On the Allocation and Impacts of Managerial Training*

Achyuta Adhvaryu Emir Murathanoglu Anant Nyshadham

November 28, 2022

Abstract

We study the allocation and productivity consequences of managerial training via a randomized controlled trial among production line supervisors in a large ready-made garment firm. We designed a program using practices identified as productive in Adhvaryu et al. (2022c), and asked middle managers – who are directly above production line supervisors in the hierarchy – to recommend which of the supervisors they manage should be prioritized for training. We then randomized access to the program within these recommendation rankings. Productivity on lines managed by treated supervisors increased by 6-7% relative to control, but these gains exhibit substantial heterogeneity across middle manager ranking categories. Highly recommended supervisors experienced no productivity gains; the average treatment effect of training is driven entirely by low-recommendation supervisors. This was not due to a lack of information about baseline skills or about who would gain the most, nor to discrimination or favoritism along observable dimensions. Instead, consistent with the fact that supervisor turnover has large personal costs for middle managers in terms of labor substitution and onboarding, middle managers prioritized the retention impacts of training. Treated supervisors were 15% less likely to quit than controls over the study period, and this gain was most pronounced for highly recommended supervisors. We show that middle manager recommendations leveraged private information, and that these unobserved factors negatively predict productivity effects while positively predicting retention effects. Heterogeneous returns and the unproductive allocation of costly training can thus help explain underinvestment in attenuating persistent within-firm gaps in managerial quality.

Keywords: managerial quality, training, productivity, soft skills, decentralization, misallocation
JEL Codes: J24, L23, M53

*Adhvaryu: University of Michigan, BREAD, NBER, J-PAL, William Davidson Institute & Good Business Lab; adhvaryu@umich.edu; Murathanoglu: University of Michigan; emirmur@umich.edu; Nyshadham: University of Michigan, BREAD, NBER, J-PAL & 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. We are grateful to David Atkin, David Autor, Daron Acemoglu, Abhijit Banerjee, Jere Behrman, Hoyt Bleakley, Esther Duflo, Hanming Fang, Robert Garlick, Mitchell Hoffman, Amit Khandelwal, Tommaso Porzio, Andrea Prat, Debraj Ray, Petra Todd, John Van Reenen, Eric Verhoogen, Shing-Yi Wang, and Dean Yang for useful discussions and comments. We also thank seminar participants at Columbia University, Duke University, Massachusetts Institute of Technology, National University of Singapore, New York University, University of California San Diego, University of Chicago, University of Michigan, University of Pennsylvania, and University of Virginia. The views expressed herein do not represent IGC, or Shahi Exports. All errors are our own.

1 Introduction

Variation in managerial quality contributes to the vast dispersion in productivity across countries and firms (Bandiera et al., 2020; Bloom et al., 2016; Bloom and Van Reenen, 2007, 2011), and even within firms across teams and workers (Bertrand and Schoar, 2003; Lazear et al., 2015). Many skills and practices have been identified as key to the productive value of managerial quality (Adhvaryu et al., 2022c), including people skills, e.g., the subjective evaluation of subordinates and effectively allocating workers to tasks (Adhvaryu et al., 2022a; Frederiksen et al., 2020; Hoffman and Tadelis, 2021). Given their importance, why then do managerial skill gaps persist? In particular, what prevents training – which is perhaps the most common lever used by organizations of all kinds to upgrade managerial practices – from successfully closing these skill gaps?

We contend that the persistence of managerial skill gaps is due in part to the potential for misallocation of – and consequently the low perceived returns to – training within the firm. Training is a scarce resource, which is allocated endogenously within organizations. Thus, *who* allocates training – and how closely the objectives of these agents are aligned with the firm's – is of paramount importance. Even if the true average treatment effect of managerial training on productivity is large for the population of managers in a firm, realized impacts could be small if returns are heterogeneous enough and if training tends to be allocated to the "wrong" managers. This concern is perhaps most salient when firms try to pilot costly investments on a subset of workers or teams to inform full-scale adoption decisions. In these instances firms rarely if ever conduct a lottery or other random assignment to choose which workers or teams will participate in the pilot, but rather explicitly ask middle managers who work most closely with the eligible candidates to nominate pilot participants. This was certainly the modal approach employed by our partner firm, and the primary motivation for our study design.

This practice harkens to the literature on the tradeoffs of decentralization of decision-making within organizations. Having multiple layers of management can be valuable (Caliendo et al., 2020, 2015; Caliendo and Rossi-Hansberg, 2012), particularly if it is possible to delegate some responsibilities and decisions to lower levels of the hierarchy (Aghion et al., 2021; Bloom et al., 2014; Bloom and Van Reenen, 2011). The argument is that middle managers may have private information or specialized understanding that makes them better equipped to make optimal resource allocation decisions. On the other hand, the classic trade-off is that this decentralization creates a principal-agent structure in which the middle manager will act according to her own incentives, which may not perfectly align with those of the organization, and that limited information at the top may make achieving first-best decision-making impossible (Acemoglu et al., 2007; Aghion et al., 2014).

In this study, we ask: what are the impacts of managerial training? And would middle managers allocate training to maximize productivity returns as the firm would want, or is their decision-making driven by other considerations? We seek to answer these questions via a randomized controlled trial in which we designed and implemented a management training program for production line supervisors in Indian garment factories. The program curriculum, which was designed by the authors along with the firm partner, emphasizes particular soft skills – such as communication, planning and organization, problem-solving, and motivation of workers – that were found to be strongly related to productivity in a nearly identical setting by Adhvaryu et al. (2022c). To study questions regarding allocation, we asked middle managers to rank the supervisors in their charge on who should be prioritized for training. We then randomized access to the training within these rankings, allowing us to recover average treatment effects as well as to study whether middle managers did indeed recommend the supervisors who ultimately gained the most from training.

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. Productivity of teams managed by trained supervisors increased substantially and persistently on average compared to controls by 7.3 percent during training and 5.8 percent over the six months after training completion. However, these average gains belie substantial heterogeneity across middle manager ranking categories. Line supervisors recommended highly by middle managers to receive the training actually experienced no significant productivity gains; the average treatment effect is driven entirely by low-recommendation supervisors.

This striking pattern of results is not due to middle managers lacking information about their supervisors baseline skills, or about who might benefit most from training. It is also not due to discrimination or favoritism on observable dimensions. We learned from conversations with middle managers as well as senior management that while middle managers are indeed incentivized to achieve high productivity (through performance-based rewards), they also face large personal effort costs from supervisor turnover. In particular, middle managers are charged with the training and onboarding of new supervisors and performing supervisory duties in the interim, all of which involves substantial effort in addition to their other day-to-day responsibilities. We administered a short survey to middle managers the results of which confirm the notion that supervisor turnover is a salient personal cost to most middle managers, and that they use training programs as in-kind benefits to help retain supervisors who they felt were close to quitting.

Accordingly, we set out a simple model to understand the tradeoffs middle managers face in allocating training resources. The effort of replacing supervisors who have quit and bringing new supervisors up to speed drives a wedge between the firms objective, which is to maximize line productivity (net of turnover), and that of the middle manager. The larger this effort cost, the farther the middle manager's optimal allocation will be from the firms. We show that when these costs are high, and when the effect of training on productivity is negatively correlated with the effect of training on retention (e.g., when supervisors who will gain the most in productivity from training will be most likely to quit afterward, as would be predicted by canonical labor market models of training (Acemoglu, 1997; Acemoglu and Pischke, 1998, 1999; Becker, 1964)), middle managers might indeed recommend supervisors who they know would gain very little in terms of productivity from training, but for whom training would greatly increase the probability of retention.

Analysis of impacts on supervisor turnover confirms these predictions. Training had a significant negative average treatment effect on turnover, with trained supervisors more than 15% less likely to quit than control supervisors. These impacts are driven entirely by highly recommended supervisors, who exhibited a 23% reduction in quitting, as compared to a 3% reduction for low-recommendation counterparts. Moreover, highly recommended supervisors in the control group are more likely to quit in the absence of training than are low-recommendation controls, indicating that middle managers indeed chose their recommendations in accordance with some private knowledge of which of the supervisors in their charge were potential "flight risks" and which supervisors might choose to stay as a result of receiving the training. Taken together, the results suggest that middle manager decision-making regarding the allocation of training was driven by turnover concerns. We confirm in calculations of returns to the firm that the productivity gains we estimate (which are intent to treat estimates already net of any impacts mediated by retention) are many times more valuable to the firm than retention gains, such that middle managers' retention priorities are indeed at odds with those of the firm as conveyed in discussions with top executives.

Leveraging a recent econometric approach by Dal Bó et al. (2021), we also show that middle manager recommendations leveraged private information, and that these unobserved factors negatively predict productivity gains while positively predicting retention effects. We use the structure inherent in this approach to estimate counterfactual allocation rules, finding that middle manager ratings of line supervisors' management and industrial engineering skills (reflecting the communication and production planning skills most centrally addressed in the training) contain informative assessments of deficiencies which could have been used to allocate the training effectively. This pattern of course contrasts sharply with the null returns achieved if allocation had followed middle managers' actual recommendations instead. That is, the middle managers indeed possess private information on critical determinants of the heterogeneous returns to training and, accordingly, could have allocated the training to achieve returns well above those from random assignment. However, the endogenous "misallocation" that arises from decentralization of the allocation decision to these middle managers who prioritize retention over productivity generates negligible returns and, as a result, strong evidence against investing in scale up of the program. Taken together, these results could help to explain persistent within-firm gaps in managerial quality.

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., 2022a,c; 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 middle managers) 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 and varied in many firms despite strong evidence of potential gains from investments in management such as the training program we evaluate. Beyond productivity, our average retention results are also complementary to the recent study by Alan et al. (2022), who find that a training program focused on the relational atmosphere within the firm, including the use of professional language in leader-subordinate interactions, has positive impacts across a range of interpersonal outcomes and, most relatedly, retention among leadership.

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 middle managers appear 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 garments 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]

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 and/or worker-task-specific productivity shocks), 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., 2022a). This recalibration of the worker-machine match (known as "line balancing"), which depends

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

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., 2022c). 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 production lines or 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 middle managers from whom we solicited training allocation recommendations. The factory level production manager works 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 production floor managers or factory level production managers whom we elicit recommendations from as the "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.

2.1.1 Middle Managerss and Supervisor Turnover

Given the central role supervisor retention plays in our upcoming analysis, we administered a short survey in September 2021 to 50 middle managers and upper managers in 5 factories in order to better understand the roles of the middle managers in relation to supervisor turnover.[2] While the sample is not representative of our study population, it provides suggestive evidence that middle managers bear personal costs when line supervisors leave. Specifically, middle managers are personally involved in many facets of replacing and onboarding line supervisors. 70% of respondents indicate that middle managers fill in for a departed supervisor before a new supervisor is assigned. 88% of respondents indicate middle managers 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 middle managers are involved in the training of new supervisors. New supervisors do not learn the necessary skills immediately. 72% of respondents indicate this process takes one week,

[2]Of the 50 respondents, 34 are designated as floor managers. The remaining 16 are Assistant Production Managers or Production Managers, who would be above floor managers in the organizational hierarchy of a large factory but would serve as floor managers themselves in smaller factories.

while the remaining responded “two or three weeks” (24%) or “a month or more” (4%). Finally, all 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 middle managers are the only employees involved in these processes. The survey results suggest, for example, that HR and upper management are also involved in replacing and training supervisors.[3] Similarly, 34% indicates floaters or assistant supervisors can be involved in filling in for a supervisor. However, collectively, the above patterns indicate that middle managers are generally involved in each step of filling in for, replacing, and training a supervisor, indicating that there is a personal cost to middle managers when line supervisors quit.

2.2 Program Details and Content

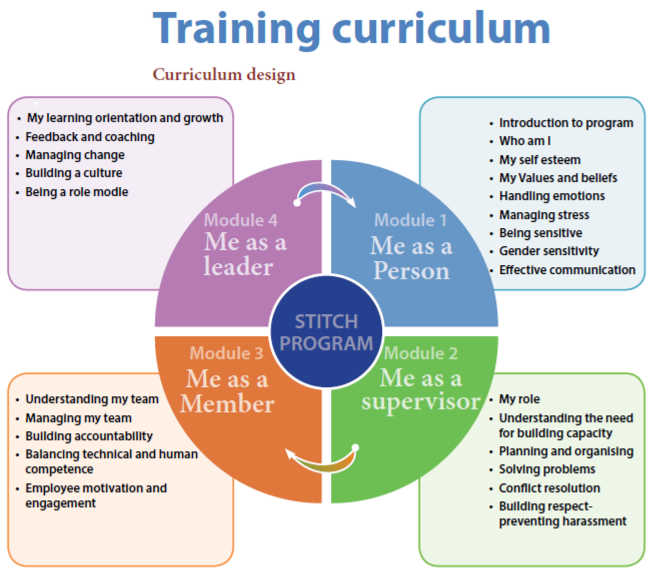
Drawing from our prior work in this specific context showing the productive value of both soft skills such as communication and specific managerial skills and practices such as control, autonomy, and attention (Adhvaryu et al., 2022b,c), the STITCH program was designed to train line supervisors in the skills and practices most likely to improve productivity, as line productivity has been identified by the firm as the main difference maker in profits.[4] The program consisted of 25 weekly hour-long sessions divided to 4 main 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 of the 25 sessions they cover. Below we give examples of how the STITCH training relates to the skills previously identified as productivity enhancing in this specific context in two studies, leaving a full discussion of the contents of each session to Appendix Section A.2.

First, Adhvaryu et al. (2022b) finds evidence that soft-skills training primarily focusing on effective communication, time/stress management, and problem solving makes garment sewing workers more productive, primarily through improved teamwork and collaboration. These exact skill sets are emphasized throughout STITCH training as they have the potential to be productivity enhancing with supervisors having a large role in enabling collaboration in our context. For example, in two sessions that directly emphasize communication skills, trainees participate in role-playing activities to learn about different communication styles, importance of communicating assertively and responsibly, and practice skills of giving and eliciting constructive feedback. For stress management, two sessions have explicitly focused on activities to understand emotional responses to situations, what positive actions can help manage emotions, identifying causes and effects of stress, and tips for effective stress management. For problem solving, there is a session where participants are trained on problem solving skills such as problem identification, analyzing the root cause, and making decisions

[3]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 middle managers with regards to replacement and training, while upper management tends to be ranked higher.

[4]In discussions with senior executives of the firm, initiatives to reduce energy consumption and worker turnover are also mentioned but are considered at least an order of magnitude less impactful for profits than labor productivity.

Figure 1: STITCH Modules and Sessions



from available options using case studies and role-playing exercises.

Second, Adhvaryu et al. (2022c) show supervisors with higher managerial control, attention, and autonomy enable higher team productivity. *Control* refers to the belief in the capacity to influence and control events and outcomes, which has been implicitly and explicitly underlined throughout the STITCH training. Broadly, the program focuses on supervisory behaviors, skills, and attitudes that enables effective team performance and instills to the trainees that their actions strongly influence workplace outcomes and productivity.

Attention broadly refers to undertaking practices that demonstrate effort and attention to accomplishing managerial tasks. One aspect of attention is active personnel management, which is related to a session focused on effective methods of employee motivation through, for example, showing appreciation and helping employees realize their value. Frequency of monitoring work is also a key component of managerial attention, and the importance of monitoring is underlined in multiple STITCH sessions on the role of the supervisor, effective planning/organizing, and building a good work culture.

Autonomy encompasses behaviors and practices that capture the degree to which the supervisor directs the team's activities proactively and without relying on input from superiors. The ability to do so also relies heavily on the nature of the rapport established with workers.[5] While communication related sessions discussed above have direct bearing on having a good rapport with subordinates (and to directing the team's activities), STITCH training had additional sessions focusing on the

[5]These two leadership styles map to the two types of behaviors identified in the leadership literature as "initiating structure" and "consideration" (Stogdill and Coons, 1957).

importance of sensitivity in interpersonal relationships and prevention of harassment by supervisors. With regards to directing team activities, a planning and organizing session focused on the importance of planning for effective team work and asks groups of participants to come up with a plan for their team to fulfill hypothetical orders.

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 middle managers, again as indicated by management, were also administered a baseline. In the middle manager baseline, among other items, middle managers were asked to rank the supervisors they managed (from 1 to 5) according to how much they believed the supervisor would gain from training. The wording of the question (presented in the next section) made clear that this ranking would indeed affect the probability that the supervisors were included in the first batch of training. We refer to this variable as the middle manager recommendation for the rest of the text. We collected this middle manager recommendations for 1175 supervisors included in the analysis.

