Improving Management through Worker Evaluations: Evidence from Auto Manufacturing *

Jing Cai

University of Maryland, NBER, and BREAD

Shing-Yi Wang

Wharton, NBER, and BREAD

August 2020

Abstract

Using a randomized experiment with an automobile manufacturing firm in China, we measure the effects of letting workers evaluate their managers on worker and firm outcomes. In the treatment teams, workers evaluate their supervisors monthly. We find that providing feedback leads to significant reductions in worker turnover and increases in team-level productivity. In addition, workers report higher levels of happiness and positive mood. The evidence suggests that these results are driven by changes in the behavior of managers and an overall better relationship between managers and workers.

*Emails: cai516@umd.edu, was@wharton.upenn.edu. We thank Qingqing Chen, Sijia Chen, Ying Chen, Tianqi Gan, Sunny Lee, Yue Li, Sai Luo, Cung Truong and Stephanie Wu for excellent research assistance; Santosh Anagol, Nick Bloom, David Lam, Dean Yang, and various seminar participants for helpful comments. We thank the University of Michigan, the Wharton Dean's Research Fund and the Wharton Katz Fund for Research on Leadership and Emotional Intelligence for funding. This study was registered in the American Economic Association Registry for randomized control trials under trial number AEARCTR-0002286. All errors are our own.

1 Introduction

There are large differences in firm productivity between developing and developed countries (Bloom and Reenen (2007)), and prior research has demonstrated that management practices may be key to explaining the observed productivity differentials (Bloom et al. (2013)). This may reflect that different firms approach the fundamental principal-agent problem between managers and firms in different ways. Theory suggests that financial incentives for managers may be key to solving the moral hazard problem of firms and managers (e.g. Grossman and Hart (1983), Holmstrom and Milgrom (1987)). However, randomized controlled studies have shown that interventions that do not alter the financial incentives of managers, including providing managers with training, feedback, or consulting, may be effective in improving managerial performance (e.g. Bruhn, Karlan, and Schoar (2018), Schoar (2013), Kelly et al. (2014), Bloom et al. (2013)).

Feedback is one potential cost-effective method of addressing these issues. A manufacturing firm usually has team-level productivity metrics such as whether the team is meeting its production targets and the quality of its output. However, workers may have information about the quality of managerial inputs, which is hard for the firm to directly observe. For example, a team may perform well because the workers happen to be high ability workers, despite having a bad manager. As a result, workers' feedback on managers can be important in measuring and improving management performance.

We investigate the impact of eliciting workers' feedback about their managers on worker, manager and firm outcomes using a field experiment in China. The experiment involved 76 production teams with 1,250 workers in a large automobile-manufacturing firm. In the randomized experiment, we allow production line workers to evaluate their managers in 50% of the teams. Workers are asked to rate their managers in five areas, including production organization, fairness, openness to suggestions, adaptability, and empathy. In treated teams, the worker evaluation component accounts for 20% of the monthly evaluation score of managers. The evaluations occurred every month for 8 consecutive months, and results are posted publicly each month along with the team's standard performance outcomes. In control teams, the company continues to evaluate managers based on hard performance metrics including meeting production targets.

Company policies where workers evaluate their managers are common. In a sample of 280 American firms, Antonioni (1996) finds about 25% of American firms in their sample do 360 degree evaluations.[1] There is an existing body of literature in psychology on upward evaluations using randomized experiments (Atwater et al. (2000), Seifert, Yukl, and McDonald (2003), Heslin and Latham (2004), Seifert and Yukl (2010)). However, the prior psychology literature does not have data on hard performance metrics but only survey reports about changes in leadership.

In our project, we study a comprehensive set of worker and manager outcomes from both administrative and survey data. We received a rich set of administrative data from the firm, including individual-level performance measures, team-level key performance indicators (KPI) evaluated monthly by the firm, and worker turnover. We also collected two waves of our own survey data where we did a baseline survey in the summer of 2016 before the intervention and a follow-up survey in the summer of 2017 shortly after the end of the intervention. In the survey we collected information on demographic characteristics, earnings, work satisfaction, interactions with managers, happiness, and social networks within the firm.

While the worker evaluations of managers occurred over 8 months, direct incentives to the managers were linked to the evaluation scores only for the first 6 months. In treatment teams, the firm gave the worker evaluations a weight of 20% in the calculation of the managers' monthly performance score and reduced the weight on hard metrics from 100 to 80%. In the control team, the firm continued to use their existing scheme based on hard metrics to determine the managers' monthly score. The performance score is important for managers because it is the key factor determining their monthly bonuses, annual pay raises, and the likelihood of future promotion. In the final two months, the worker evaluations continued to happen without entering the final performance score. This helps us to examine the importance of monetary and career incentives versus the feedback alone in changing the behavior of managers.[2] In this setting, we find evidence that the financial incentives were not necessary to see persistent changes in outcomes.

[1] The 360 degree evaluation includes upward evaluations where workers evaluate their managers, horizontal evaluations of peers, as well as the more common downward evaluations where managers give feedback on their workers.

[2] This contributes to a number of existing studies that have examined the impact of the financial incentives to managers (Coughlan and Schmidt (1985), Jensen and Murphy (1990), Lazear (2018), Murphy (1999), Mehran (1995), Carpenter and Sanders (2002), Bandiera, Barankay, and Rasul (2007), Bandiera, Barankay, and Rasul (2009)).

We estimate the impact of the intervention on the probability that workers and managers leave the firm between the baseline and follow-up surveys. Workers that participated in the manager evaluation are 6.2 percentage points less likely to quit the job relative to the control group, corresponding to a treatment effect on worker turnover of over 50%. This is important in the broader context of firms in developing countries where turnover rates are very high and employee retention is a major challenge to firms (World Bank (2013)). For example, the average employee turnover rate in China was 20.8% in 2016 with the rate being as high as 40% in some industries (Callegarin (2017)). Not surprisingly, research suggests that rapid worker turnover significantly disrupts productivity (Moon et al. (2019), Moon et al. (2020)).

We find that there is no significant change in individual productivity. However, team-level KPI does increase significantly by 2.3%. How can individual productivity not change while total team-level productivity increases? This is driven by the fact that individual-level productivity does not change for workers who stay, but treated teams have lower worker turnover than control teams. High turnover is costly to team-level KPI because new workers normally perform poorly in the beginning and need substantial training.

We also explore the impact on workers' and managers' well-being, measured by happiness, work satisfaction, mood, and health. We find that workers in teams that were given the opportunity to evaluate their managers report higher levels of happiness, and there is a significant increase in the measure of positive mood. This is consistent with the idea that the intervention led workers to be happier and less likely to quit.

Overall, our results suggest that the intervention led to improved outcomes for workers, managers and the firm with no observable downside. Workers reported higher job satisfaction and quit less, managers performed better (and increased their earnings), and the team-level performance metrics that the firm cared most about increased. After seeing the results of our research, the firm scaled up the feedback system to all of their plants, covering thousands of production teams and almost 20,000 workers.

We next test three main mechanisms that could drive the results. First, managers may change their behavior at work and these changes lead to changes in workers' outcomes. We provide direct

evidence supporting this mechanism: in treatment teams, workers report that the team manager encourages them more, criticizes them less, and is less likely to lack empathy towards the feelings of the workers. Moreover, treated managers have more social interactions with workers outside of work.[3]

A second mechanism is the idea that giving workers voice (with no change in managers' behavior) makes workers happier and reduces turnover. For example, Adhvaryu, Molina, and Nyshadham (2020) show that giving randomly selected workers the opportunity to participate in a feedback survey after a disappointing wage hike reduced turnover.[4] To test this channel in our context, we use variation in the salience of the intervention among treated teams.[5] We find that only salience of the treatment for managers has a positive and significant impact on worker outcomes and team KPI, while workers' awareness of the consequences of the evaluation does not matter, thus ruling out the voice channel as the primary mechanism.

Third, the results could be driven by changes in team composition (beyond worker turnover) in treatment relative to control teams. However, we show that there are no significant differences in manager turnover, the probability of workers shifting between treatment and control teams, and in the outcomes of interest for workers who switched teams or left the company. Taken together, our results suggest that the positive impacts of adding workers' feedback in manager evaluation are mainly driven by changes in the relationship between managers and workers.

We then test the persistence of the effect by examining new entrants to the firm who began after our intervention was terminated. Our results show that even for new hires who entered after the intervention ended and never participated in the manager evaluations, those in treated teams report that their managers gave more encouragement and provided more empathy than new entrants to control teams. This suggests that the changes in managers' interactions with workers persisted beyond the period in which subordinates evaluated them.

[3]One possible reason that managers' kindness towards workers leads to better outcomes is driven by gift exchange between workers and managers (e.g. Akerlof (1982)). For example, Bellemare and Shearer (2009) documents that giving workers a monetary gift increases productivity significantly. However, we rule out gift exchange in our setting because individual performance does not change after the intervention.

[4]The participants are told that the firm will not receive the feedback provided, so they isolate providing an avenue for worker voice absent any possible changes in firm policy or managers' behavior.

[5]Salience is measured by their recall of participating in the feedback a few months after it ended.

Finally, we discuss some issues with identification and interpretation. The main concern is that two types of spillovers might bias our estimations. First, if workers in control teams heard about the opportunity to evaluate managers in treated teams, they might react negatively to not being given this opportunity, leading to overestimates in the impacts of the intervention. We rule out this concern by showing that there is little discussion about the intervention across teams, and more importantly, control workers' outcomes are not significantly affected by the share of their social networks who are treated. Second, there can be direct productivity spillovers across teams along the production line. However, we find that having a treated team immediately upstream in the production line does not have a significant impact on the productivity of downstream control teams.

Our work builds on and contributes to several existing literatures. First, this paper sheds light on the growing literature documenting the importance of improving management practices (Bebchuk, Cohen, and Hirst (2017), Bloom and Reenen (2007), Bloom et al. (2013), McKenzie and Woodruff (2016), Adhvaryu, Kala, and Nyshadham (2019)). A number of studies have examined specific strategies to improve management, such as offering training programs to improve the soft skills of managers (Schoar (2013), Kelly et al. (2014)), performance monitoring (Frederiksen, Kahn, and Lange (2019), Gosnell, List, and Metcalfe (2019)), and giving managers more decision rights (Milgrom (1988), Bartling, Fehr, and Herz (2014)). We add to this literature by showing that letting workers provide feedback on managers can be a cost-effective way to improve management practices.

Second, our results provide one potential solution to the observed high worker turnover rates in developing countries.[6] Managing high turnover is critical because it is costly to hire and train new workers (Dube, Freeman, and Reich (2010), Blatter et al. (2015), Muehlemann and Pfeifer (2016)), and rapid turnover can lead to negative impacts on productivity and product reliability (Moon et al. (2019), Moon et al. (2020)). There is almost no experimental evidence on effective interventions that firms use to solve the problem. Our contribution to this literature is to provide

[6]The phenomenon of high worker turnover is not unique in developing countries. For example, in the early 20th century worker turnover was as high as 117.2% in the U.S. but it reduced to around 47% by the 1970s (Jacoby (1983)). Owen (1995) attributes the decline to improvements in employment relationships and benefits.

new evidence of an effective method of increasing worker satisfaction and retention.

Furthermore, our results contribute to the literature on empowering workers and changing the lines of communication between workers, managers and the firm. In addition to experiments on upwards evaluations (Atwater et al. (2000), Seifert, Yukl, and McDonald (2003), Heslin and Latham (2004), Seifert and Yukl (2010)) and voice (Adhvaryu, Molina, and Nyshadham (2020)), there is non-experimental evidence on other avenues for communication including worker cooperatives (Craig and Pencavel (1995)), unions (Cooke (1994), Freeman (1980), Batt, Colvin, and Keefe (2002)), and counterproductive work behavior (Coviello, Deserranno, and Persico (2020)). Our findings are in line with the prior research on upward evaluations in finding changes in management behavior following the feedback. However, we add to the existing literature in our access to administrative data from the firm, which allows us to be the first to consider the potential trade-off between worker satisfaction and performance. We are also the first to examine the potential mechanisms through which worker feedback about managers matters.

Lastly, our analysis provides a novel test of the theoretical idea that firms should incentivize managers based on productivity and not on any intermediate inputs (Holmstrom and Milgrom (1987)). The intuition of these models is that there is no need to incentivize intermediate inputs because managers should choose those to optimize productivity and firm profits. However, our results reject the idea that aligning financial incentives alone is the best policy; instead, profits can be increased through improved worker happiness and retention if managers are given direct feedback to be nicer to their workers.

