)	(0.115)
1(From Different State)	0.110	0.610*
	(0.130)	(0.331)
1(General Caste)	0.001	0.527**
	(0.093)	(0.207)
Observations	867	866
Pseudo R-sq	0.078	

Note: *** p<0.01, ** p<0.05, * p<0.10. The first column shows the results from the first stage probit model for supervisor level FFM selection. Second column presents results from the heterogeneous treatment effect cox regression for retention. For column 2, all shown coefficients are for the interaction of variable of interest with treatment indicator.

D.4 Alternative Explanations

Table D.6: Recommendation is not well explained by demographic and favoritism related variables

	FFM Recommendation (1)
Supervisor Age	0.015 (0.063)
Supervisor Age Squared	-0.000 (0.001)
Supervisor 1(Male)	-0.161 (0.276)
Supervisor 1(Hindu)	0.666 (0.442)
Supervisor Native Language is Kannada	-0.394*** (0.144)
Supervisor from Different State	0.173 (0.144)
Supervisor 1(General Caste)	-0.084 (0.107)
Sup and FFM Same Gender	0.137 (0.273)
Sup and FFM Same Age Group	0.111 (0.134)
Sup and FFM Same Caste	0.119 (0.106)
Sup and FFM Same Religion	-0.368 (0.362)
Sup and FFM Coincident Tenure (Years)	-0.003 (0.015)
Supervisor Hired After FFM	-0.005 (0.117)
Sup and FFM Same Cohort	-0.163 (0.232)
Constant	2.711** (1.067)
Observations	1051
R Sq.	0.016
F-stat	1.359

Note: Robust standard errors in parantheses. *** $p<0.01$, ** $p<0.05$, * $p<0.10$.

Table D.7: Baseline Productivity and FFM Assessment of Skills

	Dependent Variable = Baseline Productivity				
	(1)	(2)	(3)	(4)	(5)
Supervisor Technical Skills	2.667***				1.018
	(0.880)				(1.346)
Supervisor Industrial Engineering Skills		3.130***			3.737**
		(0.931)			(1.590)
Supervisor Management Skills			1.536		-1.483
			(0.935)		(1.437)
Supervisor Motivation				0.821	-0.239
				(1.061)	(1.203)
Constant	-12.465***	-14.126***	-7.880**	-5.193	-13.783***
	(3.549)	(3.746)	(3.867)	(4.473)	(4.893)
Observations	393	393	393	393	393
R Sq.	0.019	0.031	0.008	0.002	0.036
F-Statistic					3.725

Note: *** p<0.01, ** p<0.05, * p<0.10. Coefficients are from a linear regression of baseline productivity on the specific FFM assessment on skill. FFM assessment of supervisor skills are aggregated at the line level by taking the average of all the line supervisors. Baseline productivity is calculated as described in Appendix Section B.3.

D.5 Supervisor Attendance

We assess the day-supervisor level retention effects of training using the following DiD specification:

$$1[Attended]_{itr} = \alpha + \beta_1 T_i \times 1[During_t] + \beta_2 T_i \times 1[Post_t] + \delta_l + \mu_t + \gamma_r + \epsilon_{itr} \qquad (7)$$

where $1[Attended]_{itr}$ is an indicator for whether supervisor i attended work on date t, T_i is the treatment indicator, and the $1[During_t]$ and $1[Post_t]$ are indicators for whether training is ongoing or over in the factory of the supervisor. The results are shown in Append Table D.8. We do not observe any evidence of treatment effects on supervisor retention.

Table D.8: Treatment Effects on Supervisor Attendance

	Dependant Variable = Daily Attendance		
	(1)	(2)	(3)
Treatment X During	0.001	0.001	0.001
	(0.005)	(0.005)	(0.005)
Treatment X Post	0.004	0.004	0.003
	(0.009)	(0.009)	(0.009)
Observations	516805	516805	516805
Number of Supervisors	1636	1636	1636
Control Mean of Dependentent Variable	.895	.895	.895
Supervisor FE	X	X	X
Date FE		X	X
Relative Date FE			X

Note: *** p<0.01, ** p<0.05, * p<0.10.

D.6 Supervisor Salary

To assess the effects of treatment on salary growth, we use the following specification:

$$\%growth_i = \alpha + \beta_1 T_i + \beta_2 NumMonths_i + \mu_s + \epsilon_i$$

where $\%growth_i$ is the percent change in gross salary between January 2017 and May 2018 (or the latest month observed) for the supervisor i, T_i is the treatment indicator, and $NumMonths_i$ is the number of months after January 2017 the supervisor is in the data (with a maximum of 18 if the supervisor is with the firm until May 2018). Results are reported in Appendix Table D.9.