Randomization was stratified in multiple dimensions. For supervisors with middle manager recommendations, whether middle manager recommendation was high, moderate, or low were used as strata. Second, for supervisors that were mapped to specific production floors, the production floor was used as a strata. Finally, supervisors were grouped into similarity clusters based on personal and line characteristics for randomization.

While randomization was at the supervisor level, our key outcome of productivity is at the line level. Of the 1849 supervisors administered a baseline, a subset of 954 supervisors who (1) undertook duties directly related to production[6] 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 as shown in Appendix Figure A.2.[7]

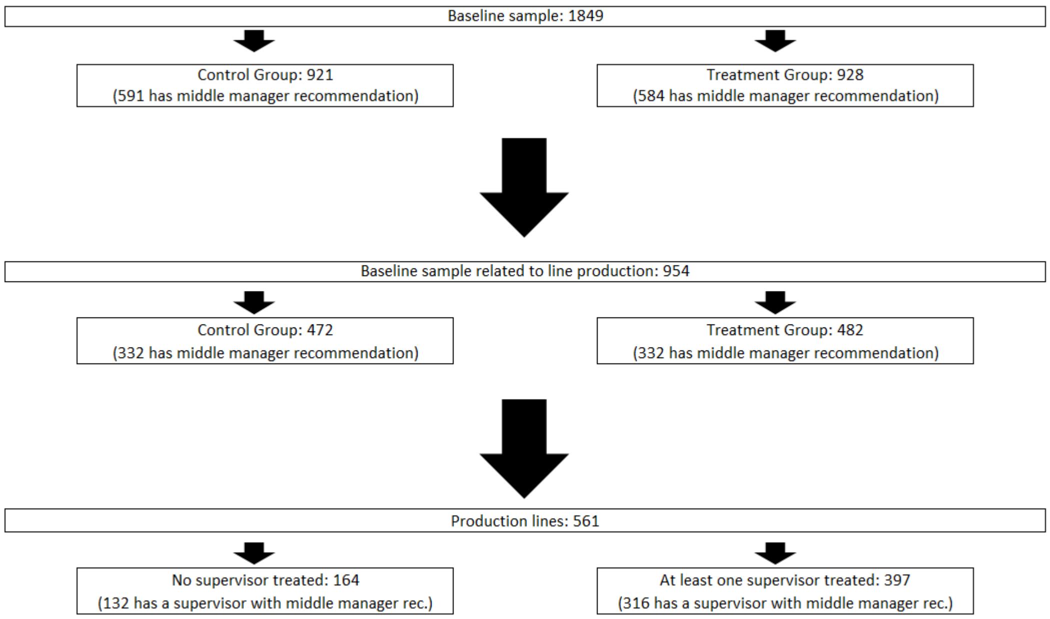
Figure 2 presents a schematic diagram of our experimental design. Of the baseline sample 1849 supervisors, 921 were randomized into control and 928 were in the treatment group. Of the

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

[7]We check robusntess of all the main results in the paper to use of binary transformations of this continuous treatment measure, including a dummy for all supervisors on a line treated and a dummy for any supervisor on a line treated.

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.

Figure 2: STITCH Study Experimental Design



Appendix Figure A.1 presents a timeline of our intervention. Middle managers and supervisors were administered the baseline surveys from December 2016 to March 2017. Training start and end dates were different across different factories.[8] 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 3.

3 Data

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

3.1 Production Data

Each production line on the sewing floors records hourly productivity data. We aggregate the hourly data to the day 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

[8]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 such that we can analyze differences between treatment and control at the same time relative to treatment within each factory.

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 data before, during, and after the training. For productivity, our analysis period spans the 6 months before training start to 6 months after training end for each production line.

3.2 Human Resources Attendance, Pay, 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.[9] We also have have access to monthly salary data which we use to see whether trained supervisors experience differential gross salary growth. The firm also has an incentive scheme where bonus payments are made to employees based on performance. Daily data on incentive payments are collected by the firm with the production line and designation of the individual who received the payment noted in the data. We use this daily data to assess whether workers in lines with treated supervisors receive higher incentive payments. Finally, using human resources personnel rosters, we further match approximately 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.

3.3 Survey Data

We complement the administrative data with three surveys. First, we administered a baseline survey of supervisors eligible for training from December 2016 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 the determinants of middle manager recommendation (discussed in the next paragraph) and the heterogeneity of treatment gains. B.4 provides a list of survey indices we use.

We also conducted a survey of middle managers, who are above the line supervisors in the firm hierarchy. We primarily use this survey for two reasons. First, and most importantly, we elicit information from the middle managers about which supervisors under them should be prioritized for training based on who they think would gain the most. We refer to this as middle manager recommendation. Second, we also elicit information from the middle managers about the managerial skills, technical skills, industrial engineering skills, and the motivation to improve of the supervisors

[9]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.

they manage. We use these skill and motivation scores in exploring the determinants of middle manager recommendations and in exploring whether the middle managers have useful knowledge about the line supervisors they manage. Specifically, we ask the following questions to measure these features, in the following order:

1. **Skill Scores**: "Imagine a ladder with 5 steps. At the lowest step is a supervisor/floater you know, who has the lowest level of [*skill of interest*]. At the highest step is a supervisor/floater you know, who has the highest level of [*skill of interest*]. On which step would you place each of the supervisors/floaters? "

   For the following [*skills of interest*]

   (a) "technical tailoring skills"

   (b) "industrial engineering (IE) skill (e.g., assigning workers to operations, meeting targets, relieving bottlenecks, line balancing)"

   (c) "non-technical management skill (e.g., communication, leadership, ability to motivate line, sense of responsibility)"

2. **Motivation to Improve**: "Imagine a ladder with 5 steps. At the lowest step is a supervisor/floater you know, who has the lowest level of motivation to improve his/her skill. At the highest step is a supervisor/floater you know, who has the highest level of motivation to improve his/her skill. On which step would you place each of the supervisors/floaters? "

3. **Middle Manager Recommendation**: "HR is planning to train supervisors and floaters in soft skills (e.g., communication, leadership, time management, problem-solving and decision-making). Shahi Management feels it cannot train all supervisors at once and would like to focus first on those who will benefit the most. We would like to know who you think will gain the most from this training. Taking into account current skill levels and ability for improvement please rank all your supervisors in order of who you think will benefit the most (Rank 1 is most benefitted). Those who you say will have the highest expected gain will have a higher chance of getting this training."

   In the rest of the paper, for consistency of exposition with other skill scores, we flip the middle manager recommendation rankings in order for higher values to mean higher recommendation.

Finally, we administered a follow-up survey to a group of 50 middle managers and upper managers in 5 factories in September 2021. This short survey focused on the role and responsibilities of middle managers in relation to supervisor turnover. We discuss the results of this survey above in section 2.1.1.

3.4 Pre- and Post-Module Test Scores

Before and after each training module, all treated supervisors and a randomly selected group of control supervisors were 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 the content covered in the training.

3.5 Summary Statistics and Baseline Balance

Table 3.1 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 3.2. 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 in the data 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 middle manager recommendation, we limit our sample to lines for which we have middle manager recommendations (i.e., lines who are mapped to supervisors with middle manager recommendations). Some incidental imbalance is introduced for the subsets.[10] 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.

4 Treatment Effects

We start our empirical investigation by showing that the treated supervisors perform better on tests administered after each training module. Interpreting this as evidence of a first-stage effect, we next analyze the effects of the STITCH training on two key outcomes: productivity and retention. For each key outcome, we first test for average training impact. Then, we check whether supervisors recommended highly for the training by their middle manager gain more in terms of productivity/retention. We leave additional treatment effects on important outcomes that our data allow us to investigate, yet are not central to our research design, to section 6.2. These include salary growth, incentive bonus payments, supervisor attendance, and worker attendance/retention.

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

Table 3.1: 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)
Middle Manager Recommendation	1175	3.05	1.61	-0.066
				(0.094)
Technical Skill (scored by Middle Manager)	1175	3.95	0.94	-0.023
				(0.055)
Industrial Engineering Skill (scored by Middle Manager)	1175	3.96	0.94	-0.004
				(0.055)
Management Skills (scored by Middle Manager)	1171	4.01	0.93	-0.010
				(0.054)
Supervisor Motivation (scored by Middle Manager)	1171	4.17	0.86	0.011
				(0.050)

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

Table 3.2: Line Level Descriptive Statistics and Balance

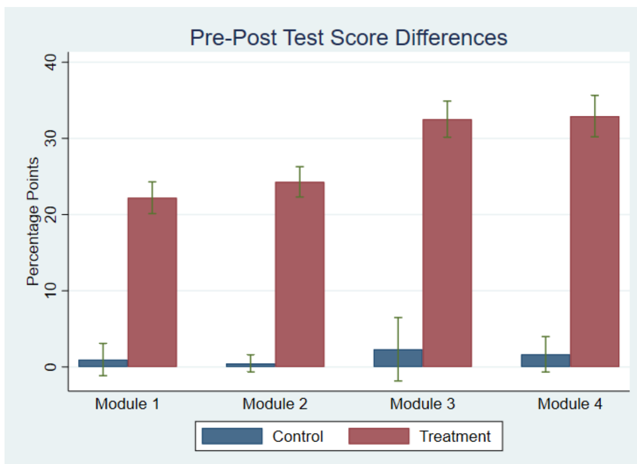
	Num Lines	Mean	SD	Coefficient/SE
Line Efficiency (Baseline)	540	54.96	9.31	0.940
				(0.978)
Line Attendance (Baseline)	541	0.89	0.05	-0.007
				(0.006)
Line Retention (Baseline)	528	0.83	0.14	0.004
				(0.016)
Line Budgeted Efficiency (Baseline)	541	60.60	7.58	-0.217
				(0.836)

Note: 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.

4.1 First Stage: Pre and Post Module Assessment

We first investigate whether the treated supervisors outperform control supervisors on tests covering the four training modules. Specifically, we compare the differences between pre- and post- module test scores for the two groups. Figure 3 presents the average difference 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 C.2 presents estimates of the treatment effect for each module using an ANCOVA specification. Consistent with the raw differences presented in the figure, treated supervisors outperform the control supervisors significantly for each module. At the low-end, treatment increases the performance of supervisors on tests of module 2 content by around 22 percentage points, which corresponds to a 40% increase from the baseline mean. The treatment more than doubles the post-training test scores for both modules 3 and 4 by inducing an increase of 32 and 39 percentage points, respectively. We interpret this as evidence that training induced substantial learning of the material covered by the modules.

Figure 3: Pre-Post Test Score Differences



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

It is also worth discussing specifically how acquisition of the content of the training might have translated into the productivity gains we document below. As discussed in Section 2.2, the training curriculum was informed by the results of a previous study in this context (Adhvaryu et al., 2022c) which identified the skills and practices of supervisors that contributed most to line productivity. In line with these previous findings, the curriculum mainly focused on managerial *control* (belief in capacity to control outcomes and effect change), *attention* (effort and attention towards accomplishing managerial tasks, particularly personnel management) and *autonomy* (planning and organizing production proactively and without relying on inputs from superiors).

Many questions in the pre- and post-module tests gauged the internalization these particular skills and practices. For example, the test pertaining to Module 1 content asked supervisors about the value of respectfully listening to workers' perspectives, managing emotions when communicating, and effective approaches to communicating with workers including non-verbal communication. These communication styles and skills are explicit inputs into both the managerial attention and autonomy factors identified in the prior study which informed curriculum development. That is, both effective monitoring and problem identification and solving depend crucially on the quality and frequency of communication between supervisors and workers. The module 2 test measured the degree to which supervisors internalize the importance of personnel management and production planning, key components of Attention and Autonomy, respectively.

Treatment supervisors exhibited the largest gains on the tests pertaining to Modules 3 and 4. The Module 3 test focused mainly on the importance of accountability and monitoring, primary components of the managerial attention factor. While true/false questions like "changes will keep happening at work in a garment factory and managing change is the responsibility of the managers" in module 4 captured the belief in the ability to control outcomes (i.e., managerial control). In our specific context, belief in the ability to control outcomes, directing effort and attention towards personnel management, and preemptive planning and proactive adjustment to plans combine to maximize productivity. That is, Adhvaryu et al. (2022c) find that the most productive supervisors engage in close and frequent monitoring of the line, open communication with workers to understand issues and bottlenecks, balancing of lines with new workers or the reallocation of workers across operations if needed, and proactively solving issues raised by workers such as poorly calibrated machines.

4.2 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 intent-to-treat (ITT) difference-in-differences (DD) specification to assess the productivity effects on a line-day level:

$$y_{ltr} = \alpha + \beta_1 T_l \times 1[During]_{lt} + \beta_2 T_l \times 1[Post]_{lt} + \delta_l + \mu_t + \gamma_r + \epsilon_{ltr} \tag{1}$$

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 $1[During]_{lt}$ and $1[Post]_{lt}$ 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 in each factory and it captures how many days have elapsed since the beginning of training in the factory of the line (with negative dates for before training). 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.[11] Our coefficients of

[11] We have experimented with using other cutoffs and results are stable under reasonable cutoff values. We further show

interest are β_1 and β_2 which estimate the causal effect of fraction of supervisors treated on line level productive efficiency.

An important feature of our design is the inclusion of training-relative date fixed effects γ_r in our preferred specification. Their inclusion underlines an important distinction between our setting and the standard staggered adoption difference-in-differences or event study designs about which there has been an active recent literature (Bilinski et al., 2022). Namely, we have within cohort randomization. While different factories start treatment on different dates, within each factory lines are randomized into treatment and control. Therefore, even for $r > 0$ we have both treated and untreated production lines. Inclusion of γ_r allows us to recover treatment effects by comparing treatment and control lines within cohort and training-relative date. The recent literature has raised concerns about the interpretation of two-way fixed effects estimators in the traditional staggered adoption setup (Borusyak et al., 2022; de Chaisemartin and D'Haultfuille, 2020; Goodman-Bacon, 2021; Sun and Abraham, 2021) in the presence of treatment effect heterogeneity across either time or treatment cohorts or both. Within factory randomization of production lines into treatment and the ability to recover treatment effects within cohort-time alleviates these concerns in our setup. We further show the robustness of our DD productivity results to running the analysis on a balanced panel in relative time (6 months before to 20 months after training start for each line) in our robustness discussion below.

As shown in Table 4.1, 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 supervisors treated are, on average, are 4.1 percentage points (7.3% of control mean) more efficient during training. Given the training was administered over a considerable duration of an average of 9 months, 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 possibly get reassigned to different lines (or successful practices can spillover to control lines) over time, we might expect the line-level treatment effect from treating supervisors dampen over time. However, the results show that the sum of these possibilities is not enough to erode the treatment effect substantially for at least 6 months after training completion. The average productivity results strongly indicate that the training increased profitability of the firm, as the firm focuses on labor productivity as the main lever for affecting profits. That is, the other major sources of costs are raw inputs such as cloth, yarn, and energy over which the firm feels it has little

robustness of our estimates to not dropping these lines.

control. We confirm this intuition in section 6.3 by undertaking an ROI analysis using information from the firm's accounting department.

Table 4.1: Effects of Training on Line Productivity

	Efficiency (Produced/Target)				
	Analysis Lines			Lines w/ Middle Manager 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

Note: 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.

To assess the dynamics of the productivity results, the Appendix Figure C.2 presents the monthly event study results, starting from 6 months before training start to 20 months after training start for each line. 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. This suggests that, even after close to a year after training end for many lines, the treatment effects do not go to zero.

Robustness Heterogeneity of productivity gains with regards to middle manager recommendation is a key focus of the study. Given this analysis can only be done on the subset of lines with middle manager information, column 4 of Table 4.1 shows that our results also hold in this subset. Column 5 shows robustness to including both the 0 productivity line-day observations and the lines with many 0 productivity days in our analysis. While results in Table 4.1 are from an unbalanced panel in relative time due to differing treatment length across factories, Appendix Table C.3 shows that the results are robust to running the analysis on a balanced panel in relative time (6 months before to 20 months after training start for each line). While there is an approximately 10% decrease in the after-training effect size, this is to be expected as the analysis period includes more months farther away from training for every line. Finally, Appendix Section C.3.4 shows robustness of our productivity results to defining line-level treatment as treating any supervisor on a line.