The paper proceeds as follows. In section 2, we explain the background for the study. In sections 3 and 4, we present the experimental design and explain a simple framework. In section 5, we discuss the data collected. In section 6, we report the main impacts of worker evaluation and results on mechanisms. Section 7 discusses the persistence of the intervention impact and some robustness checks. Section 8 concludes.

2 Background

We carry out our experiment with an automobile manufacturing firm in China. During our study period, the automobile industry was growing steadily, with an average sales growth rate of more than 5% per year and export growth of around 30%. Our project partner is a large company, mainly producing passenger vehicles, heavy duty vehicles, and new energy cars. Similar to many other manufacturers in China, the company faces problems of hiring and retaining workers, with an average annual worker turnover rate of more than 20%.

We partner with one major plant of the company, which produces passenger cars. The plant has four factories, including a stamping factory, a welding factory, a coating factory, and a final assembly factory. The production process is fairly capital intensive, but the mode of production used here is substantially less capital intensive and more labor intensive than for a similar car produced in the United States. Production line workers are organized into production teams within each plant. Every team has a team manager. Team managers are usually promoted from the pool of workers within the team, where workers need to pass a written exam to be in the eligible pool of workers for consideration for promotion.

Workers and managers are paid monthly. The monthly earnings of workers and managers are composed of two parts: the flat salary and a bonus. The flat salary accounts for around 90% of workers' monthly earnings and about 80% of team managers' monthly earnings. The monthly bonus component of their compensation depends on performance metrics at the company level and also at the team level for managers and at the individual level for workers. Workers receive public feedback on their individual performance on a daily basis at team meetings. Managers also receive public feedback on their team-level performance metrics on a monthly basis; the information about the worker and team performance is posted at the team's bulletin board regularly.[7] Workers and managers also receive semi-annual and annual bonuses and pay raises, both of which depend on their performance metrics as well as how the firm is doing in aggregate.

[7] In addition to teams' performances, the monthly ranking of team performance scores and the "star worker" of the month is usually posted at each factory's announcement board located at the entrance of the factory.

3 Experiment

Our experiment was implemented in three of the four factories at the plant.[8] The experimental factories include 76 production teams and 1251 workers. The main intervention is to randomize the introduction of workers' monthly evaluation of the performance of their manager at the team level.[9] Specifically, in half of our sample teams (the treated teams), workers evaluated their managers on five components of management (organization, adaptability, fairness, openness to suggestions and empathy) on a scale from 1 to 5 with 5 being extremely effective and 1 being not effective at all.[10] In treatment teams, the worker evaluation score accounted for 20% of the managers' monthly performance score while the hard metrics accounted for 80%. Aggregate scores from the worker evaluations were posted every month on the team's bulletin board along with the other performance metrics posted every month. The collection of the feedback from workers lasted for 8 months, from September 2016 to April 2017. However, the worker evaluation score did not enter into managers' total performance score after February 2017; thus for two months prior to the end of the feedback collection, the feedback continued without factoring directly into the managers' monthly performance score.

Managers may care about the worker evaluations for multiple reasons. First, it affects their earnings directly for 6 months; given that the monthly bonus is approximately 20% of the managers' total compensation in a month, this means the intervention alters up to 4% of their monthly salary.[11] Moreover, the performance score also influences the managers' annual bonuses and pay raises. Finally, the performance score is a key factor in determining promotions for production line managers.

The procedures of our intervention were as follows. Before beginning the intervention, the plant manager held a meeting with workers and managers of each treated team. During the meeting, they

[8] The welding factory did not participate.

[9] The randomization is stratified by team size and the average team performance score during 2015, the year before our experiment.

[10] See Appendix Table A.1 for the wording of the questions.

[11] The firm does not inform workers or managers about the specific formula determining their bonuses, but they understand what the components are and can see month-to-month how variation in their own performance and firm performance map into changes in salaries. In the data between September 2016 and April 2017, there is a positive correlation of 0.157 between managers' total evaluation scores and the team-level hard metrics.

first introduced our research assistants, and notified that the research team would visit the factory every month to collect worker evaluations of team managers. Next, the plant manager distributed the evaluation questionnaire to team managers and workers and explained each question. Finally, they informed individuals in treated teams that the evaluation was completely anonymous and that the factory would take the evaluation into account when calculating the team managers' monthly performance scores until the Chinese New Year. Specifically, the workers did not write their names on the evaluation cards, which were passed out and collected at the end of a team meeting by an individual on our research team. The manager was not present during the collection of the feedback. The data was then digitized and sent to each plant's human resources team.[12] Thus, the workers could be assured of the anonymity of their responses. The whole process took less than 10 minutes per month per team.

The firm had never asked workers to provide feedback on their managers prior to our intervention. While the policy was new, our understanding is that it did not feel out of place for several reasons. First, team-level managers were asked to provide feedback about their own production line managers on an annual basis, so managers were familiar with the concept of providing upward evaluations. Second, the firm collects and uses a lot of data at regular intervals, including some at a daily frequency.

In control teams, the pre-existing status quo was preserved. Workers did not evaluate their team managers and team managers' monthly performance scores were solely determined on the basis of the standard team-level performance metrics.

4 Framework

We provide a basic framework for understanding the incentives of workers, managers and the firm. We assume that the firm wants to maximize profits and tries to provide incentives for managers to behave in ways that help maximize firm profits.

The managers want to maximize their own earnings. The firm attempts to align incentives

[12] Nobody in the firm saw the paper copies of the individual evaluation forms filled out by workers.

by making the managers' earnings depend in part on the team's performance, $k(Q(s_m), E(s_m))$.[13] Managers (denoted by m) make choices about managerial effort, h_m and s_m, where we assume managers can spend their limited effort either pushing the workers to meet production targets (h) or providing soft management skills (s), such as expressing empathy for the life circumstances that prevent workers from getting to their shift on-time. Q and E are summations of all of the workers on the team's decisions regarding quitting (q_i) and effort (e_i), respectively.

Managers' earnings are denoted by y_m which can be broken down into a fixed component, $\overline{y_m}$, and a component that varies based on team performance. So the manager's problem is:

$$\max_{s_m} y_m \tag{1}$$

subject to

$$h_m + s_m \leq 1 \tag{2}$$

where

$$y_m = \overline{y_m} + k(Q(s_m), E(s_m)). \tag{3}$$

Workers make two decisions: whether to stay at the firm or leave, and how much effort to exert on the job. A worker i quits if

$$V(y_w^i(e_i(s_m)), a_w^i(e_i, s_m)) < v^r \tag{4}$$

where v^r is the value of their outside option. The worker's utility at the firm depends on earnings, y_w, and other amenities, a_w, which we might summarize as non-pecuniary job satisfaction. The cost of effort is reflected in the idea that job satisfaction is decreasing in effort, $\partial a/\partial e < 0$. The worker also makes a choice regarding how much effort, e, to exert on the job:

[13] This will correspond to the key performance indicators (KPI) in the data.

$$\max_{e_i} V(y_w^i(e_i(s_m)), a_w^i(e_i, s_m)). \tag{5}$$

Workers' earnings include both a fixed component, $\overline{y_w^i}$, and a bonus that depends on their effort, $b(e_i(s_m))$:

$$y_w^i = \overline{y_w^i} + b(e_i(s_m)). \tag{6}$$

We have workers' job satisfaction increasing in the managers' choice of s; in other words, $\partial a/\partial s > 0$. Note also that increasing effort leads to better performance and a higher bonus for the worker so $\partial b/\partial e > 0$. Thus, for workers, there is a trade-off where higher effort decreases their overall satisfaction at work but increases their earnings.

Let us consider how worker effort depends on the manager's decision, s_m. One possibility is that effort depends on h_m such that $\partial e/\partial h > 0$ (so $\partial e/\partial s < 0$). Under this assumption, the more managers crack the whip on their workers, the more effort is produced, but they risk workers being unhappy and quitting.[14] However, an alternative is a system of gift exchange between managers and workers such that when managers are kinder to workers, workers exert more effort.[15] Under this system of gift exchange, we would have $\partial e/\partial s > 0$.

The propensity of the worker to quit is given by

$$q_i = V(y_w^i(e_i(s_m)), a_w^i(e_i, s_m)) - v^r \tag{7}$$

where $\partial q/\partial a < 0$ and $\partial q/\partial y_w < 0$. A worker's quit rate falls with both greater job satisfaction and higher earnings.

Team-level performance depends on the levels of performance of the individual workers on the

[14] This trade-off is similar to Akerlof and Kranton (2005, 2008) where supervisors trade off between loose and strict supervision where loose supervision engenders workers' identity as part of the team.

[15] Note this set up deviates from the canonical model of gift exchange with workers where efficiency wages induce greater effort (Akerlof 1982).

team as well as the workers' turnover rates in a way defined by the function $f()$:

$$k(Q(s_m), E(s_m)) = \sum_{i=1}^{T} f(q_i(s_m), e_i(s_m)) \tag{8}$$

where T refers to the total number of workers on a team. Turnover reduces team-level performance ($\partial k/\partial Q < 0$) because team members must help with training new workers, and new workers start off with very low levels of productivity while being trained.[16] Worker effort also affects team performance, so $\partial k/\partial e > 0$.

Under the scenario where managers need to crack the whip to induce effort (i.e. $\partial e/\partial h > 0$), managers want workers to be happy because it matters for workers' turnover rates (which affect team performance). Both $\partial k/\partial s > 0$ and $\partial k/\partial h > 0$. There is a trade-off that managers face between workers' job satisfaction and workers' performance. The manager needs to decide whether to push the workers hard and get more output, but this may lead to workers being unhappy and being more likely to quit. Alternatively, under gift exchange (i.e. $\partial e/\partial s > 0$), managers choose to be nice to workers both to reduce turnover and to increase workers' performance.[17] Regardless of the relationship between worker effort and s, rational managers should choose s to maximize k.

There are several ways that the intervention may shift the incentives of managers. One possibility is that the intervention works through placing additional financial incentives on soft skills of managers through the function $j()$. We have:

$$y_m = \overline{y_m} + (1-\alpha)k(Q(s_m), E(s_m)) + \alpha j(s_m) \tag{9}$$

where before our intervention $\alpha = 0$ and afterwards $\alpha = 0.2$ and $\partial j/\partial s_m > \partial k/\partial s_m$. The key prediction here is that intervention should lead managers to shift their input decisions such that s_m increases. While a simple way to look at this is through the financial incentives given by equation 9, it is also possible that the intervention works not through y_m but through managers getting positive utility from doing well in the feedback and negative utility from doing poorly in the feedback. This

[16] See Appendix Table A.1.

[17] Note the model of gift exchange suggests that managers should simply maximize s_m with no trade-off between the s_m and h_m. In this case, we would need to replace equation 2 with a positive cost associated with s_m.

would produce the same key prediction that the intervention should lead to an increase in s and decrease in h.

The model also suggests that any observed increases in k must be driven by declines in turnover, increases in worker productivity, or both. One key implication of the framework is that if $\partial e/\partial h > 0$ and managers were optimally choosing managerial inputs prior to the intervention, we should not observe an increase in k after the intervention. As managers shift towards increasing s, workers' job satisfaction will increase and turnover will fall, but this should come at a cost to k. Under gift exchange ($\partial e/\partial s > 0$), the intervention can lead to an increase in k driven by increases in e. Thus, if we see an increase in k combined with an increase in worker productivity (reflecting an increase in e), this would provide support for the idea of a gift exchange between workers and managers where kindness by managers induces more worker effort.

If we see increases in managers' soft skills (s), worker job satisfaction (a), and team performance (k) combined with declines in turnover (Q), this suggests that we can see improvements in all outcomes at no cost and the original strategy of the firm to focus incentives on hard performance metrics may not be optimal.

5 Data

Our analysis combines data from two main sources: administrative data from the firm spanning the months May 2016 to July 2017, and survey data that we collected at baseline from June to August 2016 and for the follow-up in July and August 2017.

5.1 Administrative Data

There are two main performance data sets that we receive from the firm. First, we have monthly data on performance measures of individual production line workers. This data set provides metrics for performance at the monthly level, broken down into various components, including production, safety, quality, attitude, attendance and routine.[18] Each worker starts with a total performance

[18]Note that these aggregations can be measured differently across teams and factories. Routine is a direct translation of this category, but it is a residual that encompasses anything that does not fall into the other categories.

score of 100 and points are added or subtracted by the manager if they do well or poorly that month in each of the 6 components. Thus, the values associated with each component are positive or negative depending on the worker's performance in that component. This data is collected daily where feedback is also given to the worker daily by the manager. The data that we receive is aggregated to the monthly level by the firm (and this monthly aggregate is what is used in determining the workers' monthly bonus).