Table D.9: Treatment Effects on Salary Progression

	Dependent Variable = Salary Change (1)	(2)
Treated	0.009**	0.008**
	(0.004)	(0.004)
Num. Months Before Quitting		0.014***
		(0.001)
Observations	1411	1411
Control Mean of Dependentent Variable	.126	.126
Strata FE	X	X

Note: *** $p<0.01$, ** $p<0.05$, * $p<0.10$. The monthly salary data covers January 2017 to May 2018. For each supervisor, the percent change in salary is calculated as the percent change from the earliest to latest gross salary recorded. Supervisors who quit between January 2017 and training start are dropped from the analysis.

D.7 Incentive Bonuses

Appendix Table D.10 presents results on incentive payments.

Table D.10: Treatment Effects on Incentive Payments

	Sample: All Employees				Sample: Non - Supervisors			
	1[Any]		IHS(Amount)		1[Any]		IHS(Amount)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
During X Treatment	0.034*	0.031*	0.269*	0.241	0.033*	0.031	0.263*	0.236
	(0.019)	(0.019)	(0.155)	(0.155)	(0.019)	(0.019)	(0.154)	(0.154)
After X Treatment	0.047*	0.041*	0.377*	0.333*	0.046*	0.041*	0.372*	0.328*
	(0.024)	(0.024)	(0.200)	(0.199)	(0.024)	(0.024)	(0.198)	(0.197)
Observations	270661	270661	270661	270661	270661	270661	270661	270661
Num. Lines	476	476	476	476	476	476	476	476
Cont. Mean	.081	.081	.65	.65	.081	.081	.646	.646
Line FE	X	X	X	X	X	X	X	X
Day FE	X	X	X	X	X	X	X	X
Relative Day FE		X		X		X		X

Note: *** p$<$0.01, ** p$<$0.05, * p$<$0.10. Standard errors are clustered at line leve. $1[Any]$ indicates any incentive payments have been paid in the line on a given day. $IHS(Amount)$ is the inverse hyperbolic sine transformation of the total incentive payments in the line on a given day.

D.8 Worker Retention and Attendance

Appendix Figure D.3 shows survival curves for quitting for workers in lines with at least one supervisor treated versus none. There is no evidence of differential retention. Running a cox proportional hazard model with the preferred treatment definition of fraction of supervisors treated also yields no evidence of differential retention. For attendance, we follow an analogous approach to equation 7 for estimating the treatment effects on worker attendance, except with continuous line-level treatment. We do not find any effects.

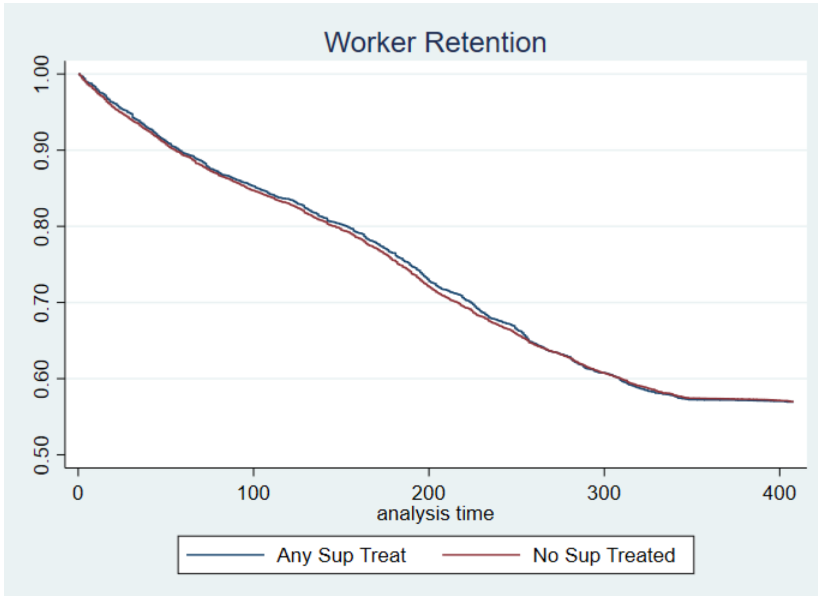


Figure D.3: Worker retention by having any treated supervisor on line.

Table D.11: Treatment Effects on Worker Attendance

	Dependant Variable = Daily Attendance			
	(1)	(2)	(3)	(4)
During Training X Treatment	-0.001	-0.001	0.002	-0.000
	(0.004)	(0.004)	(0.004)	(0.002)
Post Training X Treatment	0.005	0.005	0.009	-0.001
	(0.006)	(0.006)	(0.005)	(0.004)
Observations	10864000	10864000	10864000	10863731
Cont. Mean of Dep. Var.	.86	.86	.86	.86
Line FE	X	X	X	
Employee FE				X
Day FE		X	X	X
Relative Day FE			X	X

Note: Standard errors are clustered at line level. *** $p<0.01$, ** $p<0.05$, * $p<0.10$. Coefficients are from a linear probability model on whether the employee has attended work on a given day. Sundays and days where less than 40% of employees attend work are dropped from the analysis.