4.3 Productivity Effect Heterogeneity by Middle Manager Recommendation

Central to our research question is whether the middle managers would allocate the training to supervisors who would gain the most. If middle managers possess private information about who would gain the most from training, they could allocate training to maximize gains. However, middle managers can also have different objectives than the firm, which could conceivably lead to allocation rules that do not maximize productivity gains from training. To start exploring this question, we test if the productivity gains are higher for lines with highly recommended supervisors by their middle managers. To do so, we modify our difference-in-differences specification in equation 1 to include three way interactions between treatment, the treatment periods, and high middle manager recommendation indicator ($1[Rec_l]$). In order to go from supervisor level middle manager recommendation to the line level, we average the recommendation of every supervisor on the line. We set $1[Rec_l] = 1$ if the line level average is above the median recommendation rank of 3.[12]

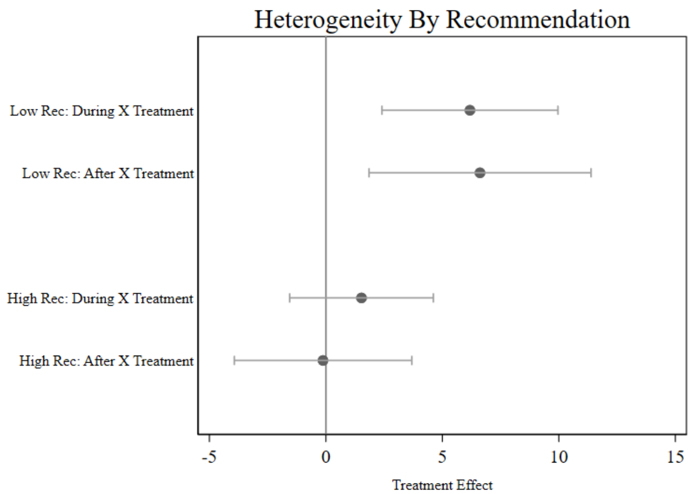
Figure 4 charts the treatment effect estimates for the high recommendation and the low recommendation lines. It is clear that there is significant heterogeneity in how much different supervisors gain from treatment. Strikingly, the productivity gains are highly concentrated among lines with low recommendation supervisors. During training, treated lines with low recommendation supervisors experience a 6.2 percentage points (11% relative to baseline) increase in productivity. For lines with recommended supervisors, the corresponding treatment effect is only 1.5 percentage points (3% relative to baseline) and statistically insignificant. 6 months following treatment, lines with highly recommended supervisors effectively do not gain from training while low recommendation lines have a treatment effect of 6.6 percentage points. Appendix Table C.4 reports the underlying regression results (column 3) along with robustness to less stringent FEs and using all production lines in our analysis.

4.4 Supervisor Retention

Next, we focus on the impacts of training on supervisor retention. To assess the retention effects of the treatment on supervisors, we estimate a Cox proportional hazard model, taking the randomization strata into account:

[12] As mentioned in Section 3.3, we flip the middle manager recommendation rankings in order for higher values to mean higher recommendation.

Figure 4: Line Productivity Heterogeneity by Middle Manager Recommendation



Note: Treatment effects on line productivity for high and low middle manager recommendation lines. A line is defined as high recommendation if the average recommendation of the supervisors on the line is above the median. 95 % CI shown.

$$q_{ist} = h_{0t} \times exp(\mu_s + \beta T_i) \tag{2}$$

where q_{ist} is the hazard function for quitting, μ_s is the strata fixed effects and T_i is the treatment indicator for supervisor i in strata s at time t. We present results for both the full set of supervisors in our study and for the subset of supervisors for whom we have middle manager recommendations, as this is the subset we use to assess heterogeneity below. We limit our sample to supervisors who are with the firm when the training starts in their factories. For each supervisor, the analysis spans from the first day of training to the end of May 2018.

Retention results are presented in Table 4.2. Both in the production sample and the full sample, treatment leads to a decrease in the hazard ratio for quitting. In the full sample, treated individuals are 15% less likely to quit. For the smaller sample with middle manager recommendations, the treated supervisors are 10% less likely to quit between training start and May 2018. While this is still an economically meaningful effect, it is imprecisely estimated and not statistically significant.

4.5 Retention Effect Heterogeneity by Middle Manager Recommendation

To explore whether the retention effects are heterogenous with regards to middle manager recommendation, we plot survival curves (with retention as the outcome) for treated and control supervisors, separately for high- and low- recommendation supervisors. Figure 5 presents the results, with figure 5a showing the curves for high-recommendation supervisors and figure 5b include the low-

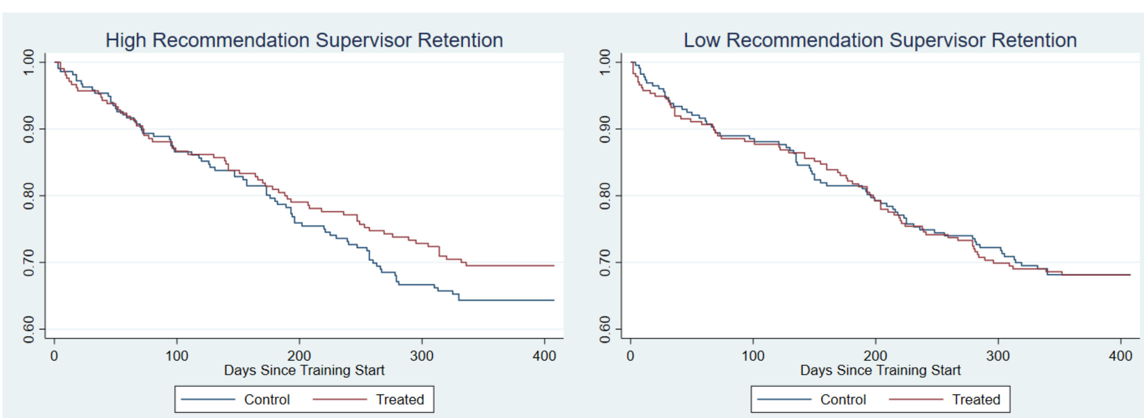
Table 4.2: Supervisor Retention

	Supervisor Quit	
	All Supervisors	Supervisors w/ Middle Manager 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

Note: *** p<0.01, ** p<0.05, * p<0.10. The sample is restricted 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 middle manager recommendations. Relative Hazard is calculated as the exponent of the coefficient on treatment.

recommendation supervisors. The results suggest that the retention effect of training is driven entirely by highly recommended supervisors. Appendix Table C.7 show results from the associated proportional hazard models, showing treated supervisors are 22% less likely to quit among high-recommendation supervisors, while only 3% less likely among low-recommendation supervisors. We are underpowered to detect statistically significant changes across the two groups, but the results are suggestive that middle managers target retention effects when allocating the training. Note that this pattern suggests that retention effects are not a major source of productivity gains from the training as the retention gains are concentrated among supervisors with high middle manager recommendations, precisely the group that exhibits negligible productivity gains.

Figure 5: Retention Results for High and Low Middle Manager Recommendation Supervisors



(a) High Middle Manager Recommendation

(b) Low Middle Manager Recommendation

Taken together, we have three key takeaways with regards to productivity and retention. First, there are large average impacts from STITCH training, especially for line-level productivity. Second, there is heterogeneity in who gains the most from training, across both productivity and retention. Third, middle managers target supervisors who gain the most in terms of retention while gaining very little in terms of productivity. Next, we turn to interpretation of these patterns.

5 Interpreting the Middle Manager Recommendations

The previous section highlights two facts about the supervisors with high middle manager recommendation: (1) they gain significantly less in terms of productivity; and (2) they gain more in terms of retention from the STITCH training. Keeping in mind that our productivity results are ITT at the line level (such that line level productivity is observed and analyzed even if a study supervisor leaves the firm), and hence provide us with treatment effects on productivity that incorporate any effects on productivity mediated through effects on retention, we interpret these results as evidence that the middle managers were targeting retention gains above and beyond productivity gains when allocating training. This is consistent with the idea that the middle managers have private incentives to target retention beyond its implications for productivity, even if the firm would rather allocate training to increase productivity itself. In our context, there are potentially sizable personal costs to middle managers in the event of supervisor turnover. First, as survey evidence we discussed in section 2.1.1 indicates, there are costs to middle managers in terms of time and effort to replace and onboard new supervisors. Further, middle managers can face professional costs if the line supervisors they manage quit frequently. With personal costs to retention and a trade-off between productivity and retention gains, we can rationalize the allocation decisions of the middle manager.

In this section, we first provide a stylized model that matches the key elements of our context to formalize how, in the presence of private costs to supervisor turnover, middle managers would misallocate the training from the perspective of a firm that primarily targets productivity. We then analyze to whom the middle managers would actually allocate the training. We start by showing that middle managers possess valuable information about line supervisors they manage, implying they have private information to target the training. We then use our rich baseline characteristics data to show the observable characteristics of highly recommended line supervisors. However, much of the variation is not explained by observables, which suggests unobservable private information drives much of the recommendation decisions. Finally, we adapt the framework proposed by Dal Bó et al. (2021) to decompose the middle manager selection to observable and unobservable components and show that the unobservable component of selection primarily predicts lower productivity gains and higher retention gains for recommended supervisors.

5.1 Simple Model of Training Allocation

Our empirical results suggest that middle managers target training to supervisors who gain below the average in terms of productivity but above the average in terms of retention. In this section, we posit a simple model of supervisor training allocation to rationalize and interpret these patterns in the context of a principle agent problem. The model consists of a firm, a middle manager, and the supervisors to whom training will be allocated. Training heterogeneously affects both the individual productivity and the retention probabilities of supervisors. We study how the firm (the principle) and the middle manager (the agent) would choose to allocate the training to supervisors with heterogeneous gains.

5.1.1 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 response τ_p^i and retention response τ_δ^i. Supervisor training takes place in the first period and its productivity effects are immediately 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 reflects that replacement supervisors can initially be less productive or that during the process of replacement the line may be less productive for a period. The replacement supervisors are not trained and their productivity are not shifted by τ_p.

Payoff for the Firm and the Middle Manager. Both the firm and the middle manager are risk neutral. The firm's objective function is to maximize line productivity.[13] The middle manager also aim to maximize line productivity, but they also incur an additional personal cost c in period 2 if the line supervisor quits after period 1. This term captures both the personal replacement/training costs the middle manager incurs, and also the fact that, in our context, part of the middle manager's job is to ensure retention of supervisors. Therefore, there can be a professional cost to high turnover of supervisors. The economic consequence is that the personal replacement cost c misaligns the principal's (the firm) and the agent's (the middle manager) objectives.

We write the middle manager's valuation of a trained and untrained supervisor i as:

[13] Conversations with the firm confirmed that, in our context, this assumption is realistic as productivity is the largest determinant of profitability which the firm feels it can influence.

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

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

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

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

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 middle managers 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).

Middle Manager Information and the Distribution of Types. Supervisor types are indexed by their productivity and retention gains. The marginal distribution of the productivity gains is $\tau_p^i \sim G(.)$. We assume the middle manager has 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 ($f'(\tau_p^i) < 0$). This negative relationship induces a trade off between gains in terms of retention and productivity, and it is consistent with what we observe in the empirical analysis. Finally, we further impose $f''(\tau_p^i) < 0$ and that the ultimate retention probability $\delta + \tau_\delta$ must lie in the unit interval.

To explore why a negative relationship may exist between retention and productivity gains, suppose that the supervisor's retention response to training has two components: a component that responds to the supervisor's productivity gain from training and an idiosyncratic component. The former component can be negatively related to productivity gains, as increasing the supervisor's productivity would make them more valuable in the labor market and (without a proportionate increase in wages, which we do not observe) could increase the likelihood they leave for another job. The idiosyncratic component could be arbitrarily related to the productivity gains. In section 5.2.1 we show suggestive evidence that control supervisors with high recommendation had a lower retention rate compared to their low recommendation counterparts. Supposing lower baseline retention rate implies a higher retention response to training, this suggests even the component of the retention gains that is not a direct response to the changing labor market outcomes (as the control supervisors were not trained) could still be negatively correlated with productivity gains.

5.1.2 Ideal Supervisor Type to Train

We consider the question of what supervisor type has the highest training payoffs from the perspective of the middle manager, which we call the *ideal type* from the perspective of the middle manager.[14] The middle manager aims to maximize the expected payoff from training a type with τ_p^i:

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

which leads to the ideal type:

$$\tau_p^* = \frac{1+\delta+f(\tau_p^*)}{-f'(\tau_p^*)} - p(1-z) - c. \tag{3}$$

and the expected productivity gain for the ideal type:

$$\mathbb{E}[\Delta_p^*] = \tau_p^*(1+\delta) + f(\tau_p^*)((1-z)p) + f(\tau_p^*)\tau_p^* \tag{4}$$

Remark The expected productivity gain from the ideal type decreases as the middle manager 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 middle manager 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 middle manager as c increases, due to the trade-off between τ_p^i and τ_c^i.

[14] 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 middle manager.

5.1.3 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.[15] 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 irrespective of supervisor retention, our ITT 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 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 Middle Manager 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.[16] This would guarantee a larger treatment effect than random allocation. However, if the firm relies on middle managers 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 middle manager allocation would approximately be the same as the firm allocation as the middle manager also wants to maximize line productivity. If, on the other hand, c is relatively large, middle managers could heavily target retention, generating productivity gains well below that of random assignment in the process.

5.2 Who Do the Middle Managers Recommend?

5.2.1 Middle Managers Have Useful Information About Line Supervisors

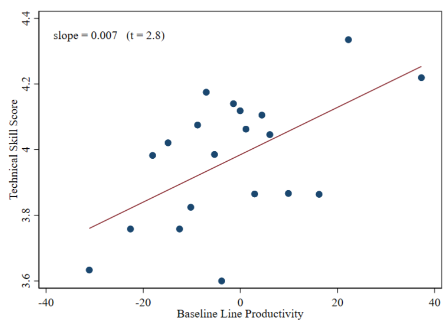
Before focusing on the determinants of middle manager recommendations, we first explore whether middle managers seem to have useful information on the supervisors which would allow them to allocate the training effectively. If middle managers lack information entirely, we might expect that the recommendations are effectively random. However, this interpretation is hard to square with the strong relationship between who the middle managers recommend and gains across productivity and retention. Nevertheless, we check whether the skill scores we elicited from middle managers about the supervisors support the notion that middle managers have useful information about their supervisors.

[15] In the stylized model, there are no dynamics to the treatment effect and the training is instantaneous. In our empirical work, we differentiate the effects of the training while the training is ongoing and the 6-month period following the training.

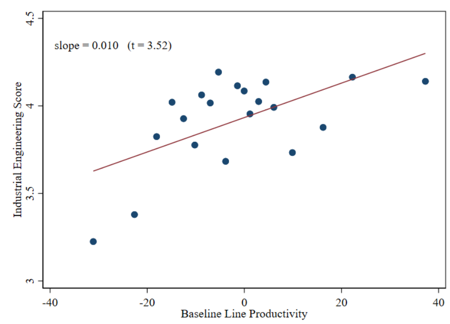
[16] This decision can include not to train supervisors who would not gain from training (or would gain less than the cost of the training if there is a cost to training an individual)

In the middle manager baseline survey, we asked the middle managers to score (from 1 to 5) all the supervisors they list as reporting to them in three dimensions: management skills, industrial engineering skills, and technical skills. In Figure 6, we show the correlation between all three of these scores and the baseline productivity of the line(s) they supervise.[17] The results indicate that the skill scores are positively correlated with baseline productivity, with the relationship more pronounced for industrial engineering and technical skill scores. Appendix Table C.12 further shows that the association is positive for all skills and statistically significant for industrial engineering and technical skills. This is suggestive that the middle managers have useful information about the skill sets of their supervisors.

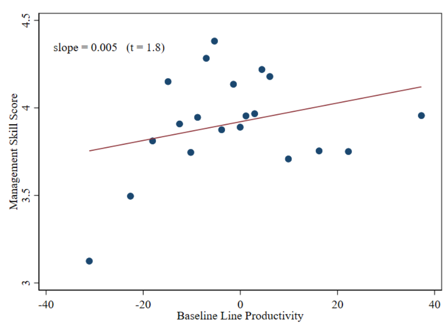
Figure 6: Middle Manager Assessment of Skills and Line Productivity



(a) Technical Skills



(b) Industrial Engineering Skills



(c) Management Skills

Note: Binned scatter plots between the middle manager 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.