The second administrative data set that we use is monthly data on team-level key performance indicators (KPI) which range from 0 to 100. Team-level KPI is also how managers' performances are assessed, and are directly linked to their bonuses. Total KPI is broken down into five main categories: production, management, equipment, quality, and safety. The production category is about meeting production targets (i.e. the completion rate) but also includes the accuracy and efficiency of information delivery, the management of inventory and the frequency of line stoppages. The management category includes easy-to-observe aspects such as whether the manager fills out the worker attendance records correctly and submits them on time, and whether the workers get the required hours of training. Safety takes into account the number of accidents and violations of safety regulations. Quality reflects the number of products that do not pass the quality test and the number of problems found in audits.

These categories are combined (with weights that vary across factories and across teams) into a total team-level performance score.[19] Appendix Table A.5 shows the average weights placed on these categories across teams, broken down by treatment and control teams. There are no significant differences in these weights between treatment and control teams. The most weight on average is placed on production, followed by quality, management, safety and equipment. We do not have data on profits (which are at the firm level) but it is reasonable to assume that the weights are based on how the firm thinks these metrics map into profits for each team.

Table 1 presents the summary statistics for the treatment group in column 1 and the control group in column 2. The statistics show the average of all pre-intervention months of data (May, June, July and August of 2016).[20] Column 3 shows the total number of observations and the last

[19]For example, the equipment category does not exist at all in one of the factories in our analysis.

[20]For cases where the data are not present for all four months, we take the average over the existing months of

column shows the p-value on the treatment dummy in a regression testing the statistical difference between the two groups. Panel A shows the statistics for the individual performance measures and Panel B the team-level KPI and its components. Across the 13 different measures of performance at the individual and the team level, only one is statistically different between the treatment and control groups; worker-level performance in terms of quality is significantly better in the treatment group than in the control group.[21]

5.2 Survey Data

We collected baseline data prior to the intervention (from June to August 2016) in addition to a round of follow-up data one year after the baseline data collection (in July and August 2017). Both workers and managers were surveyed. While the administrative data are sufficient to examine the impact of the intervention on individual and team-level performance metrics collected by the firm, our survey data are important for providing insight into the mechanisms of the changes.

Table 2 shows summary statistics for variables collected in the survey data prior to the intervention. The first set of variables show some demographic characteristics of the workers in our sample. The workers' average age is early 30's. The vast majority of these production line workers are male (approximately 90%). About half of the workers have a high school degree or more. The average tenure within the firm is over 4 years. About two-thirds of the workers are married.

We ask several questions about individual well-being. First, we ask a general well-being question that asks them to rate their life imagining a ladder with steps number 0 to 10. The workers rate their lives around 7 on this scale. This is the only summary statistic along which the treatment group and the control group are statistically different. The difference is 0.3 on a 10 point scale (or 15% of a standard deviation). We also ask them to self-report their general health on a five-point scale. The table presents an indicator for whether they answer with the top two values. Over 60% of workers report being in good or very good health.

data. For example, for a worker who entered the firm in June, we will not have data for May 2016 and calculate that worker's pre-intervention performance measures using the 3 months after the worker began.

[21] There are some large outliers in the individual performance data. Winsorization of the top and bottom 1% does not affect the comparison of means (Appendix Table A.4) or the subsequent results using individual performance (Appendix Table A.9).

While the life rating and the health question capture a longer-run sense of well-being, we also ask questions about their mood in the past 30 days. This is a widely-used set of questions called the Positive and Negative Affect Schedule (PANAS) developed by Watson, Clark and Tellegen (1988). This entails asking 20 questions, detailed in Appendix Table A.2. We follow the standard formula for converting these 20 questions into a summary measure of positive mood and negative mood.

We also ask a standard set of questions on work satisfaction. These questions are asked on a five-point scale where some questions are positive and some are negative.[22] In particular, there are four questions that refer specifically to their manager, including whether their manager encourages them, is unfair to them, criticizes them, or lacks empathy. Other questions on work satisfaction go beyond their relationship with their manager; for example, one question asks about opportunities for promotion that reflects firm-level policies and opportunities. Some questions, such as the workload, could reflect behavior by their direct manager or other higher-level managers. Table 2 shows the averages for the work satisfaction index while Appendix Table A.6 shows the summary statistics for each component of the index.

The average monthly earnings is around 3700 RMB or a little over USD$500. This survey question asks about earnings in the past two months. The average monthly bonus in the baseline survey is under USD$200.[23] Finally, we ask a question in the survey about the number of times that workers socialized with their managers outside of work in the past month. The average number of times is 0.56 in the treatment group and 0.60 in the control group, and these are statistically not different.

6 Main Results

6.1 Estimation Strategy

Our empirical strategy exploits the randomized experiment to compare outcomes in the treatment group with the control group. For outcomes at the individual level, we estimate for individual i in

[22] See Appendix Table A.3 for the wording of the work satisfaction questions.

[23] Given the timing of the survey, this is actually higher than most months because the semi-annual bonus was disbursed in the months referenced by the survey.

team t the following equation:

$$y_{it} = \alpha Treated_t + X'_{it}\beta + \epsilon_{it} \tag{10}$$

where the key regressor of interest is $Treated_t$ which is an indicator for whether workers in team t were given the opportunity to evaluate their managers. We also include a vector of control variables, X_{it}. This includes the randomization strata in all regressions. All regressions also include an indicator variable for a second randomized experiment that we conducted at the same time but is not the focus of this paper.[24] In cases where we have lagged values of the dependent variable, we include it as a control in order to improve the precision of the estimates.[25] Finally, when the dependent variable is from the survey rather than from administrative data, we also include enumerator fixed effects. The standard errors are clustered at the team level.

For the administrative data from the firm, we can estimate equation (10) by pooling the months prior to the reform and the months after the reform. In addition, we can also exploit the full panel of monthly data for the team-level KPI and individual performance and estimate the following equation with all the leads and lags around the start of the intervention where t denotes team and j denotes the month:

$$y_{it} = \sum_{j=-4}^{10} (\gamma_j Month_j \times Treated_t + \theta_j Month_j) + \phi_t + X'_{it}\beta + \epsilon_{it}. \tag{11}$$

The variables $Month_j$ are indicators for the month relative to $j = 0$ when the intervention begins. We omit the month prior to the start of the intervention ($j = -1$). We include team fixed effects, ϕ_t. With the four months prior to the intervention, we can examine whether there were any differences in the trends between the treatment and control groups prior to the intervention. With the 10 months after the intervention, we can examine how long it takes for changes to appear and whether they persist. First, we can see whether the effects persist after workers stop providing monthly

[24] In a parallel experiment, we gave randomly selected workers the right to vote on their compensation policy. Our results are not sensitive to the inclusion of the indicator variable for this randomization.

[25] Thus, our estimation strategy is similar to a differences-in-differences strategy but does not restrict the coefficient on the lagged dependent variable to be one.

feedback on their managers' performance; the feedback ends 8 months after the intervention began ($j = 7$) and we have administrative data for three additional months after the end of the collection of the feedback. Second, we can use the dynamic data to examine whether the effects are primarily driven by the financial incentives linked to the feedback. The financial incentives end 6 months after the intervention began ($j = 5$) so in periods 6 and 7, the financial incentives linked to the feedback are removed while the feedback continues.

6.2 Results on Individual Performance and Team-Level KPI

We begin with regression estimates based on equation 10. We first examine the impact of the intervention on the probability that workers and managers leave the firm between the baseline and follow-up surveys. As shown in Table 3, in treated teams relative to control teams, we see a fall in worker turnover of 6.2 percentage points, and this is significant at the 1% level. We see a 3.6 percentage points decrease in the turnover of managers but this is not significant at the standard levels. The magnitudes of the estimates are very large relative to the mean turnover in the sample of 10.7 percentage points for workers and 1.37 percentage points for managers.[26]

In Table 4, we look at the impact of evaluating managers on individual performance metrics. The dependent variable is the average across all months after the start of the intervention, and the lagged dependent variable is the average across all months prior to the intervention. For the sample of workers who were at the firm in both the baseline and the follow-up period, we see no significant changes in total productivity or in any of the components of individual productivity. Next, in Table 5, we examine the impact of evaluating managers on team-level KPI. There is a significant increase in total KPI of 2.2 (where the average KPI in the control teams in 97). We also see significant increases in two subcategories of team-level KPI: production and management.

How can we reconcile the results where individual productivity doesn't change but total team-level productivity increases? This is driven by the fact that individual-level productivity doesn't change for workers who stay, but treated teams have lower rates of turnover than control teams.

[26]Note that a turnover rate of 10.7 percentage points means that 10.7% of workers surveyed in the baseline were reported by the firm as having left the firm by the time we conducted the follow-up survey. This is an under-estimate of total turnover at the firm given this number doesn't include the many workers who began after our baseline survey and left before our follow-up survey. The firm reports that their annual turnover rate during this period was 20%.

High turnover is costly to team-level KPI because the newest workers perform poorly in the beginning and need substantial training. This is shown in Appendix Figure A.1 where we see relatively low average scores for individual production performance in the first and second quarter after starting at the firm, after which performance improves (and remains relatively flat). Furthermore, the separation rates for new workers is high as many learn that they cannot or do not want to do this kind of work.

We now turn to the estimation of dynamic effects - estimates of the leads and lags around the intervention specified in equation (11). We only have monthly panel data for individual performance and team-level KPI. The regression estimates for worker performance and KPI are shown in Appendix Tables A.7 and A.8, respectively. Given that the basic regression estimates for the performance measures and many of the KPI components are not significant in the regressions comparing post-intervention outcomes, it is not surprising that most of the estimates of the leads and lags are not significant at the standard levels. In Figure 1, we focus on the coefficients of the leads and lags for the two KPI components that were significant (in Table 5): production and management.

In looking at the coefficients on the leads in the interaction, there are no significant differences in the trends in these measures of KPI across treatment and control teams leading up to the randomized experiment. This provides reassurance that pre-existing differences in the trends between these groups are not driving the results. Next we turn to post-intervention periods. We see that the significance in the impact on production KPI shifts around 4 months after the intervention ($j = 3$). The magnitude of the coefficients for management KPI (which measures whether the managers fulfill observable aspects of their jobs such as providing the required amount of training and submitting forms promptly and completely) shifts up immediately after the intervention begins. However, the individual estimates are not significant at the standard levels, though the pooled post-intervention effect is (Table 5). This suggests that the managers begin to change their behavior almost immediately, but that the impact on team-level production may take over a quarter to fully take effect. This is consistent with the idea that effects on team-level production are driven by team turnover, which takes more time to adjust.

The financial incentives for managers to care about worker feedback end after six months ($j = 5$).

Furthermore, the provision of monthly feedback from workers ends after eight months ($j = 7$). However, we see the shift in the magnitude of the effects for production and management KPI persist both after the end of the financial incentives and after the end of the formal system of worker feedback. The estimates for production KPI remain significant at the 10% level after the end of the incentives linked to feedback for every period except one ($j = 9$) which is only significant at the 12.5% level. First, this suggests that the financial incentives were not the primary motivation for any changes in by managers and the financial incentives were not necessary for the changes to persist over time. Second, this suggests that the changes brought about by the intervention persisted even beyond the end of the formal feedback system from workers.

6.3 Results on Earnings and Other Measures of Well-Being

In Table 6, we look at the impact of evaluating managers on workers' and managers' monthly earnings. The intervention does not correspond to significant changes in workers' monthly earnings of bonus. This is not surprising given that there is no impact of the intervention on worker-level productivity, which determines their bonus. For managers, we see a 22.8% increase in their monthly bonus and this estimate is significant at the 5% level. Given that we see team-level KPI increasing in treated teams, and managers' bonus depends heavily on KPI, the impact on earnings is consistent with the other estimates. The bonus is approximately 20% of their total monthly earnings, so a 22.8% increase in bonus maps approximately into the 4% increase in total monthly earnings that we see in column 3, though this estimate is not significant at the standard levels.

The next set of results examines the impact of evaluating managers on workers' and managers' well-being. The results are presented in Table 7 in Panel A for workers and in Panel B for managers. We see no significant changes in any of the measures of well-being for managers. However, we see that workers in teams that were given the opportunity to evaluate their managers report higher levels of general happiness of the life rating scale.[27] This increases by 0.3 on a scale that goes from 0 to 10, and this increase is significant at the 1% level. We also see a significant increase in the measure of a person's positive mood. We do not see significant changes in workers' work

[27] Note these results are for the sample of workers present at the firm in both the baseline and follow-up.

satisfaction index, negative mood or a person's self-reported health status.

Overall, the main results indicate that allowing workers to provide evaluations of their managers led to a fall in worker turnover rates and a corresponding increase in team-level KPI. This is good both for treated managers, who are earning more due to the higher levels of KPI, and the firm, which cares about KPI. While worker productivity and earnings do not improve, there is evidence that the intervention is positive as they report higher levels of happiness and positive mood. Furthermore, the higher levels of retention among treated workers are also suggestive of them being happier in the job.

7 Mechanisms and Persistence

7.1 Overview of Mechanisms

We consider three main mechanisms that may drive the results. First, managers may change their behavior at work and these changes may lead to changes in workers' outcomes. There are different reasons that the intervention can lead managers to change their behavior. They may be responding to the information and feedback that they receive from workers on how to be better managers. Alternatively, managers may change their behavior to simply be nicer to their workers in order to get better evaluations from their workers, but this isn't the result of direct information from the workers. We can test whether there is evidence of observable changes in the managers' behavior.

A second mechanism is that giving workers voices, even without any real changes in what managers do, makes workers happier with the firm and reduces turnover. This is consistent with the work of Adhvaryu, Molina and Nyshadham (2020) on worker voice. To provide some suggestive evidence for this idea, we can use variation in treated teams in whether the intervention was still salient for managers or the workers in the follow-up survey, three months after the intervention had ended. We can look within treated teams at whether managers' or workers' awareness of the policy is important for changing the key outcomes of interest. If the main mechanism is about giving workers voice with no changes in managers' behavior, workers' awareness of the intervention should matter for changes in KPI, satisfaction, and mood, while managers' awareness may not be

necessary.

A final mechanism is that the results are driven primarily by changes in team composition. There are several ways that the teams may be changing differently in treatment and control teams. First, managers may be changing. For example, managers in higher KPI teams may be more likely to be promoted and new managers are closer to and more responsive to workers. Second, workers can shift across teams. The results can be driven by decisions to shift bad or unhappy existing workers from treatment teams to control teams. The higher rates of turnover in control teams may be staffed from other teams, but other managers only give up their worst workers. Finally, the results may be driven by the characteristics of the workers who leave. For example, it may be the case that the best performing and happiest workers are most likely to leave so the estimated effects on worker satisfaction are driven by the different rates of turnover. However, for all of the team composition mechanisms, there would still need to be an explanation for why there were different rates of changes in team composition in treatment versus control teams to begin with. We can look directly at whether there are differences in manager changes and worker shifts between treatment and control teams. We can also examine the characteristics of firm stayers and leavers as well as team stayers and shifters.

7.2 Results on Managers' Behavior

We first break down the work satisfaction index (shown in column 2 of Table 7) to all the individual questions. The results are shown in Table 8 where each row is a separate regression where the variable indicated in the first column is the dependent variable in the regression.[28] There are four questions that directly reference the team manager: encourages, unfairness, criticizes and no empathy. Three of the four components of work satisfaction that directly relate to the team manager change significantly in the treatment teams relative to the control teams. In treatment teams, workers report that the team manager encourages them more, criticizes them less and is less likely to lack empathy towards the feelings of the workers. The only component that references the manager specifically but does not change significantly is whether the manager is unfair towards the

[28]The wording of the question that corresponds to each dependent variable can be found in Appendix Table A.3. The corresponding summary statistics broken down by treatment and control are presented in Appendix Table A.6.

worker.

Most of the other questions about work satisfaction, such as their opportunities for promotion or their satisfaction with employer-provided benefits, do not change significantly as a result of the intervention. The changes are generally concentrated in the way that their manager treats them. The one exception is that there is a significant decrease in workers in treatment teams reporting that the incompetence of their teammates makes them have to work harder.[29] This change is consistent with the fact that there is lower turnover in treatment teams and fewer inexperienced new workers entering the team.

In Table 9, we examine questions that we asked workers about quantifiable interactions with their managers. First, we asked about the total number of times that they socialized with their managers outside of work.[30] The average in the data is socializing outside of work 5.3 times. This doesn't change significantly with the intervention. Another question that we ask in the follow-up survey is whether a worker's manager responds to the worker's WeChat posts, where WeChat is a popular form of social media in China.[31] The dependent variable is an indicator that equals one if the worker reports that their manager responds sometimes or often.[32] On average, 63% of workers report that their manager responds sometimes or often. Workers who have the opportunity to evaluate their managers report a 7% higher response rate on WeChat. This estimate is significant at the 10% level. Along with the components of work satisfaction, this suggests that managers respond to worker evaluations by being nicer and more responsive to workers.

While the evidence indicates that managers are nicer to workers after the intervention, the lack of increase in the effort of workers who stay (as measured by worker performance in Table 4) provides suggestive evidence against the idea of gift exchange between workers and managers driving the results. Rather, the results suggest that managers trade off exerting effort in cracking the whip (and make production targets) and being nice which increases worker job satisfaction and

[29]Note that this is one of two work satisfaction variables along which the treatment and control groups are statistically different at baseline at the 10% level or higher (Appendix Table A.6). However, the magnitude of the difference is fairly small (0.14 on a five point scale).

[30]This is asked since September of the previous year (after the start of the intervention) in the follow-up so the time frame is approximately almost one year. Note that the lagged dependent variable, which is the same question in the baseline survey, is asked about the past month.

[31]The options are often, sometimes, rarely or not relevant (because they do not post on WeChat).

[32]Most of the WeChat posts are not work related.

reduces turnover.

7.3 Results on Worker Voice

We now consider whether there is evidence to support the idea that the results are driven by giving workers voices, absent any changes by managers or higher management. To test this idea, we use variations in the salience of the intervention months after it has ended.[33] In particular, we make use of questions in the follow-up survey asking in the Summer of 2017 whether workers provided feedback on their managers' management skills since the Fall of 2016. Conditional on answering yes to that question, the workers are then asked whether the feedback that they provided affected in any way the earnings of their manager. We ask similar questions of managers: whether their workers provided feedback on their management skills, and conditional on answering yes to that question, they were then asked whether their workers' feedback affected their earnings.

To check that these responses make sense, we first examine the impact of the treatment on these responses in Table 10. Workers in the treatment teams are 56% more likely to report participating in giving feedback about their managers, and this estimate is significant at the 1% level (column 1). Similarly, managers in the treatment teams are 66% more likely to report that they received feedback in the past year from the workers in their team (column 4) and this is also significant at the 1% level.

In columns 2 and 5, the outcome variable is based on the question of whether the feedback affected the earnings of the managers where we fill this variable in as zero if they answered that they didn't give (for workers) or receive (for managers) worker feedback in the prior question. In columns 3 and 6, it's the same outcomes but we do not fill in based on the prior question; if they answer that they didn't participate, the question is missing. For workers, being in the treatment group significantly increases the probability that they reported being aware that the evaluations affect managers' earnings. However, the magnitudes here are much smaller than for those participating in the program. The estimates are not significant for managers, suggesting that it was not very salient to the managers that the formula determining their monthly bonus placed

[33]The feedback occurred from September 2016 to April 2017 and they were surveyed for the follow-up in July and August 2017.

weight on this feedback. This provides additional evidence to suggest that the financial incentives were not a major driver of the results, an interpretation consistent with the prior results of Figure 1 where the shift in production and management KPI in treatment teams persisted after the end of the financial incentives.

Turning to our test of the voice mechanism, we can now examine whether workers' awareness and/or managers' awareness explain the impacts of the treatment. In a sample limited to treatment teams, we estimate the impacts of whether the worker and whether the manager reported participating in the worker evaluation in the follow-up survey on the outcomes for which we saw significant impacts of the treatment. The results are presented in Table 11. The salience of the feedback for the manager has a positive and significant impact on total KPI. More specifically, in teams for which the manager finds the treatment salient, the team-level total KPI is 5 points higher than in teams where the treatment wasn't salient for the manager. The estimates of manager awareness on production KPI and management KPI are positive but not significant at the standard levels. The managers' awareness does have positive and significant effects on workers' life satisfaction and positive mood in columns 4 and 5, respectively. Controlling for managers' awareness though, workers' own awareness of the treatment is not significantly related to any of the outcomes. Overall, the results provide suggestive evidence that the main mechanism is not through worker voice and empowerment absent changes in the managers' behavior. What really matters is the managers' awareness of the feedback, and the way that they change their management behavior as a result.

7.4 Results on Team Composition

We consider whether the main effects of the treatment are explained by differential changes in the composition of the teams. We consider the following changes in team composition: manager turnover, workers shifting across teams, and the types of workers separating from the firm.

In column 1 of Table 12, the dependent variable is whether the manager for the team changed between the follow-up and baseline surveys. There are no significant differences in manager turnover between the treatment and control teams. The next two variables in Table 12 are different measures of whether workers have switched to a different team within the plant. In column 2, the measure is

at the worker level. In column 3, the dependent variable is the team-level total number of workers who have switched teams between the baseline and follow-up surveys. In both measures, we see that there are no significant differences in the probability of workers moving across teams for treatment teams as compared to control teams.

Table 13 shows the average characteristics of workers who are likely to switch teams. Not surprisingly, workers who were performing poorly are likely to be shifted to different teams. In particular, we see that workers who have switched have significantly lower individual scores on production and quality. Given that we do not see individual productivity measures changing significantly for treatment and control teams (in Table 4), it's clear that they are not moving more low performing workers to control teams. Furthermore, there are no significant differences in terms of the KPI of the teams that they were in at baseline or their well-being measures. This provides additional evidence that the results on mood and life satisfaction cannot be explained by the firm moving unhappy workers from treatment teams to control teams.

We now consider whether the characteristics of the types of people who leave versus stay might explain the results. Given that firm separations are higher in the control teams, one possible explanation for the results on the well-being measures is that workers with higher measures of well-being are more likely to leave. In Table 14, we look at baseline averages for individuals who stay at the firm as compared to individuals who leave the firm by the follow-up survey. Workers who leave the firm have significantly lower scores on attitude (a category that the firm evaluates them on) and were working in teams with higher average levels of KPI. They also have slightly lower individual total performance scores, though this difference is only statically significant at the 11% level. They are not statistically different along the metrics of well-being. This suggests that changes in the characteristics of stayers and leavers are not driving the results where workers in treatment teams are having higher levels of life satisfaction and positive mood.

A final composition-related mechanism that we consider is that teams are unable to hire to replace team members who leave the firm (either because the firm will not allow new hires or because it is difficult to find a qualified worker to take the job). Then, if the intervention leads to lower turnover in the treatment teams, they are able to have higher levels of team-level productivity and

better mood simply because the teams are better staffed and workers do not need to permanently share the burden of departing team members. This is unlikely given that we have already seen that there are no reported differences in workers feeling their workload is too much in Table 8. Column 4 of Table 12 also confirms that this mechanism is not relevant as we see no significant differences in team size between treatment and control teams at the time of follow-up.

We argue that the team-level productivity results are driven by changes in team composition, in particular by the lower turnover rates in the treatment teams as compared to the control teams. However, no other changes in the composition of teams, including changes in managers or shifting workers across teams, explains the results. Furthermore, the characteristics of the people who change teams or leave the firm do not explain the main results.

7.5 Persistence and New Workers

The primary mechanism for the results appears to be changes in the way that managers interact with their workers. We have already seen evidence that the results are persistent for several months after they stop providing feedback in treatment teams in Figure 1. This is consistent with informal conversations with workers we had two years after the follow-up survey was conducted, where many workers stated that the intervention permanently changed their working relationship with their managers. Even after they stopped formally providing feedback to their managers, they said that their managers continued to listen to them more. We can also examine the outcomes of new entrants (who began after the intervention ended in April 2017) to test for whether there is evidence that the improvements in the relationship between managers and workers persisted. There were 69 new entrants defined in this way.

The results are presented in Table 15 where we focus on individual-level variables that were significantly affected by the intervention: two measures of well-being (life satisfaction and positive mood) and four components of job satisfaction (the incompetence of others, managers' encouragement, managers' criticisms, and managers' empathy). The results indicate that the new entrants to the treatment teams, who never participated in the feedback, report that their managers give more encouragement and provide more empathy than new entrants to control teams. These two

estimates are significant at the 5% and the 10% levels, respectively. We also observe higher levels of life satisfaction and positive mood among the new entrants to treatment teams, though these are only statistically different from zero at the 13% and 12% levels, respectively. Overall, these results not only suggest additional support for the mechanism that managers changed their interactions with workers, but also suggest that these changes persisted beyond the period in which they were receiving feedback from their subordinates.[34]

8 Robustness Checks

We consider whether the results are robust to survey attrition and spillovers from the treatment teams to the control teams.