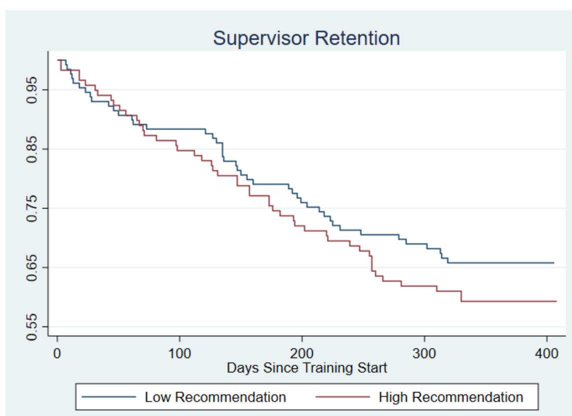
We further investigate the information content of the middle manager skill assessments by focusing on the heterogeneous productivity gains from training with regards to the three skill types. We augment our difference-in-differences specification in 1 to include the three way interactions between treatment, treatment periods, and the average skill scores of the line supervisors. If the skill scores capture meaningful information about the skill sets of the supervisors which are then

[17] The productivity of the line is calculated following an AKM-style two-way fixed effect model, following Adhvaryu et al. (2022c).We implement the model on productivity data from the three months preceding survey (January-March 2017). Further description of the procedure can be found in Appendix Section B.3.

augmented by training, we would see differential effects of training by baseline stock of skills as indicated by middle managers. Further, given the focus of the training is managerial skills, we would expect that baseline level of managerial skills and industrial engineering skills (which most closely map to the content of the training) would be be particularly related to treatment gains.[18] Table 5.1 presents our results. As expected, baseline industrial engineering and management skills have a large and significant effect on productivity gains. Specifically, supervisors with lower level of baseline skills gain more from the training, indicating that training is a substitute for baseline skills in these dimensions as assessed by the middle managers.[19] Technical skills are also negatively related to treatment gains, but the effect is smaller and statistically insignificant (especially during training). These results suggest that middle managers have nuanced information about the baseline skill sets of the supervisors they manage and this information can be leveraged to allocate training to maximize productivity gains relative to full randomization.

Finally, we check if the middle manager recommendation is related to quitting rates in the control group. While statistically insignificant, Figure 7 shows that highly recommended supervisors in the control group are more likely to quit than their low recommendation counterparts (recommended supervisors are 15% [p = 0.29] more likely to quit in a hazard regression). That the middle manager recommendations are predictive of future quitting rates in the control group is consistent with the idea that middle managers possess private information about their supervisors. It is further consistent with the idea that middle managers could be using the training to target retention gains, as the recommended supervisors are more likely to quit the firm absent the treatment.

Figure 7: Control Group Retention by Middle Manager Recommendation



Note: Within the control group, highly recommended supervisors are relatively more likely to quit the firm

[18]Industrial engineering skills we underline in our middle manager survey includes ”assigning workers to operations”, ”meeting targets” and ”line balancing” which are skills that are related to managerial and leadership capacities covered in the training.

[19]This is consistent with Adhvaryu et al. (2022b)’s conclusions about a similar soft-skills training aimed at sewing workers.

Table 5.1: Heterogeneous Productivity Effects by Supervisor Skill

	Efficiency (Produced/Target)		
	(1)	(2)	(3)
During Training X Treatment	6.917	17.576***	16.716***
	(6.430)	(6.280)	(5.851)
After Training X Treatment	12.774*	18.362***	19.349***
	(6.955)	(6.678)	(6.229)
During Training X Treatment X Technical Skill	-0.818		
	(1.525)		
After Training X Treatment X Technical Skill	-2.476		
	(1.688)		
During Training X Treatment X Ind. Eng. Skill		-3.679**	
		(1.513)	
After Training X Treatment X Ind. Eng. Skill		-4.138**	
		(1.604)	
During Training X Treatment X Management Skill			-3.605**
			(1.428)
After Training X Treatment X Management Skill			-4.564***
			(1.550)
Observations	189380	189380	189380
Number of Lines	395	395	395
Control Mean of Dependentent Variable	55.279	55.279	55.279
Line FE	X	X	X
Date FE	X	X	X
Relative Date FE	X	X	X

Note: Standard errors are clustered at floor 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. 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.

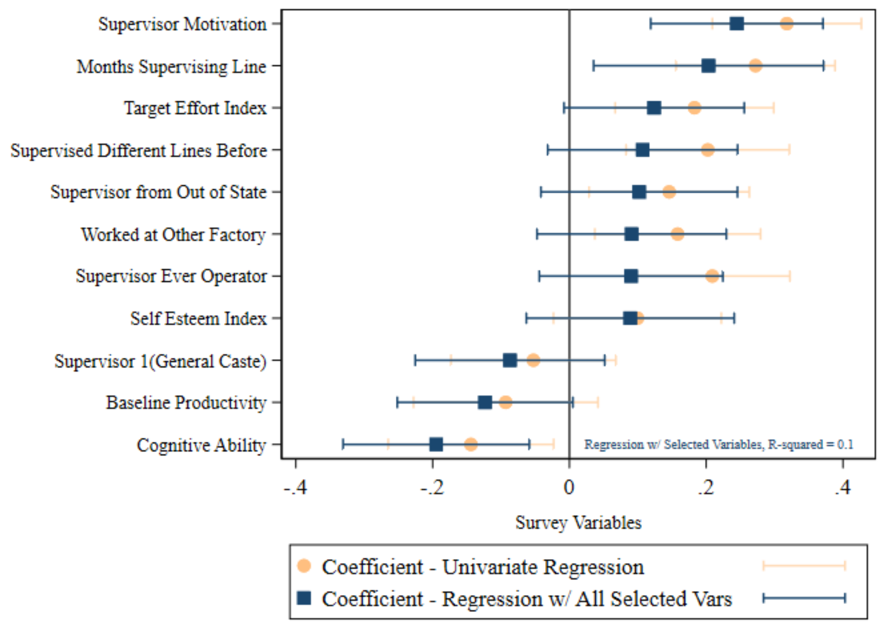
5.2.2 Observable Determinants of Middle Manager Recommendation

Having established that middle managers seem to possess information to select which supervisors would gain more from training in terms of productivity, we now turn to who the middle managers have actually recommended. Specifically, we focus on the observable determinants of middle manager recommendation. Given the rich set of baseline information we have about the supervisors and middle managers, we use a simple LASSO procedure to see which variables out of 52 supervisor characteristics and supervisor-middle manager joint characteristics (for example, whether they share the same religion or are from the same state) are associated with high recommendations.[20] Given we are primarily interested in the negative relationship between recommendations and the productivity gains from training, the analysis below focuses on the production sample (i.e. the set

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

of supervisors who undertook duties directly related to production in a specific line) and includes line-level covariates such as baseline productivity of the line. The outcome of interest is middle manager recommendation (ranging 1 to 5).

Figure 8: Lasso Selected Variables



Note: Variables selected from the lasso procedure. The light orange coefficients are from a regression of middle manager recommendation on the selected variable of interest. The dark blue coefficients are from a regression of middle manager recommendation on all selected variables. 95% CI are shown using robust SEs.

As presented in Figure 8, 11 variables are selected using this procedure. Before looking at the selected variables, we note that linearly regressing the middle manager recommendations on all the selected variables yields an R^2 of only 0.1. Further, 8 out of the 11 variables have coefficients insignificant at the 5% level. We take these as evidence that, on average, much of the variation in the middle manager recommendations are driven by unobservable factors, even with the rich set of baseline variables we observe.

While our main takeaway is the overall importance of unobservable factors, a few patterns emerge with regards to observable characteristics. First, variables indicating high tenure and variety of experience consistently predict higher middle manager 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, middle managers seem to recommend individuals who they view as motivated, as evidenced by not only the positive coefficient on supervisor motivation, but also that on 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 middle manager recommendations, implying supervisors may view the training as substitutes to baseline stocks of these characteristics.

5.2.3 Decomposing Middle Manager Selection

The previous section established that much of what drives the middle manager recommendation is unobservable to the econometrician, even with the rich baseline data we collected. In this section, we use a simple framework to decompose middle manager selection into observable and unobservable components and to investigate which components drive the positive/negative relationship between middle manager selection and retention/productivity. The framework we use closely follows Dal Bó et al. (2021).[21] We apply the model to both productivity and retention outcomes separately. Therefore, in the general framework, gains can refer either to productivity and retention.

Suppose the middle managers are perfectly knowledgeable about the gains of the supervisors from training, and they recommend supervisors both based on the gains from training (Δ_p^i) and for other idiosyncratic reasons (Δ_c^i). For example, when we are focusing on productivity gains from training, the idiosyncratic component can include retention gains above and beyond productivity. Denote the value of recommending supervisor *i* as $\tilde{\Delta}^i$:

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

where X_i is the observable characteristics of supervisor i, $\tilde{\Delta}_p^i = \beta' X_i + \eta_i$ is the gains from training for supervisor i, and $\tilde{\Delta}_c^i = \psi' X_i + \theta_i$ is the idiosyncratic middle manager preferences for recommending *i*. Both the gains from training and the middle 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 gain (derived in Appendix Section B.2):

$$\mathbb{E}[\tilde{\Delta}_p^i | X_i, Rec_i] = \beta' X_i + \rho_{u\eta}\sigma_\eta \times \lambda(X_i, Rec_i) \tag{5}$$

[21] Our setup differs from theirs 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.

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 and be plugged in as a covariate to the estimating equations.[22] If the unobservable component of gains from treatment is negatively correlated with the entire unobserved component of the middle manager 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 middle manager selection is negatively related to treatment gains. β captures the effects of all the observable components. The estimated model can be used to compare different training allocation schemes.

We employ this general framework to decompose the relationship between middle manager selection and treatment gains to observable and unobservable components, both for gains in terms of productivity (where the middle manager selection negatively predicts treatment gains) and retention (where the middle manager selection positively predicts treatment gains). Because our productivity analysis is at the level of production lines, we perform this decomposition for productivity at the line level as well, where we use the averages of observable characteristics of supervisors tied to specific lines. Further, for simplicity, we collapse the during-training and post-training treatment effects into a single after-treatment-start treatment effect for the lines.

For each outcome, the analysis proceeds in two steps. First, we use a probit model to regress middle manager high-recommendation indicator on a set of observable characteristics. We use the estimates from the first stage to calculate the inverse mill ratio (IMR), denoted $\lambda(X_l, Rec_l)$ for lines and $\lambda(X_i, Rec_i)$ for individual supervisors . We then plug in the inverse mills ratio in the following modified versions of estimating equations 1 and 2, corresponding to our productivity and retention specifications:

$$\begin{aligned} \text{Productivity:} \quad & y_{ltr} = \alpha'(1[After]_{lt} \times T_l \times X_l) + \rho_{u\eta} \\ & \sigma_\eta \times \lambda(X_l, Rec_l) \times 1[After]_{lt} \times T_l + \delta_l + \mu_t + \gamma_r + \epsilon_{ltr} \qquad (6) \\ \text{Retention:} \quad & q_{tsi} = h_{0t} \times exp\left(\beta'(T_i \times X_i) + \rho_{u\eta}\sigma_\eta \times \lambda(X_i, Rec_i)\right) \qquad (7) \end{aligned}$$

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

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

negative.[23]

First columns of Appendix Tables C.8 and C.9 presents the probit results from the first stage. 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), middle manager 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 this first stage is around 19.7% at the line level (for production results) and 8% at the individual level. This is consistent with our earlier conclusion that observables explain only a small fraction of the overall middle manager selection patterns.

However, the fact that we cannot explain the majority of the variation in middle manager 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. The second column of Table C.8 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 of middle manager recommendation are partially driving the heterogeneity in treatment effects. For retention, we get a corresponding coefficient of -0.05, which implies that the unobservable elements of middle manager recommendations are positively correlated with retention gains (keeping in mind that in the retention model a negative coefficient implies lower quitting rates, hence a higher retention).

We use the estimated model to obtain predicted treatment effects for different allocation rules to investigate the role of the unobserved component of middle manager recommendations on the pattern of treatment effects. We present treatment effects under three alternative allocation rules following Dal Bó et al. (2021): (1) random assignment, (2) assignment based on middle manager recommendations, and (3) assignment based on middle manager recommendation with the effects of unobservables shut down.[24] The third allocation allows us to assess whether middle managers private knowledge of the supervisor unobservables is a driver of the treatment effects that would be generated under middle manager allocation. Appendix Figure C.4 presents the resulting treatment effects for line productivity and retention outcomes.

We first focus on retention. The model implies that, if half of the supervisors are treated, allocating the training based on middle manager recommendations outperforms random allocation by 35%. Importantly, when we shut-down the effects of unobservables on the treatment effects we get that middle manager recommendation yields treatment effects almost identical to random allocation.

[23]For retention, a negative treatment effect implies a lower probability of quitting, so a negative coefficient on the IMR implies middle manager selection is positively correlated with retention gains.

[24]Specifically, to shut down the effects of unobservables, we set the IMR to 0.

This suggests that the private information of the middle managers allow them to target supervisors with higher retention response to treatment. The firm would presumably not be able to replicate this allocation based on observable characteristics alone (even though observed characteristics in this instance include costly measures like skill scores elicited from the middle managers).

For line productivity, as expected from earlier analysis, the model suggests that random allocation significantly outperforms allocation by middle manager recommendation. With half of the lines treated, the average treatment effect is approximately 0 with middle manager allocation, while the random allocation yields an average treatment effect of 2.8 percentage points. With the effects of unobservables shut down, middle manager allocation yields an average treatment effect of 1.6. This suggests that private information of middle managers, which allows them to target retention, is negatively correlated with productivity gains. Taken together, it is clear that unobservables or private information held by the middle managers drive the pattern in the heterogeneity results: that recommended supervisors gain more in terms of retention while gaining little in terms of productivity.

Targeting Allocation Using Skill Scores. Suppose the firm is considering large scale adoption of the STITCH training, and decides to pilot the training to decide whether the investment is worth it. Piloting of this sort is indeed common in this context and it is rare that the firm would create a randomized pilot to assess average effects of the program. Rather the firm often pilots by letting middle managers choose some lines or workers to receive a particular intervention being considered. Accordingly, suppose in this instance the firm pilots by treating the supervisors of half the lines based on nominations of middle managers, our results suggest that they would get an average productivity effect of approximately zero and likely not adopt the program. Given the large average effect we document (and the large returns to investment we document in Section 6.3 below, not adopting the training is a costly error caused by the decentralization of the decision. This raises the question of whether there is a way to use the information that middle managers have about the supervisors without fully decentralizing the result.

We consider an alternate allocation rule that uses information gleaned from the middle managers about the skill stocks of workers. We focus on two simple and ex-ante reasonable allocation rules: allocate training to production lines with supervisors who have the lowest average baseline score for (1) management skills and (2) industrial engineering skills as elicited by the middle managers.[25] Appendix Figure C.5 presents the results treatment effects. As expected from previous analysis, these allocation schemes substantially outperform random assignment in terms of productivity gains. With half the lines treated, allocation based on management or industrial engineering skills leads to approximately 93% and 88% larger average treatment effects, respectively (ATE of 5.2 pp

[25] We consider these allocation rules as ex-ante reasonable because the STITCH training explicitly focuses on the management and industrial engineering skills we emphasize in our survey such as "assigning workers to operations", "meeting targets", and "line balancing." In contrast, technical sewing skills are not related to the content of the STITCH training.

and 4.9 pp, compared to the ATE of 2.8 pp under randomization). These result suggest that just eliciting information from the middle managers and making allocation decisions based on the elicited information is preferable to full decentralization or randomization in this context. However, this allocation scheme would likely not be incentive compatible over time. Middle managers can learn that elicited information is used for allocating training and distort the information they provide, undoing the informational content of their answers.

5.3 Alternative Explanations

Below we explore alternative explanations for the observed middle manager recommendations as they relate to productivity and retention gains.