8.1 Survey Attrition

Workers attrite from our analysis when they leave the firm, and we conclude that these firm separations are different in treatment and control teams and explain the results. Here, we consider the idea that there is attrition from the survey (among workers who still work at the firm).[35] This could potentially explain the results based on survey responses, including job and life satisfaction and mood, but this issue cannot explain the results based on administrative data from the firm, including team-level productivity and individual-level performance. Survey attrition in the follow-up survey (not including firm separations) is extremely low at 1.65% of the sample. More importantly, there are no differences in survey attrition between the treatment and control teams.[36]

[34] An alternative possible explanation is that new entrants are not randomly assigned to teams but that new applicants who were happier were assigned to teams that had lower turnover rates (i.e. treatment teams). In Appendix Table A.10, we examine characteristics that are fixed since the time of hire or very likely not to have changed (i.e. married) since being hired within the past couple of months. The observable characteristics of new entrants were not significantly different for treatment and control teams.

[35] This could occur if the workers were absent from work when we were interviewing the team.

[36] The coefficient in a regression including 1272 observations on treatment where the dependent variable is survey attrition is 0.010 with a standard error of 0.009.

8.2 Spillovers

We will consider two potential ways that spillovers could affect the results. One primary concern regarding spillovers is that individuals in the control group found out about the intervention in the treatment teams and reacted negatively to not being given the opportunity to evaluate their managers. Thus, the results could be driven by workers in the control group leaving the firm or reporting lower satisfaction and worse relationships with their managers because they were upset about being excluded from the evaluations. This concern could lead to overestimates in the impact of the intervention and a misattribution of the mechanism. We do two things to consider this concern. First, we asked questions about whether individuals discussed the intervention with others in the follow-up survey. Second, we collected information about social networks in the baseline survey and use that information to estimate spillover effects by looking at whether the treated share of individuals in one's social network matters for the results.

A second concern is that there are direct productivity spillovers across teams along a production line. This type of spillover could lead to underestimates of the productivity effects of the intervention if the control teams are positively affected. To examine this concern, we estimate the impact of having a treated team immediately upstream in the production line.

Appendix Table A.11 shows the summary statistics for the questions in the follow-up survey related to discussing and hearing about the intervention. The three questions on discussing the intervention are asked conditional on answering yes to the question about participating in the feedback (referred to in column 1 of Table 10). We asked whether they discussed the intervention with their managers, with their colleagues (on the same team) and with employees on other teams. Overall, most workers didn't report talking about the feedback system to anyone. As shown in Panel A, among those who reported participating in giving feedback on their managers, 13.57% reported discussing the feedback in any way with their managers, and less than one-quarter discussed it with workers on their team. And only 3.2% of workers reported discussing the feedback with workers on other teams. Panel B shows the responses to the same questions for managers. The vast majority of managers didn't talk about the feedback program with their workers (82%), managers on other teams (73%) and their higher-level supervisors (75%). Conditional on answering no to participating

in the feedback intervention, workers were asked whether they had heard of workers on other teams being able to do this. Only 11% of workers who didn't participate in the feedback themselves said that they had heard of other teams doing this. This partially mitigates concern that there was a lot of discussion across teams about the intervention.

To further test the importance of spillovers in the control teams, we also make use of social network information that we collected in the baseline survey. We ask all survey respondents to identify any relatives who are currently working at the firm, friends on other teams, people who referred them to the firm and people whom they referred to the firm.[37] In Table 16, we include a variable for the number of people in their social network outside of their own team that were treated plus its interaction with the treatment status of the individual.[38] Thus, the key coefficient of interest is the one on $Network\ Treated$ which provides information on whether the outcomes of individuals in control teams are affected by having more people in their social network exposed to the treatment. For all of the outcome variables, we see no significant effects of having more treated social connections on the outcomes for control group individuals. This provides additional evidence that the main results are not being driven by negative spillover in the control group.

Finally, we consider the possibility that there are productivity spillovers across teams along the production line. To some extent, manufacturing teams are organized to minimize production spillovers, but we consider the possibility that higher productivity from the team directly upstream affect a team's own KPI outcomes. To do this, we include a variable on whether the team upstream in the production line within a plant is treated in the regressions of team-level KPI. It is also common for multiple teams to feed into the production process for a team; in this case, the variable is equal to zero if less than half of the adjacent upstream teams were treated and equal to one if more than half of the adjacent upstream teams were treated.

The results are presented in Table 17. Whether the upstream teams are treated or not has no significant effects on total KPI, production KPI or management KPI. The results indicate that

[37] In the survey, we allow respondents to list up to 6 family members, 8 friends on other teams, 5 people they referred, and 2 people who referred them. These maximums are hit in less than half of one percent of cases. The mean number of family or friends on other teams is 1.41. They have referred on average 0.13 people to other teams and received on average 0.13 referrals from people on other teams.

[38] An alternative measure is the *share* of their social network treated rather than the total level. The results, shown in Appendix Table A.12, are not substantively different.

there are no production spillovers resulting from the intervention across teams. Overall, there is no evidence to suggest that spillovers bias the results either through negative effects on the control group or through positive spillover effects across the production line.

9 Conclusion

The results of our paper show that letting workers participate in evaluating their managers reduces worker turnover by over 50%. Although the intervention does not affect individual productivity, team-level productivity does increase significantly. This is driven by the fact that treated teams have less worker turnover than control teams, and high turnover reduces team productivity because new workers normally perform poorly in the beginning and need substantial training. Furthermore, we find that workers in teams that were given the opportunity to evaluate their managers report higher levels of happiness, and there is a significant increase in the measure of positive mood. This is consistent with the idea that the intervention led workers to be happier and less likely to quit.

Examining factors driving the observed results, we find direct evidence for one mechanism: changes in managers' behavior and their relationship with workers. Specifically, treated workers report that their team managers encouraged them more, criticized them less, and provided more empathy for the feelings of the workers. Moreover, treated managers had more social interactions with workers outside of work. We also rule out two other potential mechanisms, including empowering workers by giving them voices and changes of team composition.

An important question for future research is: why didn't managers engage in a more empathetic and encouraging relationship with their workers prior to the intervention? Given that the less critical behavior reduced worker turnover, increased team-level KPI and increased the managers' monthly earnings, rational economic models suggest that they should have already been choosing their managerial inputs to reduce turnover and maximize team performance. One possibility is that managers may have underestimated the value of retention. Related to the idea that managers are making mistakes, it is possible that behavioral biases may explain the results. For example, the manager may know that improving their relationships with workers will lead to better outcomes, but is present-biased where there may be costs to productivity today in order to increase future

retention and productivity.

An alternative possibility is that managers may estimate correctly the value of retention but are not able to gather the information needed to know how to increase retention. Indeed, it is quite possible that if managers try to directly solicit feedback from their workers about their managerial skills, workers would not feel comfortable being honest. Thus, it may be the case that outside intervention (either by a third party or other parties in the firm that are not the manager) is necessary for workers to feel confident that their individual responses are anonymous. These are thought-provoking questions that we leave for future research.

Importantly, our results pass the market test. After seeing our findings, our partner company, the large automobile-manufacturing firm in China, decided to change company policy to incorporate worker feedback in all their plants, covering thousands of production teams and almost 20,000 production-line workers. Our results thus suggest that letting workers participate in their managers' evaluation is an effective and inexpensive way of improving management practices, increasing worker satisfaction, and reducing turnover.

References

Adhvaryu, Achyuta, Namrata Kala, and Anant Nyshadham. 2019. "Management and Shocks to Worker Productivity." *NBER Working Paper No. 25865* .

Adhvaryu, Achyuta, Teresa Molina, and Anant Nyshadham. 2020. "Expectations, Wage Hikes, and Worker Voice." *Working Paper* .

Akerlof, George A. 1982. "Labor contracts as partial gift exchange." *The quarterly journal of economics* 97 (4):543–569.

Antonioni, David. 1996. "Designing an effective 360-degree appraisal feedback process." *Organizational Dynamics* 25 (2):24–38.

Atwater, E. Leanne, David A. Waldman, David Atwater, and Priscilla Cartier. 2000. "An Upward Feedback Field Experiment: Supervisors' Cynicism, Reactions, and Commitment to Subordinates." *Personnel Psychology* 53 (2):275–297.

Bandiera, Oriana, Iwan Barankay, and Imran Rasul. 2007. "Incentives for Managers and Inequality Among Workers: Evidence from a Firm-level Experiment." *The Quarterly Journal of Economics* 122 (2):729–773.

———. 2009. "Social Connections and Incentives in the Workplace: Evidence from Personnel Data." *Econometrica* 77 (4):1047–1094.

Bartling, Barting, Ernst Fehr, and Holger Herz. 2014. "The Intrinsic Value of Decision Rights." *Econometrica* 82 (6):2005–2039.

Batt, Rosemary, Alexander J.S. Colvin, and Jeffrey Keefe. 2002. "Employee Voice, Human Resource Practices, and Quit Rates: Evidence from the Telecommunications Industry." *Industrial and Labor Relations Review* 34 (1):573–594.

Bebchuk, Lucian A., Alma Cohen, and Scott Hirst. 2017. "The Agency Problems of Institutional Investors." *Journal of Economic Perspectives* 31 (3):89–112.

Bellemare, Charles and Bruce Shearer. 2009. "Gift giving and worker productivity: Evidence from a firm-level experiment." *Games and Economic Behavior* 67 (1):233–244.

Blatter, Marc, Samuel Muehlemann, Samuel Schenker, and Stefan C. Wolter. 2015. "Hiring Costs for Skilled Workers and the Supply of Firm-Provided Trainings." *Oxford Economic Papers* 68 (1):238–257.

Bloom, Nicholas, Benn Eifert, Aprajit Mahajan, David McKenzie, and John Roberts. 2013. "Does Management Matter? Evidence from India." *The Quarterly Journal of Economics* 128 (1):1–51.

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

Bruhn, Miriam, Dean Karlan, and Antoinette Schoar. 2018. "The impact of consulting services on small and medium enterprises: Evidence from a randomized trial in Mexico." *Journal of Political Economy* 126 (2):635–687.

Callegarin, Myriam. 2017. "High Employee Turnover in China is not a Matter of Money." URL `https://www.newpointdeview.com/pro-en/intercultural-compentence/high-employee-turnover-in-china-not-matter-of-money/`.

Carpenter, Mason A. and WM. Gerard Sanders. 2002. "Top Management Team Compensation: the Missing Link Between CEO Pay and Firm Performance?" *Strategic Management Journal* 23 (4):367–375.

Cooke, Williams. 1994. "Employee Participation Programs, Group-Based Incentives, and Company Performance: A Union-Nonunion Comparison." *Industrial and Labor Relations Review* 47 (4):594–609.

Coughlan, Anne T. and Ronald M. Schmidt. 1985. "Executive Compensation, Management Turnover, and Firm Performance: An Empirical Investigation." *Journal of Accounting and Economics* 7 (1-3):43–66.

Coviello, Decio, Erika Deserranno, and Nicola Persico. 2020. "Minimum Wage and Individual Worker Productivity: Evidence from a Large US Retailer." *Working paper* .

Craig, Ben and John Pencavel. 1995. "Participation and Productivity: A Comparison of Worker Cooperatives and Conventional Firms in the Plywood Industry." *Brookings Papers: Microeconomics* :121–174.

Dube, Arindrajit, Eric Freeman, and Michael Reich. 2010. "Employee Replacement Costs." *Working Paper* .

Frederiksen, Anders, Lisa B. Kahn, and Fabian Lange. 2019. "Supervisors and Performance Management System." *NBER Working Paper No. 23351* .

Freeman, B. Richard. 1980. "Unionism and the Dispersion of Wages." *Industrial and Labor Relations Review* 34 (1):3–23.

Gosnell, Greer K., John A. List, and Robert D. Metcalfe. 2019. "The Impact of Management Practices on Employee Productivity: A Field Experiment with Airline Captains." *NBER Working Paper No. 25620* .

Grossman, Sanford J. and Oliver D. Hart. 1983. "Implicit Contracts Under Asymmetric Information." *The Quarterly Journal of Economics* 98:123–156.

Heslin, A. Peter and Gary P. Latham. 2004. "The Effect of Upward Feedback on Mangerial Behavior." *Applied Psychology: an International Review* 53 (1):23–37.

Holmstrom, Bengt and Paul Milgrom. 1987. "Aggregation and Linearity in the Provision of Intertemporal Incentives." *Econometrica* 55 (2):303–328.

Jacoby, Stanford M. 1983. "Industrial Labor Mobility in Historical Perspective." *Industrial Relations* 22 (2):261–282.

Jensen, Michael and Kevin Murphy. 1990. "Performance Pay and Top-Management Incentives." *Journal of Political Economy* 98 (2):225–264.