Discrimination and Favoritism in Middle Manager Recommendations. One explanation for the negative relationship can be that middle managers engage in discrimination based on demographics or favoritism. If the characteristics middle managers 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 do not see strong evidence of discrimination/favoritism in our data. Many demographic characteristics (gender, age, caste, etc.) and measures that may relate to favoritism (coincident tenure, whether the supervisor and the middle manager 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 middle manager recommendation) and whether the supervisor is from out of state (which is associated with higher middle manager recommendation). In Appendix Table C.10, we directly look for evidence by regressing the middle manager recommendation on many demographic and coincident tenure related characteristics. The first column shows results for all supervisors. Across the 15 covariates, only whether the supervisor's native language is Kannada (native language of the region) is significantly and negatively related to middle manager recommendation. The R^2 of the model is only 1.6% and the joint F-statistic is 1.36 ($p = 0.17$). The second column focuses on the supervisors who are in our productivity analysis sample, a similar restriction to the LASSO exercise . The only variable with a statistically significant relationship is whether the supervisor is from out of state (which is associated with higher middle manager recommendation), consistent with the lasso exercise. The R^2 is still low (2.5%) and the joint F-statistic is insignificant ($p = 0.36$). Overall, 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 for the middle managers when allocating training. Yet, the available measures in our data do not indicate discrimination or favoritism as clear ulterior motives driving the recommendations.

Rewarding or Hoarding Productive Supervisors. An alternative view is that middle managers might view the training program as a reward for supervisors who are performing well. However, such a selection would presumably lead to a positive correlation between either baseline productivity or the skill scores of the middle managers. None of the three skill scores show up in our LASSO analysis as a good predictor of the middle manager recommendations. Baseline productivity of the supervisor's line does show up in this analysis, but is (weakly) negatively related to the recommendation. These patterns provide no evidence for rewarding productive supervisors through training allocation. In Appendix Table C.11, we present results from regressing the middle manager recommendation on the three skill scores and supervisor's motivation to improve, as elicited from the middle managers. The first two columns show results for the full sample. First, the three skill scores alone explain an exceedingly small fraction of the variation in middle manager recommendation with $R^2 = 0.5\%$. Second, after controlling for supervisor's motivation to improve, which is positively correlated with middle manager recommendations consistent with the LASSO analysis, there is a negative relationship between the middle manager recommendation and the technical skills of the supervisors. This is contrary to what we would expect to see if middle managers aimed to reward good supervisors with the training.[26]

Conversely, if middle managers believe that trained supervisors are more likely to get promoted and stop being a supervisor on the floor, they may recommend low-skill supervisors for training to hoard talent, similar to the mechanism explored in Haegele (2022). This is a harder interpretation to dismiss given the negative relationship between middle manager recommendation and baseline productivity and baseline technical skill score. However, these relationships are relatively weak. That is, the fact that none of the three skill scores show up as a predictor of rankings in the LASSO makes it unlikely that hoarding motive is the primary driver of the middle manager recommendations.

One possibility that would undermine this analysis is that middle managers may be strategically misreporting the skill scores of the supervisors under them, either to justify their recommendations or to mis-identify talented workers to avoid detection of hoarding. Two facts go against this possibility. First, the middle managers are asked about the supervisor skill scores before they are told about the training and asked about its allocation, making it unlikely that the training allocation is inducing them to misreport. Second, as we discuss in Section 5.2.1, the skill scores are meaningfully related to both baseline productivity and gains from training, making it unlikely that there is large scale strategic mis-reporting. Finally, in our context, it is unlikely that the middle managers believed the trained supervisor would leave the floor due to a promotion as such promotions are rare for supervisors in this firm.

Mistaken Beliefs About the "Production Function" of Training. In Section 5.2.1, we argue that supervisors possess private information about supervisor skills that, if used properly, can be utilized to allocate training effectively. Specifically, allocating training to supervisors who have been indicated by their middle managers to have a low level of baseline managerial and industrial

[26]For the subsample included in the productivity analysis (columns 3 and 4), this negative relationship between middle manager recommendation and technical score vanishes.

engineering skills beats randomization in terms of productivity gains. However, this does not rule out the possibility that middle managers have information about their supervisors, but they systematically misunderstand the production function of the training. For example, supervisors may believe that training is a complement to baseline stock of skills, instead of a substitute. Misallocation would then result not from a lack of knowledge about supervisors themselves, but about the training. While possible, we note that in a world where such a misunderstanding is the main mechanism behind the negative relationship between recommendation and productivity gains, we would expect there to be a strong relationship between supervisor skill scores (specifically for skills that middle managers believe are complementary to training gains) and recommendation. As discussed above, we do not observe this strong relationship.

Multiple Types of Middle Managers. Of course, the evidence we discuss above is not necessarily dispositive. In particular, it is possible that there are multiple types of middle managers ranking supervisors in countervailing ways. For example, some middle managers may be recommending high skill supervisors to reward them while others may be recommending low skill supervisors so as to hoard the high skill supervisors. Then, the average relationship between middle managers' perceptions of supervisors' skills and their recommendations might appear null. This is a difficult possibility to entirely refute, but we look for evidence that more than one approach or strategy is being employed in the data. We describe our approach in Appendix Section C.6.1. In short, we repeat our LASSO analysis of observable determinants of middle manager recommendations a 1000 times on random subsamples of middle managers and their reporting supervisors to see if there exists groups of middle managers that seem to employ different strategies or, in particular, have opposite relationships between key observable determinants such as skill ratings and training recommendations. Overall, we do not find much evidence to support this notion.

6 Additional Analysis

6.1 Spillovers

Given many production lines co-exist in the production floors in close proximity and these lines interact with and influence each other in many ways, productivity spillovers from training may be present in our context. For example, Adhvaryu et al. (2021) and Adhvaryu et al. (2020) demonstrates worker mobility and sharing across production lines is common in Shahi factories. Further, as our unit of analysis is the production line, if training leads to differential worker mobility patterns across lines in a production floor 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 (such as a decrease in abusive practices like motivation through yelling) across the production floor can further lead to positive spillover effects. Conversely, we might be concerned that the allocation of treatment to a subset of supervisors can lead to discouragement of the control group which negatively influences their productivity, or that middle managers can direct resources and attention to production lines with trained supervisors. If present, such negative spillovers can lead us to erroneously conclude that training leads to an increase in production, while in actuality what we observe is a decrease in productivity of the control group.

Our experimental design originally meant to study these spillover effects by randomly varying production floor level treatment saturation for a subset of the lines. For this subset, the original design was to have floors with 70% of the supervisors treated (high saturation) and 30% of the supervisors treated (low saturation). This design was imperfectly implemented due to complications with the matching of production lines to floors, leading to significant variation in the fraction of supervisors treated within saturation groups.[27] However, we still use the variation induced on the fraction of supervisors treated on a floor (saturation) to check for spillover effects. First, we divide the 54 production floors with saturation variation into three groups based on terciles of saturation level.[28] We then update our main productivity specification equation 1 to include the triple interaction between the training periods, floor level saturation tercile, and the line level treatment.

Figure 9 presents the results of floor saturation on productivity and the treatment effects.[29] The left panel charts the productivity effects of being on the second and third terciles of floor saturation for control lines, relative to the first tercile. While there is no evidence for spillovers during the training, there is suggestive evidence for positive spillovers after training. Specifically, the coefficient for being on the second and third terciles are 3.7pp (statistically significant at 10%) and 3.6pp (not statistically significant), respectively. The right panel of Figure 9 charts the treatment effects of training for different terciles. We do not see a clear pattern that suggests treatment effects are systematically different for higher or lower saturation. We note however that the treatment effects for the after training period is smaller now, which implies the long-run productivity effects we previously estimated are partly driven by the positive spillovers. Again, our results are imprecise and therefore largely suggestive. However, they overall point against the presence of negative spillovers in our context, ruling out that the productivity effects we observe are driven by negative spillovers to the control groups.

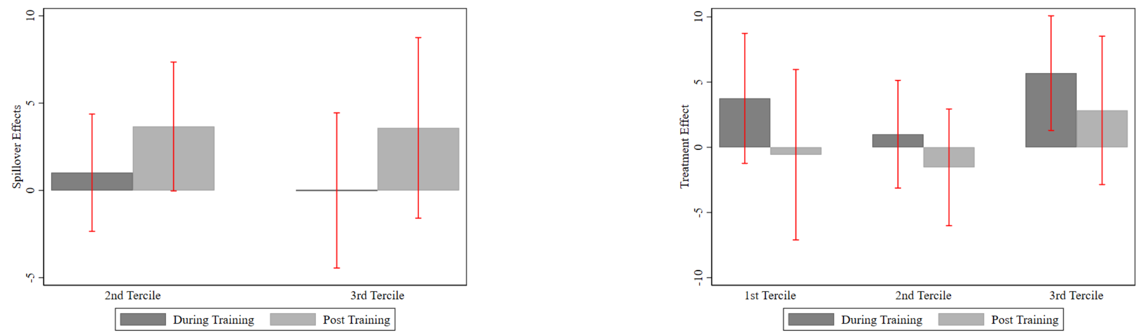
It is worth noting that the proposed framework to decompose the middle manager selection into observed and unobserved components in Section 5.2.3 does not take any possible productivity

[27]Many production lines were not properly matched to production floors at the time of randomization. Supervisors of these lines were not randomized with the correct treatment probability corresponding to the saturation group of the floors they worked on. Matching these supervisors to the correct floors after randomization lead to the dispersion of floor level treatment saturation.

[28]The first tercile covers saturation levels from 0 % to 33 %, the second tercile covers 35% to 63 %, and the third tercile covers 66 % to 100 %.

[29]The full regression results are reported in Appendix Table C.13. The values underlying Figure 9 correspond to our preferred specification with line, date and relative date fixed effects, presented in column 3.

Figure 9: Productivity Effects of Treatment Saturation



(a) Saturation Effect (Spillover)

(b) Treatment Effect

Note: Estimates underlying the figures are presented in Appnedix Table C.13 column 3. Left panel depicts the effects of saturation on productivity, reporting coefficients on the interaction between during/after training and second/third saturation terciles. The right panel depicts the treatment effect of training for different saturation terciles. Saturation is defined as the fraction of supervisors treated on the production floor.

spillovers into account. We proceed with this simplification for two reasons. First, our research question is about ascertaining whether the middle managers 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 pertinent to the question at hand. Second, our experimental design asks middle managers (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 middle manager 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. To confirm that controlling for floor-level saturation do not change the key result that highly recommended supervisors gain less in terms of productivity, in Table C.14, we run the middle manager recommendation heterogeneity specification presented in Section 4.3, with additional controls for event period and saturation tercile interactions. Consistent with our main results, we still see that lines with highly recommended supervisors gain less from training.

6.2 Additional Treatment Effects

6.2.1 Supervisor Salary

Appendix Table C.15 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 0.8 percentage points after controlling for retention effects (6 % of baseline). As shown in Appendix Table C.16, we do not find this effect to be heterogeneous by middle manager recommendation.

6.2.2 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 aggregate the 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 1, 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.[30]

Appendix Table C.17 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 26% increase in incentive payments during the training period (not statistically significant), and 37% increase six months after training (significant at 10 %).[31]. 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.[32] The results are very similar to the results using the full sample, indicating that the effects also accrue to the workers, not just the supervisors who have been trained or to managers.

[30]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.

[31]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 change from treating all the supervisors on the line using the formula $e^{\hat{\beta}-0.5\hat{V}(\hat{\beta})} - 1$ where $\hat{\beta}$ is the coefficient of interest and $\hat{V}()$ is the estimate of the variance.

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

6.2.3 Supervisor Attendance

Using the attendance roster, we also analyze the impact of training on daily attendance of supervisors. We use a specification similar to the difference-in-difference specification in equation 4, except with supervisor-day level observations with daily attendance as the outcome. The details and results are presented in Appendix Section C.10 and Appendix Table C.18. Overall, we do not find evidence that training increases supervisor retention.

6.2.4 Worker Attendance and Retention

Finally, we investigate whether workers who were subjected to the treatment through treated supervisors have differential attendance and retention outcomes. Details of the analysis and the results are presented in Appendix C.11. 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.

6.3 Returns on Investment

Finally, we quantify the profit and rate of return to the firm from STITCH training by combining our effect estimates on productivity, wage growth, and incentive payments with program cost data and inputs from the accounting department of the firm. Table 6.1 presents our net profit and rate of return calculations. On the cost side, we combine the direct costs of developing and implementing the STITCH training, additional incentive payments to treated lines, and increased salaries of supervisors.[33]

On the benefit side, we exclusively focus on the 480 production lines included in our main productivity analysis sample. This is conservative as it implicitly assumes the gains for the lines not included in this sample is 0.[34] We combine our productivity effect estimates with the target quantity information for each line-day and revenue/profit margin we obtained from the firm. Overall, counting flow benefits up till six months after training end, the NPV of the benefits from additional productivity is more than $4.5 million. On the cost side, we consider direct costs of the program, such as trainer salaries, equipment, food, and program development costs ($13,085), costs associated with additional incentive payments for lines with treated supervisors ($31,976), and increased salary of treated supervisors ($ 36,815). Overall, we estimate the total costs to be around $82,000. The net profit from the program considering cost and benefit flows up through 6 months after training end is $4,460,669 (corresponding to $4807 per overall treated supervisor and $10,302 per treated supervisor working at one of the 480 analysis lines). The net rate of return is thus around 54 times

[33]Because the training took place on Sundays, the off day of supervisors, costs associated with lost production hours are not a part of our cost calculation.

[34]We also do not consider spillovers in this analysis. This is also likely conservative as available evidence implies some positive spillovers.

the training cost.[35]

Finally, we undertake a simple back-of-the-envelope calculation to give a rough estimate of how much additional supervisor turnover would need to cost (above and beyond its productivity effects which are already internalized in our line-level intent-to-treat productivity impact estimate) for the middle manager allocation to be justified from the perspective of the firm. To do this, we compare the avoided supervisor turnover to the missed productivity gains that would result from the middle manager allocation. On the retention side, we note that around 35% of the supervisors in the control group leave the firm by the end of our period.[36] Applying this percentage, we assume 325 of the 928 supervisors in the treatment group would leave the firm absent treatment. The quitting hazard ratio for treated among the highly recommended supervisors, presented in Appendix Table C.7, is 22% (while it is almost 0 for low treatment supervisors). Using this point estimate, we conclude that the middle manager allocation would avoid 72 of the 325 turnovers that would take place absent training. We conservatively assume that the middle manager allocation would lead to half the productivity benefits that we observe from the random allocation, while keeping the costs the same.[37] This implies that each additional turnover would need to cost the firm approximately \$31,000 for the middle manager allocation to be profitable from the firms perspective, as compared to the random allocation.[38] In the extreme case of approximately zero productivity gains under middle manager allocation (which is not unrealistic given the patterns we document above), each supervisor replacement would need to cost the firm around \$62,000 to make the middle manager allocation better than random allocation from the firm's perspective. To put this in perspective, these costs are roughly 11 to 23 times the average annual earnings of a supervisor in our study, which are around \$2,700 at the start of the training. Given the large magnitudes, it is implausible that the additional retention from the middle manager allocation would make up for the foregone productivity gains from the firm's perspective. Note that these calculations compare the middle manager recommendation allocation to random assignment, but the middle manager recommendation assignment rule is even more starkly suboptimal when compared to an alternative assignment rule which outperforms random assignment (e.g., one that uses baseline skill deficiencies elicited from the middle managers as discussed in section 5.2.3).

[35]Note however the cost calculations do not include the cost of the authors' expertise in conducting the prior study (Adhvaryu et al., 2022c) which informed the curriculum. One could argue that prior to this study the cost of acquiring the content for this curriculum would be extremely high for the firm.

[36]This value only includes supervisors who were present at the start of the training.

[37]In fact, we know that the productivity gains are heavily concentrated in lines with lower average middle manager recommendations, so the productivity benefits would likely fall substantially more.

[38]The firm reports that quantifying the cost of replacing workers who leave is in general difficult. Though they had no such calculation for supervisors, their best assessment of the cost of replacing a machine operator was roughly 20,000 INR which amounts to roughly 300 USD at the time of the study. This is less than 1% of the cost needed to justify the middle manager training allocation with respect to returns to the firm. The cost of replacing supervisors would, if anything, be less given that machine operators are mostly recruited from distant villages, trained for several months and relocated at the firm's expense to the city, while supervisors are generally hired from other nearby factories or promoted from within the firm.