Kelly, Erin L., Phyllis Moen, J. Michael Oakes, Wen Fan, Cassandra Okechukwu, Kelly D. Davis, Leslie B. Hammer, Ellen Ernst Kossek, Rosalind Berkowitz King, Ginger C. Hanson, Frank Mierzwa, and Lynne M. Casper. 2014. "Changing Work and Work-Family Conflict: Evidence from the Work, Family, and Health Network." *American Sociological Review* 79 (3):1–32.

Lazear, Edward P. 2018. "Compensation and Incentives in the Workplace." *Journal of Economic Perspectives* 32 (3):195–214.

McKenzie, David and Christopher Woodruff. 2016. "Business Practices in Small Firms in Developing Countries." *Management Science* 63 (9):2773–3145.

Mehran, Hamid. 1995. "Executive Compensation Structure, Ownership, and Firm Performance." *Journal of Financial Economics* 38 (2):163–184.

Milgrom, Paul R. 1988. "Employment Contracts, Influence Activities, and Efficient Organization Design." *Journal of Political Economy* 96 (1):42–60.

Moon, Ken, Patrick Bergemann, Daniel Brown, Andrew Chen, James Chu, Ellen A. Eisen, Gregory M. Fischer, Prashant Loyalka, Sungmin Rho, and Joshua Cohen. 2019. "Manufacturing Productivity with Worker Turnover." *Working paper* .

Moon, Ken, Prashant Loyalka, Patrick Bergemann, and Joshua Cohen. 2020. "The Hidden Cost of Worker Turnover: Attributing Product Reliability to the Turnover of Factory Workers." *Working paper* .

Muehlemann, Samuel and Harald Pfeifer. 2016. "The Structure of Hiring Costs in Germany: Evidence from Firm-Level Data." *Industrial Relations* 55 (2):193–218.

Murphy, Kevin J. 1999. "Executive Compensation." *In Orley Ashenfelter and David Card (eds), Handbook of Labor Economics* 3b (38):2485–2563.

Owen, Laura. 1995. "Worker Turnover in the 1920s: The Role of Changing Employment Policies." *Industrial and Corporate Changes* 4 (3):499–530.

Schoar, Antoinette. 2013. "The Importance of Being Nice: Supervisory Skill Training in the Cambodian Garment Industry." *Working paper* .

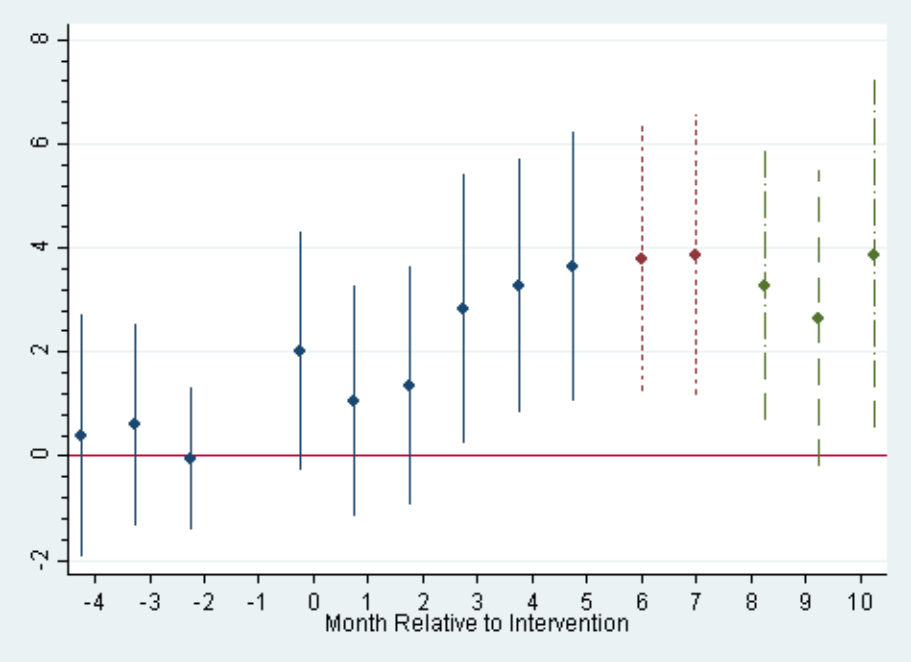
Seifert, Charles and Gary Yukl. 2010. "Effects of Repeated Multi-source Feedback on the Influence Behavior and Effectiveness of Managers: A Field Experiment." *The Leadership Quarterly* 21 (5):856–866.

Seifert, Charles, Gary Yukl, and Robert A. McDonald. 2003. "Effects of Multisource Feedback and Feedback Facilitator on the Influence Behavior of Mangers Toward Subordinates." *Journal of Applied Psychology* 88 (3):561–569.

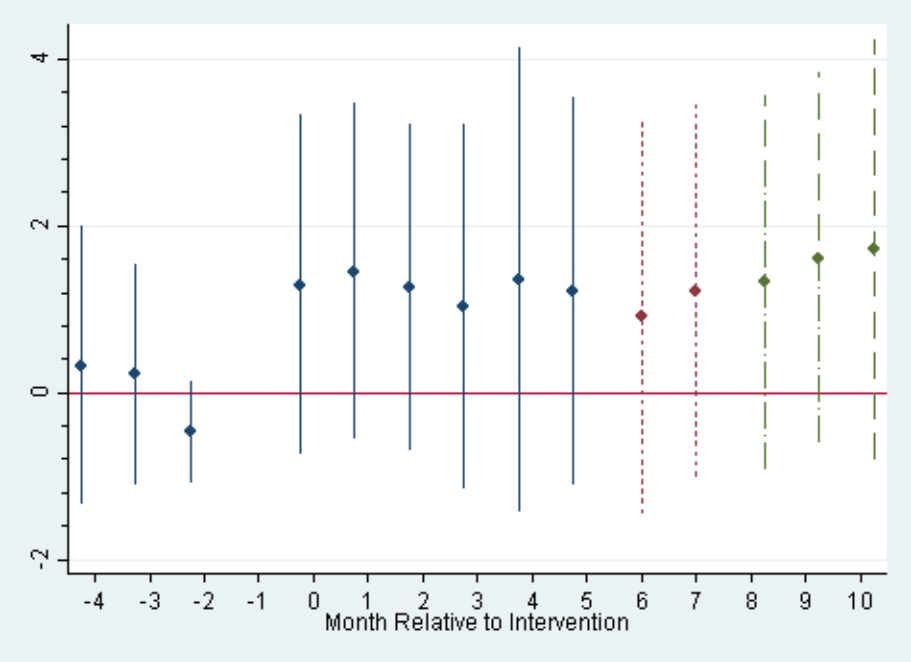
World Bank. 2013. "World development report 2013: Jobs."

Figure 1: Coefficients on Treatment × Month around Intervention

(a) Production KPI



(b) Management KPI



Note: The dependent variable is Production KPI in Panel A and Management KPI in Panel B. The dots indicate the coefficients on the interactions between treatment and month around intervention. The regressions include indicators for month around intervention, randomization strata, KPI form version and team. The bars show the 90% confidence intervals around the estimates. The periods shown in red with the short dashed line as the confidence intervals (6 and 7) indicate that the feedback is being provided but the incentives have ended. The periods 8 through 10 shown in green with the alternating long dashes and dots line as the confidence intervals indicate periods where the feedback has ended.

Table 1: Pre-Intervention Summary Statistics for the Administrative Data

	Treatment	Control	N	p-value
Panel A: Worker Performance Measures				
Total	99.55	99.08	1094	.6520
	(4.697)	(17.64)		
Production	.4453	.1327	1075	.2006
	(3.500)	(1.822)		
Safety	-.071	-.098	1069	.7248
	(.4473)	(.4161)		
Quality	.2017	-.031	1069	.0806
	(.9646)	(.6412)		
Attitude	.0688	.0702	1069	.9788
	(.6315)	(.6324)		
Routine	-.376	.0164	1070	.1106
	(1.422)	(1.134)		
Attendance	.0764	.7278	1075	.4446
	(1.948)	(17.37)		
Panel B: Team-Level KPI				
Total	98.11	97.15	76	.5556
	(6.930)	(7.147)		
Production	32.98	33.83	74	.7544
	(11.09)	(12.00)		
Management	20.40	20.12	76	.8115
	(5.386)	(4.600)		
Equipment	10.22	10.05	30	.9145
	(4.721)	(3.428)		
Quality	26.87	29.07	70	.5670
	(15.96)	(15.97)		
Safety	19.96	16.21	76	.1720
	(13.25)	(10.07)		

The mean of each variable is shown for the average of all pre-invention months with standard deviations below in parentheses. The p-value is taken from a regression testing the statistical difference between the two groups.

Table 2: Pre-Intervention Summary Statistics for the Survey Data

	Treatment	Control	N	p-value
Age	32.80	33.73	1251	.4982
	(9.359)	(9.591)		
Male	.9226	.8800	1250	.2211
	(.2672)	(.3251)		
High School or More	.4703	.4525	1251	.5975
	(.4995)	(.4981)		
Tenure at Firm	4.490	4.947	1251	.4985
	(4.967)	(5.091)		
Married	.6474	.6671	1247	.6746
	(.4781)	(.4715)		
Life Rating	7.092	6.795	1248	.0361
	(1.892)	(1.974)		
Good Health	.6546	.6345	1251	.6402
	(.4758)	(.4819)		
Positive Mood	29.27	28.60	1247	.2414
	(6.614)	(7.183)		
Negative Mood	17.96	17.38	1246	.1443
	(6.372)	(5.905)		
Work Satisfaction	35.65	35.81	1250	.8489
	(8.870)	(9.453)		
Monthly Earnings	3690.6	3652.1	1235	.7030
	(802.9)	(769.1)		
Monthly Bonus	1357.0	1315.2	1228	.5930
	(594.2)	(562.1)		
Socialize with Manager	.5570	.5981	1247	.7580
	(1.325)	(1.341)		

The mean of each variable is shown for the average of all pre-invention months with standard deviations below in parentheses. The p-value is taken from a regression testing the statistical difference between the two groups.

Table 3: Impact of Evaluating Managers on Leaving Firm

	Workers	Managers
	(1)	(2)
Treated	-0.0622***	-0.0361
	(0.0178)	(0.0272)
Obs	1424	73
Mean of Dep Var	0.107	0.0137

Notes: Regressions include controls for randomization strata and second treatment. Standard errors clustered at the team level are shown in parentheses. ***, **, * denotes significance at the 1, 5 and 10% levels respectively.

Table 4: Impact of Evaluating Managers on Individual Performance

	Total	Production	Safety	Quality	Attitude	Routine	Attendance
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treated	0.142	0.0184	-0.347	0.0596	-0.144	-0.0888	0.0358
	(0.375)	(0.0608)	(0.320)	(0.0447)	(0.189)	(0.0609)	(0.0695)
Lag DepVar	1.090***	0.725***	0.671	0.596***	0.438***	0.804***	1.181***
	(0.143)	(0.0311)	(0.405)	(0.0754)	(0.0724)	(0.0697)	(0.109)
Obs	963	949	946	944	945	945	946

Notes: Regressions include controls for randomization strata and second treatment. Standard errors clustered at the team level. ***, **, * denotes significance at the 1, 5 and 10% levels respectively.

Table 5: Impact of Evaluating Managers on Team KPI

	Total	Production	Management	Equipment	Quality	Safety
	(1)	(2)	(3)	(4)	(5)	(6)
Treated	2.264**	3.714***	1.899**	0.566	0.419	0.541
	(1.075)	(1.289)	(0.808)	(0.603)	(1.070)	(1.184)
Lag DepVar	0.471***	0.929***	0.476***	0.876***	0.951***	0.827***
	(0.0898)	(0.102)	(0.0783)	(0.106)	(0.0320)	(0.133)
Obs	70	68	70	26	58	70

Notes: Regressions include controls for randomization strata and second treatment. Standard errors clustered at the team level. ***, **, * denotes significance at the 1, 5 and 10% levels respectively.

Table 6: Impact of Treatment on Earnings

	Workers		Managers	
	Earnings	Bonus	Earnings	Bonus
	(1)	(2)	(3)	(4)
Treated	-0.00264	-0.153	0.0388	0.228**
	(0.0166)	(0.125)	(0.0288)	(0.106)
Lag DepVar	0.762***	0.664***	0.741***	0.590***
	(0.0436)	(0.190)	(0.112)	(0.177)
Obs	1114	1020	70	70

Notes: The dependent variables, total monthly earnings and monthly bonus, are transformed using the inverse hyperbolic sine (IHS) function. Regressions include controls for randomization strata, enumerators, survey month and second treatment. Standard errors clustered at the team level are shown in parentheses. ***, **, * denotes significance at the 1, 5 and 10% levels respectively.