Table 6.1: Return on Investment Calculations for 6 Months After Program End

Total Benefit (Only For the 480 Sewing Lines Included in Analysis Sample)	$4,542,544
Additional Productivity (Lines with STITCH Trained Supervisors)	$4,542,544
Total Cost	-$81,875
STITCH Training Cost (Development, Trainer Salary, Materials, and Refreshments)	-$13,085
Additional Incentive Payments (Lines with STITCH Trained Supervisors)	-$31,976
Increased Salary (STITCH Trainees)	-$36,815
Net Benefit	**$4,460,669**
Net Rate of Return	**54X**
Assumptions	
Revenue per Additional Garment	$7
Profit Margin on Revenue from Additional Productivity	20 %
Interest Rate	10 %
Exchange Rate (INR per 1 USD)	65

Note: All values in April 2017 present values. Productivity calculations only covers the 480 lines included in our main analysis sample. Period of interest is from training start to 6 months after training end for each factory, consistent with our analysis. Additional garments due to training is calculated by multiplying the average target quantity for a given line-month with the relevant coefficient (based on during/post training) and the fraction of line supervisors treated. We then assume there are 25 production days on a given month. Revenue per additional garment is taken from the accounting department of the firm. Profit margin on revenue from additional productivity is calculated as 80% of the percent labor contribution to cost (25%) as guided by the accounting office. For each line-day, we find the treatment effect coefficients on incentive payments are 26 INR and 16 INR for during/after training. We add this cost multiplied by fraction of supervisors treated for each line-day for the included lines to get additional incentive payments. For increase in salary, we multiply treatment effect on salary growth (0.008) with average baseline salary of supervisors in April 2017 (14,855 INR) with the number of trained supervisors for each month. Observe that this is conservative as it includes all trained supervisors (not just the ones for the analysis lines) and assumes percent increase in salaries take place immediately. Materials and food cost amounted to 30,000 INR and 150,000 INR respectively. Cost of development was 70,500 INR. Trainer salaries were 600,000 INR. Exchange rate is the rate in the beginning of April 2017.

7 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 Van Reenen, 2011). These studies argue that middle managers may 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 elicited from middle managers rankings of which line supervisors should be prioritized for training and then 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 middle managers to receive the training actually gaining little to nothing from the training in terms of productivity.

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 were more likely to quit in the absence of training than were low recommendation control supervisors. We adapt a recent approach by Dal Bó et al. (2021) to decompose the allocation decisions of middle managers into observable and unobservable components. This analysis confirms that substantial variation (at least 80%) in middle manager 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 middle managers 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, middle managers 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 middle managers 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 middle managers appear 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. 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 middle managers) the gains and return on investment would have been negligible.

Indeed, we note that the very design of the trial reported on in this paper was motivated by anecdotal conversations with upper management at the firm regarding how investments like the one we evaluate get piloted and rolled out in the firm. These conversations revealed that many such programs are proposed and considered over the course of the year, often from buyers with whom the firm wants to maintain a strong relationship (Adhvaryu et al., 2020). Given these programs are costly particularly in terms of time and effort for their coordination and implementation, the firm often pilots these programs with a subset of production lines or workers before deciding to roll them out across the entire firm.

The most likely way these pilot lines and workers are selected is via a decentralized recommendation much like the one we elicited in the study. Accordingly, if the firm were to undertake exactly this pilot approach with respect to the program we study here, we note that they would have believed the gains to be null and would have aborted the program after the pilot, forfeiting more than $4.5M in gains. In this sense, our results provide one potential explanation for why managerial quality remains low on average and highly varied in many firms despite strong academic evidence of potential gains from investments in management such as the training program we evaluate. That is, when returns to such investments are heterogeneous (as might be expected given the heterogeneity in baseline stocks of these skills and productivity across supervisors and teams within the firm) and allocation of costly resources is decentralized (potentially as a means of piloting to inform investment decisions), investment decisions may be made on the basis of inaccurate estimates of returns leading to underinvestment.

References

Acemoglu, D. (1997). Training and innovation in an imperfect labour market. *The Review of Economic Studies*, 64(3):445–464.

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.

Acemoglu, D. and Pischke, J.-S. (1998). Why do firms train? Theory and evidence. *The Quarterly Journal of Economics*, 113(1):79–119.

Acemoglu, D. and Pischke, J.-S. (1999). Beyond Becker: Training in imperfect labour markets. *The Economic Journal*, 109(453):112–142.

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. (2022a). Management and shocks to worker productivity. *Journal of Political Economy*, 130(1):1–47.

Adhvaryu, A., Kala, N., and Nyshadham, A. (2022b). Returns to on-the-job soft skills training. *Journal of Political Economy*, forthcoming.

Adhvaryu, A., Nyshadham, A., and Tamayo, J. (2022c). Managerial quality and productivity dynamics. *The Review of Economic Studies*, forthcoming.

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.

Alan, S., Corekcioglu, G., and Sutter, M. (2022). Improving Workplace Climate in Large Corporations: A Clustered Randomized Intervention*. *The Quarterly Journal of Economics*. qjac034.

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

Becker, G. S. (1964). *Human Capital*. New York: Columbia University Press.

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.

Bilinski, A., Poe, J., Roth, J., and SantAnna, P. (2022). Whats trending in difference-in-differences? a synthesis of the recent econometrics literature. Technical report.

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., 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.

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

Borusyak, K., Jaravel, X., and Spiess, J. (2022). Revisiting event study designs: Robust and efficient estimation. Technical report.

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.

de Chaisemartin, C. and D'Haultfuille, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, 110(9):2964–96.

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

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, 225(2):254–277. Themed Issue: Treatment Effect 1.

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.

Haegele, I. (2022). Talent hoarding in organizations. Technical report.

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.

Stogdill, R. M. and Coons, A. E. (1957). Leader behavior: Its description and measurement.

Sun, L. and Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, 225(2):175–199. Themed Issue: Treatment Effect 1.

APPENDIX

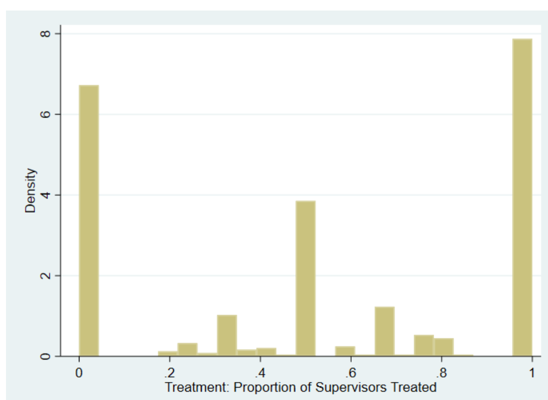
Not for publication.

A Experiment Details

A.1 Experiment Timeline

Figure A.1: Timeline of Experiment and Data Collection

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



A.2 STITCH Modules

The STITCH training is composed of 4 modules and 25 hour-long sessions. Below, we summarize the contents of each session:

A.2.1 Module 1: Me as a Person

- **Introduction:** Introduction to the STITCH program. Includes interactive activities for participants to introduce themselves and to reflect on and discuss their aspirations.
- **Who am I:** Focus on enabling participants to think and reflect about themselves and what they value both in work and in life. Underlines the importance of health, knowledge/skills, and good relations for contentment in all spheres of life.

- **My Self Esteem:** Pair activity focusing on building self-esteem and confidence by better understanding of own strengths. Importance of building self-esteem within team members at work and how the fast pace of work and demand for productivity can lead to an environment detrimental to self-esteem.
- **My Behaviours and Values:** Focus on understanding behaviors required for an effective work place and the link between values and behaviors. Participants are asked to list their commitments towards developing and maintaining effective behaviors.
- **Handling Emotions:** Role-play activity to understand emotional responses and their impact. Focus on the importance of holding immediate reactions to situations and positive actions to manage emotions.
- **Managing Stress:** Role-play activity to better understand causes, reactions, and effects of stress. Tips for effective stress management, at work and in other spheres of life.
- **Being Sensitive:** Focus on the importance of sensitivity in interpersonal relationships to managing the emotions and stress of others. Role-play activity focusing on interactions between workers, supervisors, and managers to reflect on and build sensitivity.
- **Gender Sensitivity:** Focus on enabling the participants to understand the impact of socialization on attitudes and mind-sets of men and women. Reflect on how these attitudes effect behavior and outcomes at work. Participants are asked to identify one action they commit to undertaking for sensitivity.
- **Effective Communication:** Focus on understanding communication styles and develop skills to communicate assertively and responsibly. Role-playing and brainstorming activity to understand outcomes of different communication styles.

A.2.2 Module 2: Me as a Supervisor

- **My Role:** Focus on understanding the roles supervisors perform on a day-to-day basis. Highlight the importance of both technical responsibilities and people management responsibilities. Group discussion about what knowledge, skills, and attitudes supervisors need at possess in order to be effective in both their production and human resource management roles.
- **Understanding the Need for Building Capacity:** Focus on the importance of and tips for using time more effectively in order to find time to develop and enhance new skills. A time mapping-exercise to identify all the activities supervisors do in a given day and see how much time is spent on activity categories such as planning, problem solving, communication with management etc.
- **Planning and Organizing:** Focus on the importance of planning and organizing, especially while working as a team, and the options available within given work situations. Session composed of a team game to underline the need for planning in team work and a group activity where the groups are asked how they would plan for a hypothetical scenario (such as a specific order quantity they have to fulfill in 20 days) and discuss their options.
- **Solving Problems:** Focus on problem solving skills using case studies and role play. Underlines skills such as problem identification, analyzing the root cause of the problem, making decision based on available options, implementing the decision, and reviewing the outcomes of the decision. Includes a session where participants think of creative solutions to presented problems.

- **Conflict Resolution:** Focus on enabling supervisors to understand the causes/consequences of unresolved conflicts and helping them understand different styles of handling conflict situations (competing, collaborating, compromising, avoiding, and accommodating). Includes case studies for participants to work through and discuss.
- **Building Respect-Preventing Harassment:** Focus on helping participants understand and reflect on what constitutes harassment and its impacts, clarifying company policy regarding sexual harassment, and finding better ways to be effective at work. Case studies to clarify and discuss what constitutes harassment and how it can be prevented.

A.2.3 Module 3: Me as a Member

- **Understanding My Team:** Introductory session focused on the importance and benefits of teamwork. Session is mainly composed of a team game that needs to be completed without any effective communication to underline the importance of teamwork. The trainer than discusses with the participants their experiences with the activity and importance of teamwork in their work.
- **Managing My Team:** Focus on understanding various stages of team development and different leadership styles. The emphasis is on demonstrating Tuckman's stages of group development (forming, storming, norming, performing) using role playing exercises. The trainer than discusses appropriate leadership styles for the various stages.
- **Building Accountability:** Focus on the importance of responsibility and accountability for their work and functioning of their teams. At the end of the session, the participants are asked to make an action plan to create ownership and accountability within their teams.
- **Balancing Technical and Human Competence:** Focus on the importance of working on human management skills alongside technical skills to create better performing teams. Session is composed of a role-playing activity and a self-assessment questionnaire for participants to complete and reflect on.
- **Employee Motivation and Engagement:** Focus on the need for motivation/encouragement and tips for effective motivation and engagement. The session covers the Herzberg's theory of motivation and a role-playing activity to demonstrate the varying motivations of different members and how a leader can support such motivations to be effective. It emphasizes the need for showing appreciation, helping workers realize their value and encouraging learning and ownership.

A.2.4 Module 4: Me as a Leader

- **My Learning Orientation and Growth:** Introductory session focusing on factors that contribute to learning and growth and how to create an environment that facilitates growth. Importance of a growth-mindset as opposed to a fixed mindset.
- **Feedback and Coaching:** Focus on the importance both giving and receiving constructive feedback for developing required skills. Session is mainly composed of a role-playing activity to practice the skill of giving and eliciting both positive and negative feedback.

- **Managing Change:** Focus on the need and the impact of change and how to manage the process of change. The session includes an activity where the participants are divided into groups and each group is given a task to complete. In the middle of the process, a change is introduced to the task and participants are asked questions about the process of adaptation to this change.
- **Building a Culture:** Focus on the importance and the formation of work culture. Participants are asked to brainstorm about the traits of a good work culture and then are asked in groups to develop an action plan towards achieving a positive work culture. The trainer then highlights the importance of the supervisors for work culture and discusses how good management practices, such as providing timely feedback, can contribute to a better culture.
- **Being a Role Model:** Closing session that focuses on the concept of self-management and on developing skills and knowledge required to be a role model. The trainer explains how all components discussed in the program helps individuals develop and become role models. The participants are asked to reflect on their leanings and how these leanings can help them with their aspirations they shared in the first session.

B Model and Analysis Details

B.1 Derivation of Remark

Before we establish remark 1, we show the following lemma holds:
Lemma: $\frac{d\tau_p^*}{dc} < 0$, 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 middle managers 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 the remark 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. (2022c), 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.[39] 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.

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

B.4 Lasso Procedure and Included Variables

We use Stata's $lasso$ command to implement a linear lasso model where the penalty term λ is selected through 10-fold cross validation. The selected $\lambda = 0.094$ with an out of sample R^2 of 2.4$\%$. The list of variables included in the lasso procedure is below. 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
- **Middle Manager and Supervisor Joint Characteristics**: From same state, Same religion, Same gender, Supervisor hired after middle manager, 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 tailoring skill, Industrial engineering skill, Managerial skill, Training interest, Expected gain from training, Amount supervisor would allocate to training, Self efficacy index, Instrumentality of training
- **Moddle Manager Assessment of Supervisor:** Technical tailoring skill, Industrial engineering skill, Managerial skill, Motivation to improve, Months supervising current line
- **Other:** Cognitive ability, Risk preference, Discount index, Baseline line productivity, Suggested hires last month

B.5 Creation of Survey Indices

Below table outlines what questions are used and how they are combined for the creation of the indices from the baseline supervisor surveys.

Index	Method	Question Text
Conscientiousness	Supervisors were asked to rate how much they agreed with the following statements (5 = "strongly agree", 1 = "strongly disagree"). The index is created by adding up the points. "(-)" next to the question implies that the scores were reversed before adding up.	I am always prepared
		I pay attention to details
		I get chores done right away
		I carry out my plans
		I make plans and stick to them
		I procrastinate and waste my time (-)
		I find it difficult to get down to work (-)
		I do just enough work to get by (-)
		I don't see things through (-)
		I shirk my duties (-)
Locus of Control	Supervisors were asked to rate how much they agreed with the following statements (5 = "strongly agree", 1 = "strongly disagree"). The index is created by adding up the points. "(-)" next to the question implies that the scores were reversed before adding up.	I believe that my success depends on ability rather than luck
		I believe that unfortunate events occur because of bad luck (-)
		I believe that the world is controlled by a few powerful people (-)
		I believe some people are born lucky (-)
		I believe in the power of fate (-)
Perseverence	Supervisors were asked to rate how much they agreed with the following statements (5 = "strongly agree", 1 = "strongly disagree"). The index is created by adding up the points. "(-)" next to the question implies that the scores were reversed before adding up.	I don't quit a task before it is finished
		I am a goal-oriented person
		I finish things despite obstacles in the way
		I am a hard worker
		I don't get sidetracked when I work
		I don't finish what I start (-)
		I give up easily (-)
		I do not tend to stick with what I decide to do (-)
Self Esteem	Supervisors were asked to rate how much they agreed with the following statements (5 = "strongly agree", 1 = "strongly disagree"). The index is created by adding up the points. "(-)" next to the question implies that the scores were reversed before adding up.	On the whole, I am satisfied with myself
		At times, I think I am no good at all (-)
		I feel that I have a number of good qualities
		I am able to do things as well as most other people
		I feel I do not have much to be proud of (-)
		I certainly feel useless at times (-)
		I feel that I am person of worth, at least on an equal plane with others
		I wish I could have more respect for myself (-)
		All in all, I am inclined to feel that I am a failure (-)
		I take a positive attitude toward myself
Consideration	Supervisors were asked to rate how much they agreed with the following statements (5 = "strongly agree", 1 = "strongly disagree"). The index is created by adding up the points. "(-)" next to the question implies that the scores were reversed before adding up.	I do personal favors for workers in my line
		I do little things to make it pleasant to be a member of my line
		I am easy to understand
		I find time to listen to members of the line
		I keep to myself (-)
		I look out for the personal welfare of individual workers on my line
		I refuse to explain my actions (-)
		I act without consulting the line (-)
		I back up the workers in my line in their actions
		I treat all workers in my line as my equals