Table 7: Impact of Evaluating Managers on Well-Being

	Life Rating	Work Satisfaction	Positive Mood	Negative Mood	Health Rating
	(1)	(2)	(3)	(4)	(5)
Panel A: All Workers Sample					
Treated	0.324***	0.665	0.633*	-0.0145	0.0402
	(0.116)	(0.635)	(0.378)	(0.319)	(0.0295)
Lag DepVar	0.292***	0.451***	0.418***	0.384***	0.413***
	(0.0348)	(0.0246)	(0.0305)	(0.0250)	(0.0317)
Obs	1126	1128	1125	1124	1129
Panel B: Managers Sample					
Treated	0.454	3.848	2.002	0.777	0.135
	(0.442)	(2.318)	(1.788)	(1.522)	(0.0935)
Lag DepVar	0.601***	0.401*	0.578***	0.369*	0.716***
	(0.146)	(0.171)	(0.0873)	(0.152)	(0.101)
Obs	62	63	62	62	63

Notes: Regressions include controls for randomization strata, enumerators and second treatment. Standard errors clustered at the team level are shown in parentheses. ***, **, * denotes significance at the 1, 5 and 10% levels respectively.

Table 8: Impact of Evaluating Managers on the Components of Work Satisfaction

	Treated		Lag Dep Var		
	Coefficent	Std Error	Coefficient	Std Error	N
Promotion	0.044	(0.062)	0.257***	(0.062)	1129
Other Incompetence	-0.152*	(0.078)	0.201***	(0.078)	1129
Job Meaningless	-0.081	(0.051)	0.255***	(0.051)	1129
Encourages	0.128**	(0.049)	0.249***	(0.049)	1129
Politics	0.010	(0.081)	0.264***	(0.081)	1129
Workload	0.062	(0.082)	0.222***	(0.082)	1129
Unfairness	-0.051	(0.072)	0.169***	(0.072)	1056
Raises	0.020	(0.077)	0.317***	(0.077)	1129
Switch Team	-0.005	(0.065)	0.187***	(0.065)	1128
Criticizes	-0.162**	(0.064)	0.194***	(0.064)	1129
Bad Rules	-0.113	(0.071)	0.239***	(0.071)	1129
No Empathy	-0.164**	(0.066)	0.223***	(0.066)	1056
Effort Rewarded	-0.020	(0.079)	0.280	(0.079)	1129

Notes: Regressions include controls for randomization strata, enumerators and second treatment. Standard errors clustered at the team level are shown in parentheses. ***, **, * denotes significance at the 1, 5 and 10% levels respectively.

Table 9: Impact of Evaluating Managers on Socializing with Managers

	Socialize Manager	WeChat Manager
	(1)	(2)
Treated	0.518	0.0713*
	(0.903)	(0.0390)
Lag DepVar	4.340***	
	(1.448)	
Obs	1125	922
Mean of Dep Var	5.277	0.632

Notes: Regressions include controls for randomization strata, enumerators and second treatment. Standard errors clustered at the team level are shown in parentheses. ***, **, * denotes significance at the 1, 5 and 10% levels respectively.

Table 10: Impact of Treatment on Awareness of Scoring

	Workers' Awareness			Managers' Awareness		
	Scored Manager	Affected Wage (Fill)	Affected Wage (Unfill)	Scored Manager	Affected Wage (Fill)	Affected Wage (Unfill)
	(1)	(2)	(3)	(4)	(5)	(6)
Treated	0.564***	0.301***	0.129**	0.664***	0.194	0.0966
	(0.0349)	(0.0417)	(0.0621)	(0.117)	(0.139)	(0.387)
Obs	1031	820	378	63	62	39

Notes: Regressions include controls for randomization strata, enumerators and second treatment. Standard errors clustered at the team level are shown in parentheses. ***, **, * denotes significance at the 1, 5 and 10% levels respectively.

Table 11: Impact of Awareness of Policy (Manager Versus Worker)

	Total KPI	Product KPI	Manage KPI	Life Satis.	Positive Mood
	(1)	(2)	(3)	(4)	(5)
Manager Aware Scoring	5.088***	1.546	1.782	0.583***	1.760*
	(1.434)	(1.785)	(1.330)	(0.207)	(0.965)
Worker Aware Scoring	0.379	0.0571	0.893	0.252	0.0226
	(0.725)	(0.672)	(0.705)	(0.190)	(0.676)
Lag DepVar	0.526***	0.866***	1.110***	0.275***	0.426***
	(0.127)	(0.0830)	(0.216)	(0.0477)	(0.0495)
Obs	605	605	605	562	561

Notes: The sample is limited to the treatment teams. Regressions include controls for enumerators and second treatment. Standard errors clustered at the team level are shown in parentheses. ***, **, * denotes significance at the 1, 5 and 10% levels respectively.

Table 12: Impact of Evaluating Managers on Team Composition

	Manager Changes (1)	Worker Changes (2)	Total Worker Changes (3)	Team Size (4)
Treated	-0.0303	0.0140	0.0791	-0.0918
	(0.0814)	(0.0444)	(0.499)	(0.755)
Lag Dep Var				0.830***
				(0.0719)
Obs	66	1060	71	71
Mean Dep Var	0.152	0.167	2.493	18.45

Notes: The dependent variable in column 1 is whether the manager of the team changed between the two survey rounds. In column 2, it is whether the worker changed teams. In column 3, it is the total number of workers who shifted teams at the team-level. In column 4, it is total team size at follow-up. Regressions include controls for randomization strata, and second treatment. Standard errors clustered at the team level are shown in parentheses. ***, **, * denotes significance at the 1, 5 and 10% levels respectively.

Table 13: Baseline Summary Statistics by Team Composition Changes

	Team Stayers Mean	Team Changers Mean	p-value Diff	Stayers Obs	Changers Obs
Total KPI	98.038	97.901	0.889	904	270
Production KPI	33.971	34.822	0.610	884	256
Management KPI	20.591	21.056	0.430	904	270
Equipment KPI	8.920	10.679	0.285	256	116
Quality KPI	27.174	27.821	0.775	877	248
Safety KPI	19.848	15.956	0.002	904	270
Total PFM	99.163	99.514	0.599	820	256
Production PFM	0.382	0.011	0.020	808	254
Safety PFM	-0.079	-0.105	0.743	802	254
Quality PFM	0.124	-0.041	0.034	802	254
Attitude PFM	0.074	0.025	0.166	802	254
Routine PFM	-0.199	-0.235	0.801	802	254
Attendance PFM	0.538	-0.095	0.264	807	255
Firm Tenure	4.388	4.589	0.719	909	270
Life Rating	6.910	6.900	0.947	907	270
Job Satisfaction	35.903	35.000	0.302	908	270
Positive Mood	28.881	28.597	0.617	908	268
Negative Mood	17.653	17.899	0.590	907	268

Notes: Column 1 shows the mean value at baseline for workers who are in the same team at follow-up. Column 2 shows the mean value at baseline for workers who have changed teams between the baseline and follow-up. Column 3 shows the p-value from a regression testing the statistical difference between the two groups.

Table 14: Baseline Summary Statistics of Firm Stayers and Leavers

	Stayers Mean	Leavers Mean	p-value Diff	Stayers Obs	Leavers Obs
Total KPI	97.986	99.486	0.028	1246	152
Production KPI	34.178	34.798	0.505	1211	151
Management KPI	20.666	20.807	0.720	1246	152
Equipment KPI	9.554	9.381	0.617	400	42
Quality KPI	27.308	28.563	0.409	1191	148
Safety KPI	18.864	17.791	0.313	1246	152
Total PFM	99.321	97.785	0.100	1094	142
Production PFM	0.289	0.336	0.894	1075	140
Safety PFM	-0.085	-0.097	0.817	1069	137
Quality PFM	0.085	0.002	0.128	1069	137
Attitude PFM	0.070	-0.139	0.025	1069	137
Routine PFM	-0.180	-0.189	0.947	1070	137
Attendance PFM	0.402	0.623	0.806	1075	140
Firm Tenure	4.725	1.789	0.000	1251	152
Life Rating	6.940	6.796	0.368	1248	152
Job Satisfaction	35.737	36.224	0.591	1250	152
Positive Mood	28.929	29.704	0.148	1247	152
Negative Mood	17.670	17.724	0.924	1246	152

Notes: Column 1 shows the mean value at baseline for workers who are still at the firm in the follow-up survey. Column 2 shows the mean value at baseline for workers who have left the firm by the follow-up survey. Column 3 shows the p-value from a regression testing the statistical difference between the two groups.

Table 15: Impact of Past Intervention on New Entrants' Outcomes

	Life Satis.	Positive Mood	Others Incompet.	Manager Encourage	Manager Criticize	Manager Empathy
	(1)	(2)	(3)	(4)	(5)	(6)
Treated	2.388	4.622	0.143	1.368**	-0.291	2.629*
	(1.516)	(2.861)	(0.602)	(0.521)	(0.683)	(1.364)
Obs	66	66	66	66	66	66

Notes: Regressions include controls for randomization strata, and second treatment. Standard errors clustered at the team level are shown in parentheses. ***, **, * denotes significance at the 1, 5 and 10% levels respectively.

Table 16: Heterogeneity by Treatment of Networks

	Total KPI	Product KPI	Manage KPI	Life Satis.	Positive Mood
	(1)	(2)	(3)	(4)	(5)
Network Treated	0.152	-0.0571	-0.0172	0.0770	0.280
	(0.145)	(0.128)	(0.0808)	(0.0632)	(0.268)
Treated X Network Treated	0.257	0.743***	0.0981	0.0794	0.154
	(0.180)	(0.238)	(0.120)	(0.101)	(0.370)
Treated	1.199	3.221***	0.714	0.289**	0.579
	(0.789)	(1.016)	(0.539)	(0.128)	(0.428)
Lag DepVar	0.420***	0.755***	0.509***	0.294***	0.414***
	(0.0710)	(0.0699)	(0.0583)	(0.0351)	(0.0305)
Obs	1123	1098	1123	1126	1125

Notes: Regressions include controls for randomization strata, enumerators and second treatment. Standard errors clustered at the team level are shown in parentheses. ***, **, * denotes significance at the 1, 5 and 10% levels respectively.

Table 17: Impact of the Treatment Status of Upstream Teams

	Total KPI	Product KPI	Manage KPI
	(1)	(2)	(3)
Upstream Treated	2.235	2.120	1.575
	(2.336)	(2.174)	(1.320)
Lag Dep Var	0.466***	0.930***	0.613***
	(0.0935)	(0.117)	(0.0796)
Obs	70	68	70

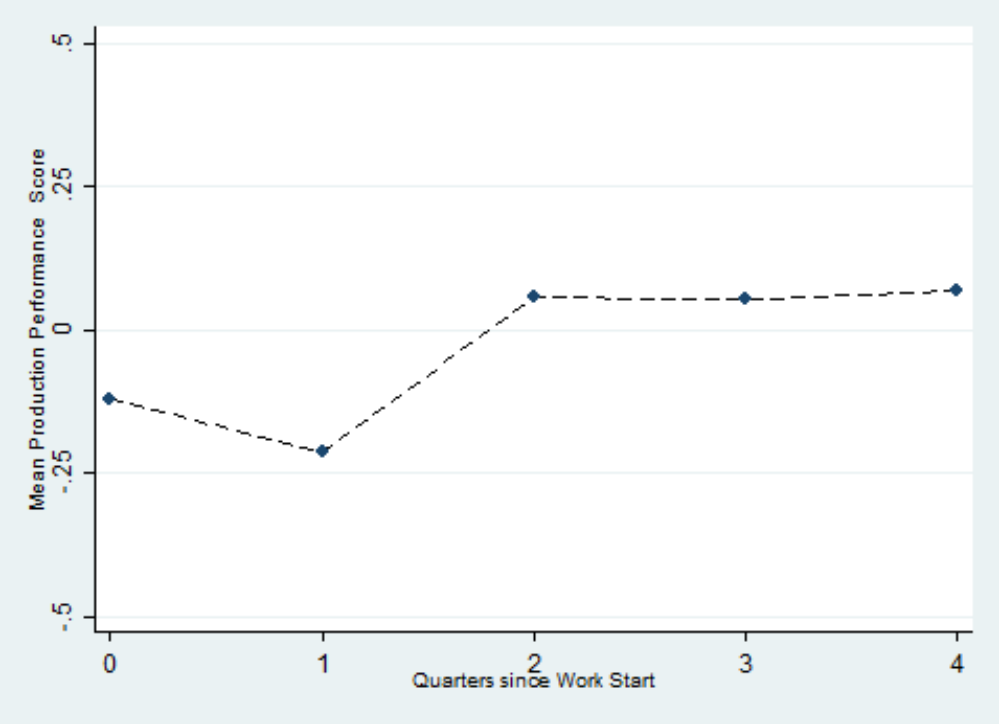
Notes: Regressions include controls for randomization strata as well as the primary and secondary treatments. Standard errors clustered at the team level are shown in parentheses. ***, **, * denotes significance at the 1, 5 and 10% levels respectively.