Initiating Structure	Supervisors were asked to rate how much they agreed with the following statements (5 = "strongly agree", 1 = "strongly disagree"). The index is created by adding up the points. "(-)" next to the question implies that the scores were reversed before adding up.	I make my attitudes clear to the workers in my line I try out new ideas with my line I rule with an iron hand I criticize poor work I assign workers on the line to particular tasks I speak in a manner not to be questioned I schedule the work to be done I maintain definite standards of performance I emphasize the meeting of deadlines I encourage the use of uniform procedures
Conflict Index	Supervisors were asked to rate how frequently following statements would occur (5 = "very frequently", 1 = "very infrequentlye"). The index is created by adding up the points.	Had direct reports that resisted your initiatives Had interpersonal conflicts between you and at least one of your key direct reports Had employees who were used to doing things they way they had been done and were reluctant to change Had key members of your staff that were incompetent, unmotivated, technically obsolete, or otherwise poor performers Had key direct reports that lacked the experience to do their jobs without close supervision
Problem Index	Supervisors were asked to select which of the following production related problems they encounter in day-to-day operation. The index is created by summing the number of distinct problems supervisors reported facing.	Malfunctioning of machines Shortage of inputs or inventory Production errors made by workers Worker absenteeism Workers arriving late Unmotivated or shirking workers Unrealistic targets Other (Specify)
Autonomous Problem Solving	Supervisors indicate whether they agree or disagree with the following statements. The formula for the index is as follows: 1[Supervisors ID Problem] + 1[Supervisors solve problem] - (1[Workers ID Problem] + 1[Workers solve problem]) - (1[Middle Managers ID Problem] + 1[Middle Managers solve problem]) - 1[Problems Fix Themselves]	Problems are brought to my attention by workers I discover problems while doing rounds of the line The floor-manager points out the problems to me I usually interact with the line workers and find a solution I usually find solutions on my own I usually involve the floor-incharge to find a solution I usually don't do anything, the problems fix by itself
Target Effort Index	Supervisors were asked to select which of the following actions they take to ensure production targets are met. The index is created by summing the number of distinct actions supervisors reported taking.	Do rounds of the line to ensure things are in order Talk to workers individually Provide positive reinforcement to high-performing workers Make low-performing workers aware of their work Demonstrate the way of work by example Other (Specify)
Monitoring Frequency		On a normal day, how often do you make rounds of the line to ensure that production is running smoothly?
Communication Index	(Worker discussion frequency * 1[Supervisor initiates discussion]) + (Middle Manager discussion frequency * 1[Supervisor initiates discussion]) + (Upper manager discussion frequency * 1[Supervisor initiates discussion]) +	How often do you discuss efficiency and performance with the workers in your line? Typically who initiates the discussion about efficiency and performance? [with workers] How often do you discuss efficiency and performance with the floor-incharge? Typically who initiates the discussion about efficiency and performance? [with floor-incharge] How often do you discuss efficiency and performance with someone superior like a manager (higher than the floor-incharge)?

		Typically who initiates the discussion about efficiency and performance? [with managers]
Active Personnel Management	Supervisors were asked to select which of the listed actions they would take under the following scenarios. The index is created by summing the number of distinct actions report they would take.	**Suppose you have a worker who is an under-performer. He or she may not be very motivated, and may not have the right skills for the job. What steps would you take to address this issue?** Talk to the worker directly and in person Talk to the worker directly in presence of other workers Talk to the floor-incharge about the worker Try to replace the worker **Now suppose you have a worker who is a star-performer. He or she is extremely productive, does all tasks on time and always meets targets. What steps would you take to retain this worker?** Commend the worker on their work/effort Praise the worker in presence of other workers Put in good word for workers to floor-incharge or other superiors Recommend the worker for a promotion Do you publicly or privately commend such workers for their effort? Do you talk to the floor-incharge or other superiors about this worker? Do you recommend such workers for promotion?
Self Efficacy Index	Supervisors were asked to rate how much they agreed with the following statements (5 = "strongly agree", 1 = "strongly disagree"). The index is created by adding up the points. "(-)" next to the question implies that the scores were reversed before adding up.	I have strong learning abilities It takes me time to assimilate the contents of training (-) If the course is too abstract, I easily get lost (-) I can easily memorize training materials I am able to follow even if the trainer goes quickly I know that I will be able to learn the contents of the training
Instrumentality Index	Supervisors were asked to rate how much they agreed with the following statements (5 = "strongly agree", 1 = "strongly disagree"). The index is created by adding up the points.	This training is important for helping me acquire skills This training is important to my self confidence at work This training is important to my efficiency at work This training is important for helping me grow as a person This training is important for helping me perform my job better This training is important to increase my job security This training is important to increase my salary
Cognitive Ability	The cognitive ability index is created by averaging the (normalized) scores from a digit span recall test, an arithmetic test, and whether the supervisors has chosen any dominated options in the risk and time preference questions.	

C Additional Checks and Results

C.1 Additional Line Balance

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

	Analysis Subsample				Analysis Subsample w/ Middle Manager			
	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)

Note: 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 middle manager 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.

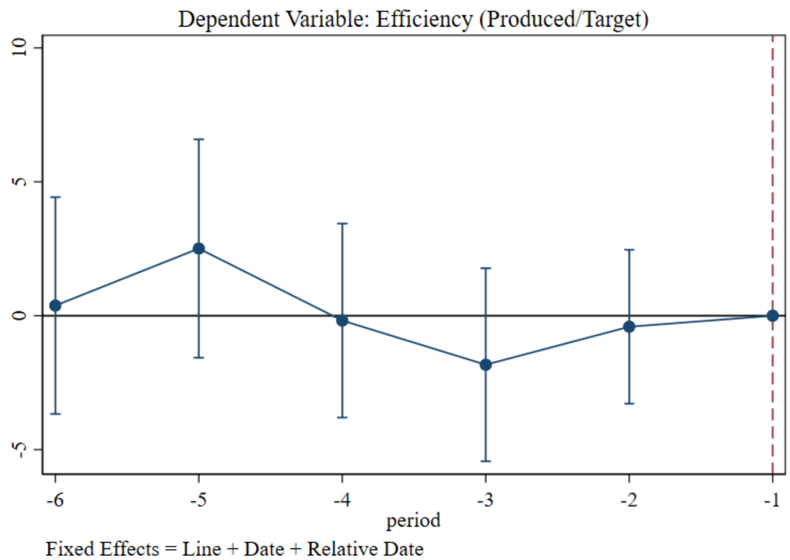
C.2 Pre-Post Test Scores

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

$$s_{i2} = \beta_0 + \beta_1 T_i + \beta_2 s_{is1} + \mu_s \tag{8}$$

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 C.2. Across all modules, treatment leads to a significant gain in the post-module tests.

Figure C.1: Pre-Period Event Study Coefficients



Note: Event study coefficients for months preceding training. The model is run for the full analysis period.

Table C.2: Treatment Effect on Post-Module Exam Scores

	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

Note: 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.

C.3 Additional Productivity Results

C.3.1 Balanced Relative Month Panel

Table C.3: Effects of Training on Line Productivity - Panel Balanced on Relative Month

	Outcome: Efficiency (Produced/Target)				
	Analysis Lines			Lines w/ Middle Manager Match	All Lines
	(1)	(2)	(3)	(4)	(5)
During Training X Treatment	4.002***	4.009***	4.092***	3.865***	4.300***
	(1.109)	(1.111)	(1.111)	(1.247)	(1.286)
After Training X Treatment	2.789**	2.791**	2.893**	2.848**	2.829*
	(1.260)	(1.262)	(1.228)	(1.433)	(1.575)
Observations	274300	274299	274299	226915	305341
Number of Lines	480	480	480	395	553
Cont. Mean of Dep. Var.	55.533	55.533	55.533	55.533	55.533
Line FE	X	X	X	X	X
Month FE	X				
Day FE		X	X	X	X
Relative Date FE			X	X	X

Note: Standard errors are clustered at line level (*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$). The analysis covers six months prior to twenty months after training start for each line. 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 line-days with 0 efficiency.

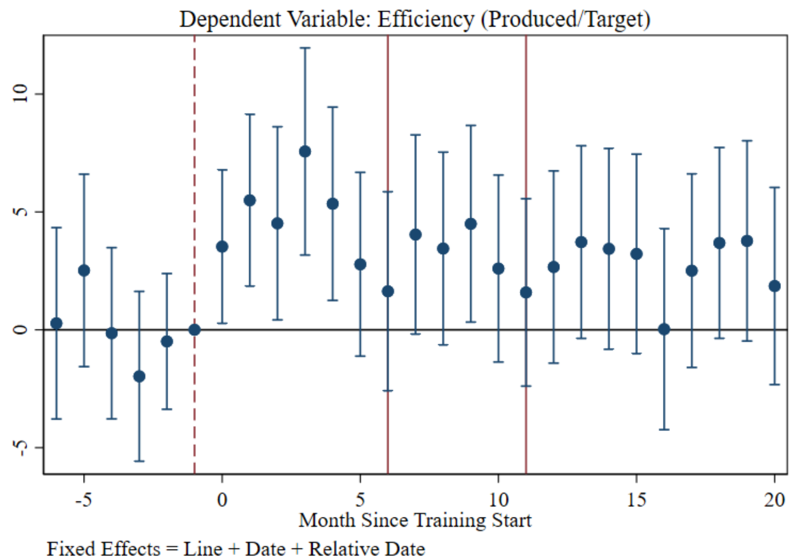
C.3.2 Dynamic Specification

Appendix Figure C.2 charts the monthly coefficients estimates β_m from the following event study specification:

$$y_{ltr} = \sum_{m=-6}^{20} \beta_r (T_l \times D_{ltr}^m) + \delta_l + \mu_t + \gamma_r + \epsilon_{ltr} \tag{9}$$

where y_{ltr} is productive efficiency in line l, date t, and relative date r, T_l is line level treatment, D_{ltr}^m is an indicator for whether the date is within m months since treatment start for the factory, and δ_l, μ_t, γ_r are fixed effects.

Figure C.2: Event Study Results



Note: Figure shows β_m from estimating equation 9. Month 0 signifies treatment start. β_{-1} is normalized to 0. Shortest training duration is 6-months (first solid red line). Longest training duration is 11 months (second solid red line). 95% confidence intervals are shown.

C.3.3 Middle Manager Ranking Heterogeneity

Table C.4: Productivity Effect Heterogeneity by Middle Manager Recommendation

	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

Note: 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.

C.3.4 Results with Binary Treatment Definition

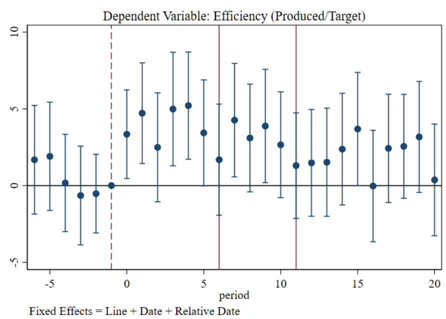
We replicate our line-level productivity results with an alternative binary line-level treatment definition: at least one line supervisor is treated. Given lines have varying number of supervisors, the probability of having at least one treated supervisor differs substantially across lines. Therefore, for each line, we calculate the probability of having at least one treated supervisor given our randomization scheme. We then show results for three weighting schemes: (1) unweighted, (2) inverse probability weighted (IPW) to recover ATEs, and (3) IPW with lines with extreme weights dropped. Appendix Table C.5 shows average productivity results, corresponding to Table 4.1 in the body. Appendix Figure C.3 shows monthly event studies estimated using estimating equation 9. Appendix Table C.6 show middle manager recommendation heterogeneity. Overall, our results are consistent across continuous and binary treatment definitions.

Table C.5: Effects of Training on Line Productivity - Binary Treatment

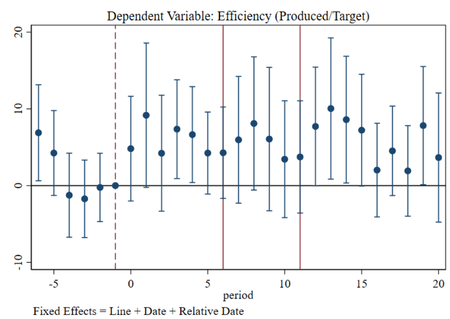
	Outcome: Efficiency (Produced/Target)								
	No Weighting			IPW			IPW, Drop If Weight >20		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Treatment X During Training	3.215***	3.221***	3.095***	4.141*	4.156*	4.585**	3.517***	3.520***	3.561***
	(1.023)	(1.025)	(1.020)	(2.189)	(2.202)	(2.039)	(1.201)	(1.204)	(1.206)
Treatment X After Training	2.450**	2.439**	2.274**	5.704**	5.657**	5.993**	2.703*	2.686*	2.805**
	(1.131)	(1.133)	(1.119)	(2.622)	(2.640)	(2.614)	(1.378)	(1.383)	(1.362)
Observations	228167	228166	228166	228167	228166	228166	222350	222349	222349
Number of Lines	480	480	480	480	480	480	468	468	468
Control Mean of Dependentent Variable	55.865	55.865	55.865	55.865	55.865	55.865	55.865	55.865	55.865
Line FE	X	X	X	X	X	X	X	X	X
Month FE	X			X			X		
Day FE		X	X		X	X		X	X
Relative Date FE			X			X			X

Note: Standard errors are clustered at line level (*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$). Treatment is a binary indicator for having at least one treated supervisor on the line. 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.

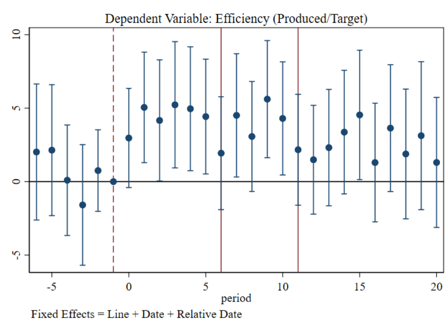
Figure C.3: Monthly Event Study Results - Binary Treatment



(a) Unweighted



(b) IPW



(c) IPW - Drop if Weight > 20

Note: Figure shows β_m from estimating equation 9. Month 0 signifies treatment start. β_{-1} is normalized to 0. Shortest training duration is 6-months (first solid red line). Longest training duration is 11 months (second solid red line). 95% confidence intervals are shown.

Table C.6: Productivity Effect Heterogeneity by Middle Manager Recommendation - Binary Treatment

	Outcome: Efficiency (Produced/Target)								
	No Weighting			IPW			IPW, Drop If Weight > 20		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Treatment X During Training	4.320**	4.312**	4.314**	7.763**	7.755**	8.032**	6.087***	6.088***	6.068***
	(1.694)	(1.699)	(1.679)	(3.386)	(3.392)	(3.206)	(2.196)	(2.200)	(2.183)
Treatment X After Training	4.457**	4.427**	4.429**	10.843***	10.731***	11.060***	6.621***	6.598***	6.541***
	(1.975)	(1.978)	(1.965)	(4.005)	(4.036)	(4.116)	(2.226)	(2.231)	(2.218)
Treatment X During X High Rec	-3.263	-3.235	-3.548	-8.434**	-8.413**	-8.706**	-5.550**	-5.552**	-5.587**
	(2.224)	(2.229)	(2.233)	(3.809)	(3.806)	(3.610)	(2.651)	(2.652)	(2.652)
Treatment X After X High Rec	-4.364*	-4.319*	-4.718*	-11.276***	-11.140***	-11.404**	-7.636**	-7.624**	-7.632**
	(2.509)	(2.513)	(2.499)	(4.269)	(4.285)	(4.418)	(3.000)	(3.003)	(2.969)
Observations	189381	189380	189380	189381	189380	189380	183564	183563	183563
Number of Lines	395	395	395	395	395	395	383	383	383
Control Mean of Dependentent Variable	55.279	55.279	55.279	55.279	55.279	55.279	55.279	55.279	55.279
Line FE	X	X	X	X	X	X	X	X	X
Month FE	X			X			X		
Day FE		X	X		X	X		X	X
Relative Date FE			X			X			X

Note: Standard errors are clustered at line level (*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$). Treatment is a binary indicator for having at least one treated supervisor on the line. 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.

C.4 Additional Retention Results

Table C.7: Retention Effects by Middle Manager Recommendation

	Supervisor Quit	
	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

Note: *** $p < 0.01$, **$p < 0.05$, * $p < 0.10$. The sample is restricted 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 middle manager recommendation. Second column limits the sample to supervisors with low middle manager recommendation.

C.5 Model Results

Table C.8: 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 to Imporove (Scored by Middle Manager)	0.243*	0.574
	(0.129)	(1.921)
Months as Supervisor (Answered by Middle Manager)	-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 Middle Manager)	-0.041	3.136
	(0.129)	(2.188)
Industrial Engineering Skills (Scored by Middle Manager)	-0.180	-1.521
	(0.142)	(2.065)
Management Skills (Scored by Middle Manager)	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	

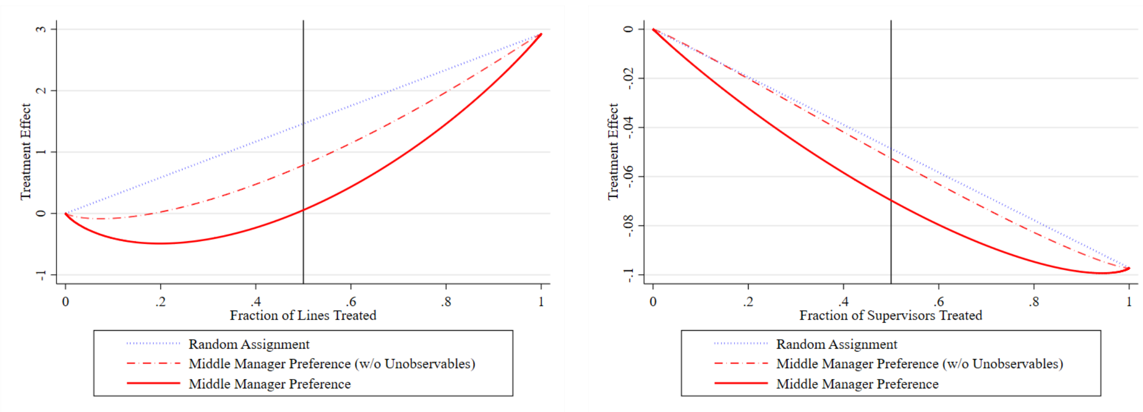
Note: *** p<0.01, ** p<0.05, * p<0.10. The first column shows the results from the first stage probit model for line level middle manager selection. Second column presents results from the heterogeneous treatment effect regression. For column 2, all shown coefficients are for the triple interaction of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

Table C.9: Selection and Retention Effect Heterogeneity

	First Stage **Middle Manager 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 to Improve (Scored by Middle Manager)	0.273***	-0.155
	(0.065)	(0.153)
Months as Supervisor (Answered by Middle Manager)	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 Middle Manager)	-0.159**	0.390
	(0.065)	(0.246)
Industrial Engineering Skills (Scored Middle Manager)	-0.151**	-0.153
	(0.067)	(0.167)
Management Skills (Scored by Middle Manager)	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 middle manager 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.

Figure C.4: Random Allocation vs. Middle Manager Allocation

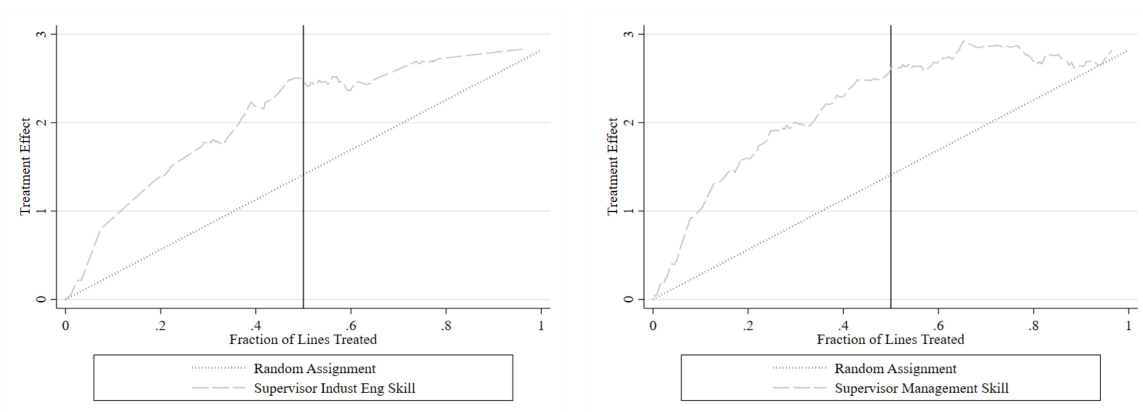


(a) Line Productivity

(b) Supervisor Retention

Note: Treatment effects with random allocation and middle manager allocation. Left panel shows results for line productivity. Right panel shows results for supervisor retention. Horizontal line signifies the point where 50% of lines (for productions) or supervisors (for retention) are treated.

Figure C.5: Training Allocation Based on Middle Manager Skill Scores



(a) Allocate Based on Industrial Engineering Skill Score

(b) Allocate Based on Management Skill Score

Note: Productivity effects with allocation rules based on middle manager assessment of supervisors. Left panel allocates based on the industrial engineering score. Right panel allocates based on the management skill score. Training is allocated first to lines with lowest average supervisor skill scores. Horizontal line signifies the point where 50% of lines are treated.

C.6 Alternative Explanations

Table C.10: Middle manager recommendation is not well explained by demographic and favoritism related variables

	Middle Manager Recommendation	
	All Supervisors	Productivity Sample
	(1)	(2)
Supervisor Age	0.015	0.036
	(0.063)	(0.106)
Supervisor Age Squared	-0.000	-0.001
	(0.001)	(0.002)
Supervisor 1(Male)	-0.161	0.318
	(0.276)	(0.334)
Supervisor 1(Hindu)	0.666	0.803
	(0.442)	(0.688)
Supervisor Native Language is Kannada	-0.394***	0.102
	(0.144)	(0.229)
Supervisor from Different State	0.173	0.508***
	(0.144)	(0.190)
Supervisor 1(General Caste)	-0.084	-0.176
	(0.107)	(0.142)
Sup and Middle Manager Same Gender	0.137	-0.277
	(0.273)	(0.330)
Sup and Middle Manager Same Age Group	0.111	0.052
	(0.134)	(0.179)
Sup and Middle Manager Same Caste	0.119	0.190
	(0.106)	(0.140)
Sup and Middle Manager Same Religion	-0.368	-0.724
	(0.362)	(0.592)
Sup and Middle Manager Coincident Tenure (Years)	-0.003	0.002
	(0.015)	(0.022)
Supervisor Hired After Middle Manager	-0.005	-0.141
	(0.117)	(0.166)
Sup and Middle Manager Same Cohort	-0.163	0.031
	(0.232)	(0.349)
Constant	2.711**	2.251
	(1.067)	(1.761)
Observations	1051	585
R Sq.	0.016	0.025
F-stat	1.359	1.090

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

Table C.11: Middle Manager Recommendation and Middle Manager Skill Scores

	Middle Manager Recommendation			
	All Supervisors		Productivity Sample	
	(1)	(2)	(3)	(4)
Technical Tailoring Skills	-0.102*	-0.145**	0.009	-0.060
	(0.060)	(0.058)	(0.087)	(0.083)
Industrial Engineering Skills	-0.058	-0.089	-0.021	-0.050
	(0.067)	(0.064)	(0.097)	(0.090)
Management Skills	0.093	-0.046	0.077	-0.110
	(0.060)	(0.063)	(0.086)	(0.087)
Motivation to Improve		0.347***		0.489***
		(0.061)		(0.078)
Constant	3.343***	2.748***	2.812***	1.904***
	(0.213)	(0.236)	(0.292)	(0.317)
Observations	1289	1289	695	695
R Sq.	0.005	0.027	0.002	0.046
F-stat	2.150	10.605	0.413	10.004

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

Table C.12: Baseline Productivity and Middle Manager Assessment of Skills

	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: Robust standard errors in parantheses. *** p<0.01, ** p<0.05, * p<0.10. Coefficients are from a linear regression of baseline productivity on the specific middle manager assessment on skill. Middle manager 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.

C.6.1 Observable Determinants of Middle Manager Recommendation in Random Subsamples of Middle Managers

In this section, we rerun the LASSO analysis described in Section 5.2.2 and Appendix section B.4 on random subsets of middle managers to assess any evidence of distinct and possibly countervailing ranking strategies by middle managers, especially with regards to their skill scores for the supervisors. To do this, we randomly select 50% of middle managers in our data, rerun the analysis with the same large set of possible determinants, and collect the selected variables. We repeat this exercise 1000 times.

With 50% subsamples, the LASSO analysis tends to select more variables and have higher R^2 than the analysis on the full sample. While analysis on the full sample selects 11 variables that, when included in a regression explaining the middle manager recommendation, give an R^2 of around 10%, the subsample analyses on average select 37 variables and produce an R^2 of 32% (both mean and median). However, we still interpret this relatively small R^2 on average as evidence that even in subsamples with potentially more homogeneous ranking strategies among middle managers we struggle to explain even half the variation in middle manager rankings with a very rich set of observables. Indeed the R^2 never reaches 50% in any of the 1000 iterations.

To identify observables that capture dimensions that middle managers might use to recommend supervisors in potentially countervailing ways, we look for observables that, when selected by the LASSO procedure, have both positive and negative coefficients a substantial fraction of the runs. Of course, if an observable has no relationship with the training recommendation, it would have positive and negative coefficients each about half the time due to noise. Therefore, we focus on observables that get selected by the LASSO at least half the time, and, when they are selected, they are significantly (at 5%) associated with the middle manager recommendation at least half the time.[40] The idea is that, if, for example, half the middle managers use an observable to positively recommend supervisors and the other half negatively, certain random samples that favor one group or the other due to chance would provide significant associations. In the case where there is no relationship between the characteristic and recommendation, this variable should not be significantly associated with the recommendation in the subsamples. This leaves us with 16 variables out of 50. All of these variables are either negative or positive over 90% of the runs in which they are selected, providing no indication that there is substantial countervailing heterogeneity or a bimodal distribution of coefficient signs in how middle managers recommend based on these characteristics.

Further, the variables that are most often selected by the LASSO have similar interpretation to the results from the pooled sample. For example, middle manager perceptions of supervisors motivation, supervisor effort (as measured by target effort index), and the supervisors quantity and variety of experience (tenure at line, whether operator before, worked different line before) are consistently selected and often significant but essentially always positive such that to the degree that middle managers rank on this criteria they always reward motivation and tenure positively. Similarly, cognitive ability is often selected and consistently with a negative coefficient, consistent with the pooled analysis. Taken together we see no clear evidence of countervailing ranking strategies for any variables which are often selected and/or significant. Further, we also see that the set of variables which are often selected and significant is generally similar to those from the full

[40]For each run, we run a linear regression of middle manager recommendation on all the selected variables and use the statistical significance of the coefficient from this regression.

sample exercise and still amount to a relatively small explanatory power (though larger than the pooled analysis) even in this analysis across smaller subsamples.

Specifically looking at skill scores given by middle managers, baseline management skills is most consistently selected by the LASSO analysis (though it ranks 20 out of 50 variables). This is followed by baseline industrial engineering skills (at rank 39). When selected, both these variables have consistently negative coefficients (around 90% of the time), suggesting that middle managers tend to allocate training to supervisors lacking these skills at baseline. That these variables are not selected in the pooled sample and are selected relatively less often in this subsample analysis likely reflects their low explanatory power as opposed to strong opposing ranking choices by middle managers. Finally, supervisor technical tailoring skill score is among the least selected variables, and, even when selected, is significant only a quarter of the time. As opposed to other skills, it has a positive coefficient 75 % of the times it is significant.

C.7 Spillover Results

Table C.13: Spillovers for Line Productivity Effects

	Efficiency (Produced/Target)		
	(1)	(2)	(3)
During Training X Treatment	4.476*	4.559*	3.749
	(2.555)	(2.578)	(2.543)
After Training X Treatment	0.264	0.326	-0.568
	(3.309)	(3.328)	(3.333)
During Training X Second Tercile of Saturation	1.729	1.818	1.014
	(1.706)	(1.723)	(1.715)
During Training X Third Tercile of Saturation	0.108	0.073	-0.002
	(2.304)	(2.316)	(2.268)
After Training X Second Tercile of Saturation	4.402**	4.374**	3.660*
	(1.859)	(1.868)	(1.883)
After Training X Third Tercile of Saturation	3.487	3.365	3.579
	(2.759)	(2.767)	(2.639)
During Training X Second Tercile of Saturation X Treatment	-3.254	-3.362	-2.749
	(3.340)	(3.365)	(3.304)
During Training X Third Tercile of Saturation X Treatment	0.894	0.886	1.935
	(3.442)	(3.465)	(3.393)
After Training X Second Tercile of Saturation X Treatment	-1.618	-1.643	-0.970
	(4.004)	(4.033)	(4.038)
After Training X Third Tercile of Saturation X Treatment	2.276	2.218	3.399
	(4.515)	(4.537)	(4.421)
Observations	197639	197638	197638
Number of Lines	422	422	422
Number of Floors	54	54	54
Control Mean of Dependentent Variable	54.447	54.447	54.447
Line FE	X	X	X
Month FE	X		
Day FE		X	X
Relative Date FE			X

Note: Standard errors are clustered at line level (*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$). Saturation is defined as the fraction of supervisors treated on the production floor. 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.

Table C.14: Treatment Effect Heterogeneity by Middle Manager Recommendation Controlling for Spillovers

	Efficiency (Produced/Target)		
	(1)	(2)	(3)
During Training X Treatment	5.616***	5.628***	5.480***
	(2.119)	(2.131)	(2.091)
After Training X Treatment	3.787	3.730	3.628
	(2.615)	(2.625)	(2.593)
During Training X High Rec X Treatment	-4.096*	-4.024	-4.342*
	(2.481)	(2.501)	(2.450)
After Training X High Rec X Treatment	-6.140*	-5.967*	-6.447**
	(3.343)	(3.363)	(3.274)
Observations	168335	168334	168334
Number of Lines	356	356	356
Number of Floors	51	51	51
Control Mean of Dependentent Variable	54.447	54.447	54.447
Line FE	X	X	X
Month FE	X		
Day FE		X	X
Relative Date FE			X

Note: Standard errors are clustered at line level (*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$). All specifications include controls for the interaction of floor saturation tercile and event period (during/after training). "High Rec" is an indicator for whether a line has average supervisor recommendation above the median. The analysis covers six months prior to training start month and the six months post the training end month for each factory. 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.

C.8 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 C.15. Heterogeneity of the treatment effect by middle manager recommendation is reported in Table C.16.

Table C.15: Treatment Effects on Salary Progression

	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.

Table C.16: Heterogeneity of Treatment Effects on Salary Progression

	Salary Change (1)	(2)
Treated	0.003	0.005
	(0.007)	(0.006)
Treated X High Rec	0.001	-0.002
	(0.010)	(0.009)
Num. Months Before Quitting		0.014***
		(0.001)
Observations	884	884
Control Mean of Dependentent Variable	.124	.124
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.

C.9 Incentive Bonuses

Appendix Table C.17 presents the results on incentive bonus payments.

C.10 Supervisor Attendance

We assess the day-supervisor level retention effects of training using the following difference-in-differences specification:

Table C.17: 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 level. $1[Any]$ 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.

$$1[Attended]_{itr} = \alpha + \beta_1 T_i \times 1[During_t] + \beta_2 T_i \times 1[Post_t] + \delta_l + \mu_t + \gamma_r + \epsilon_{itr} \tag{10}$$

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 C.18. We do not observe any evidence of treatment effects on supervisor retention.

Table C.18: Treatment Effects on Supervisor Attendance

	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.

C.11 Worker Retention and Attendance

Appendix Figure C.6 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 evidence of any treatment effects.

Figure C.6: Worker Retention by Supervisor Treatment

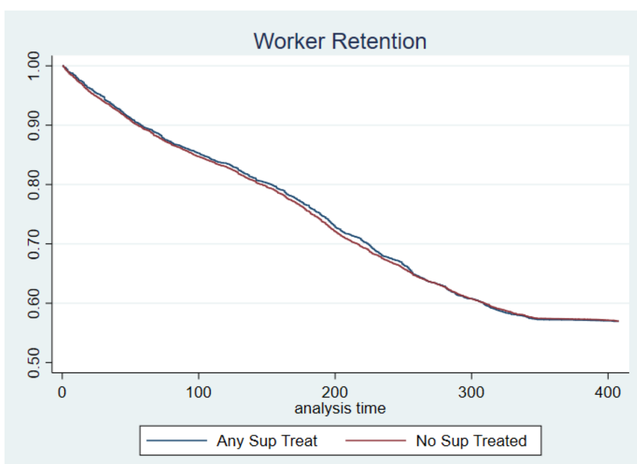


Table C.19: Treatment Effects on Worker Attendance

	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.