Table A.1: Manager Evaluation Questions

1	Organizing production
2	Exhibiting good judgement in emergencies and conflicts between workers
3	Treating all team members fairly, without showing favoritism
4	Listening to opinions and feedback from team members
5	Demonstrating understanding/compassion for workers in their work and lives

Notes: The question was "How effective was your team manager in the last month in the following dimensions?" The responses were a five point scale with 1 being not effective at all and 5 being extremely effective.

Figure A.1: Individual Productivity by Worker Tenure



Note: The figure shows the average production productivity score of individuals by the quarter of tenure within the firm.

Table A.2: Mood Questions

Positive	Negative
Interested	Distressed
Excited	Upset
Strong	Guilty
Enthusiastic	Scared
Proud	Hostile
Alert	Irritable
Inspired	Ashamed
Determined	Nervous
Attentive	Jittery
Active	Afraid

Notes: The PANAS mood questions are phrased in the following way. I would like to ask questions about your feelings and thoughts during last month. In each case, you will be asked to indicate how often you felt or thought a certain way in the last 30 days. Although some of the questions are similar, there are differences between them and you should treat each one as a separate question. The best approach is to answer each question fairly quickly. That is, do not try to count up the number of times you felt in a particular way, but rather indicate the option that seems like a reasonable estimate. For each question choose from the following: very slightly or not at all, a little, moderately, quite a bit, extremely.

Table A.3: Work Satisfaction Questions (5 Point Scale - Totally Disagree to Totally Agree)

1	There is really too little chance for promotion on my job.
2	The goals of this organization are not clear to me.
3	I am not satisfied with the benefits I receive.
4	I find I have to work harder at my job because of the incompetence of my teammates.
5	I sometimes feel my job is meaningless.
6	My manager often encourages me.
7	There is too much bickering and fighting at work.
8	I have too much to do at work.
9	My manager is unfair to me.
10	Raises are too few and far between.
11	I wish I could switch teams because I don't get along with my coworkers.
12	I'm often criticized by my manager.
13	Many of our rules and procedures make doing a good job difficult.
14	My manager shows too little interest in the feelings of subordinates.
15	I don't feel my efforts are rewarded the way they should be.

Table A.4: Winsorized Pre-Intervention Summary Statistics for the Administrative Data

	Treatment	Control	N	p-value
Total	99.65	99.85	1094	.6565
	(3.452)	(2.372)		
Production	.3037	.0902	1075	.1577
	(1.673)	(.8569)		
Safety	-.068	-.087	1069	.7744
	(.3364)	(.3340)		
Quality	.1750	-.012	1069	.0908
	(.7851)	(.4880)		
Attitude	.0673	.0735	1069	.8816
	(.4540)	(.4878)		
Routine	-.401	-.045	1070	.1297
	(1.365)	(.4380)		
Attendance	.0243	-.010	1075	.7243
	(.9938)	(.9618)		

The data are winsorized at the top and bottom 1%. The mean of each variable is shown for the average of all pre-invention months with standard deviations below in parentheses. The p-value is taken from a regression testing the statistical difference between the two groups.

Table A.5: Summary Statistics for KPI Weights

	Treatment	Control	N	p-value
Production	35	34.70	69	.9265
	(14.24)	(11.86)		
Management	20.85	21.14	71	.8159
	(4.615)	(5.568)		
Equipment	10.66	11.11	25	.7324
	(2.581)	(3.333)		
Quality	29.03	25.14	65	.3110
	(15.56)	(15.04)		
Safety	14.85	17.85	71	.3498
	(10.60)	(15.59)		

The mean of each variable is shown is the average across all treatment or control teams with standard deviations below in parentheses. The p-value is taken from a regression testing the statistical difference between the two groups.

Table A.6: Baseline Summary Statistics for Work Satisfaction Questions

	Treatment	Control	p-value	Obs.
Promotion	0.441	0.355	0.010	1350
Other Incompetence	2.374	2.518	0.057	1350
Job Meaningless	2.221	2.199	0.758	1350
Encourages	0.373	0.338	0.274	1350
Politics	2.557	2.512	0.556	1349
Workload	0.199	0.216	0.576	1350
Unfairness	0.109	0.115	0.732	1348
Raises	2.979	2.926	0.547	1350
Switch Team	1.848	1.870	0.756	1350
Criticizes	0.062	0.065	0.869	1350
Bad Rules	2.174	2.207	0.626	1350
No Empathy	0.125	0.120	0.786	1348
Effort Rewarded	2.392	2.443	0.581	1350

Notes: Column 1 shows the mean value at baseline for treatment group workers. Column 2 shows the mean value at baseline for control group workers. Column 3 shows the p-value from a regression testing the statistical difference between the two groups.

Table A.7: Leads and Lags Estimates of Impact of Evaluating Managers on Individual Performance

	Total	Production	Safety	Quality	Attitude	Routine	Attendance
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
4 months pre X Treat	0.106	-0.0695	-0.298*	0.0450	-0.282**	-0.0701	-0.333*
	(0.351)	(0.115)	(0.177)	(0.116)	(0.139)	(0.106)	(0.179)
3 months pre X Treat	0.418	-0.0359	-0.256	0.162	-0.224	-0.127	-0.139
	(0.290)	(0.113)	(0.183)	(0.101)	(0.151)	(0.099)	(0.170)
2 months pre X Treat	0.131	-0.0711	-0.115	0.00996	-0.240*	-0.0715	-0.301*
	(0.308)	(0.154)	(0.139)	(0.054)	(0.141)	(0.095)	(0.176)
0 months post X Treat	0.288	-0.286*	-0.186	0.0454	-0.184	-0.162*	-0.221
	(0.321)	(0.151)	(0.143)	(0.070)	(0.162)	(0.086)	(0.166)
1 months post X Treat	0.676	-0.0703	-0.169	-0.0484	-0.0639	-0.145	-0.549*
	(0.448)	(0.108)	(0.160)	(0.120)	(0.157)	(0.090)	(0.290)
2 months post X Treat	0.247	-0.0647	-0.195	-0.0856	-0.0549	-0.139	-0.785**
	(0.422)	(0.216)	(0.141)	(0.114)	(0.133)	(0.094)	(0.367)
3 months post X Treat	0.204	-0.253*	-0.227	0.118	-0.199	-0.142	-0.423**
	(0.298)	(0.145)	(0.159)	(0.096)	(0.132)	(0.095)	(0.211)
4 months post X Treat	0.411	0.118	-0.254	-0.0675	-0.315	-0.126	-0.578*
	(0.482)	(0.203)	(0.178)	(0.105)	(0.221)	(0.106)	(0.300)
5 months post X Treat	1.001	-0.755	-0.544	0.0269	-0.126	-0.146*	-0.447*
	(0.679)	(0.526)	(0.460)	(0.132)	(0.134)	(0.084)	(0.248)
6 months post X Treat	-0.106	-0.183	-0.127	0.0775	-0.297*	-0.0254	0.743
	(0.620)	(0.120)	(0.201)	(0.106)	(0.158)	(0.122)	(0.557)
7 months post X Treat	-0.462	-0.207*	-0.127	-0.0122	-0.127	-0.106	-0.109
	(0.500)	(0.117)	(0.170)	(0.068)	(0.151)	(0.104)	(0.397)
8 months post X Treat	0.477	-0.0918	-0.254	-0.0416	-0.184	-0.0539	-0.196
	(0.780)	(0.140)	(0.186)	(0.104)	(0.142)	(0.093)	(0.440)
9 months post X Treat	0.236	-0.0810	-0.171	0.129	-0.174	-0.138	-0.188
	(0.471)	(0.169)	(0.140)	(0.113)	(0.133)	(0.098)	(0.313)
10 months post X Treat	0.154	-0.198	-0.300*	-0.00503	-0.223	0.176	-0.220
	(0.486)	(0.243)	(0.158)	(0.094)	(0.184)	(0.201)	(0.189)
Obs	14974	14392	14328	14304	14291	14266	14369

Notes: Regressions include controls for month around intervention, randomization strata, and team fixed effects. Standard errors clustered at the team level. ***, **, * denotes significance at the 1, 5 and 10% levels respectively.

Table A.8: Leads and Lags Estimates of Impact of Evaluating Managers on Team-Level KPI

	Total	Product	Manage	Equipment	Quality	Safety
	(1)	(2)	(3)	(4)	(5)	(6)
4 months pre X Treat	0.344	0.395	0.323	1.329	-0.141	-2.232
	(1.785)	(1.399)	(1.003)	(1.030)	(1.332)	(3.068)
3 months pre X Treat	1.297	0.595	0.213	0.271	0.179	-0.138
	(1.519)	(1.161)	(0.801)	(1.016)	(0.947)	(2.781)
2 months pre X Treat	0.580	-0.0702	-0.477	0.970	1.144	0.318
	(1.342)	(0.819)	(0.370)	(0.984)	(0.772)	(0.864)
0 months post X Treat	0.532	2.007	1.287	2.283**	0.602	-1.940
	(2.010)	(1.374)	(1.225)	(0.935)	(1.560)	(1.622)
1 months post X Treat	0.743	1.050	1.454	1.797*	1.110	-1.890
	(1.882)	(1.340)	(1.205)	(1.006)	(1.427)	(1.537)
2 months post X Treat	0.0744	1.347	1.259	2.134**	0.308	-1.095
	(2.140)	(1.372)	(1.181)	(1.002)	(2.288)	(1.628)
3 months post X Treat	2.081	2.822*	1.033	2.160**	0.493	-0.686
	(2.918)	(1.542)	(1.315)	(0.848)	(2.544)	(1.762)
4 months post X Treat	-0.574	3.260**	1.351	1.923*	0.967	-0.390
	(6.115)	(1.471)	(1.674)	(1.038)	(3.456)	(2.611)
5 months post X Treat	2.322	3.635**	1.218	2.624**	-0.484	0.352
	(2.767)	(1.544)	(1.393)	(1.064)	(1.998)	(2.937)
6 months post X Treat	1.271	3.798**	0.904	1.983*	-0.843	1.021
	(2.512)	(1.532)	(1.424)	(1.071)	(1.770)	(2.590)
7 months post X Treat	3.457	3.869**	1.220	1.640*	1.159	0.546
	(2.836)	(1.618)	(1.340)	(0.953)	(2.127)	(2.774)
8 months post X Treat	1.546	3.260**	1.320	1.473	-0.495	0.512
	(2.225)	(1.553)	(1.351)	(0.989)	(1.753)	(2.766)
9 months post X Treat	1.419	2.651	1.612	2.059*	-0.582	1.436
	(2.262)	(1.708)	(1.330)	(1.049)	(1.672)	(2.600)
10 months post X Treat	2.987	3.864*	1.716	1.823*	-1.262	0.900
	(3.298)	(2.004)	(1.514)	(1.038)	(1.950)	(2.709)
Obs	923	905	922	351	684	923

Notes: Regressions include controls for month around intervention, randomization strata, KPI form version and team fixed effects. Standard errors clustered at the team level. ***, **, * denotes significance at the 1, 5 and 10% levels respectively.

Table A.9: Impact of Evaluating Managers on Individual Performance

	Total	Production	Safety	Quality	Attitude	Routine	Attendance
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treated	-0.135	0.0630	-0.0207	0.0601	-0.00310	-0.114*	0.00652
	(0.206)	(0.0709)	(0.0407)	(0.0433)	(0.0275)	(0.0578)	(0.0593)
Lag DepVar	0.562***	0.386***	0.191**	0.517***	0.321***	0.767***	0.314***
	(0.0467)	(0.0406)	(0.0794)	(0.0595)	(0.0524)	(0.0720)	(0.0243)
Obs	963	949	946	944	945	945	946

Notes: The data are winsorized at the top and bottom 1%. Regressions include controls for randomization strata and second treatment. Standard errors clustered at the team level. ***, **, * denotes significance at the 1, 5 and 10% levels respectively.

Table A.10: Summary Statistics for Characteristics of New Entrants at the Time of Hire

	Treatment Mean	Control Mean	p-value Diff	Treatment Obs.	Control Obs.
Age	25.606	24.972	0.669	33	36
Male	0.939	0.972	0.642	33	36
High School or More	0.636	0.528	0.293	33	36
Married	0.333	0.444	0.341	33	36
Referred to Company	0.545	0.611	0.607	33	36
Prior Experience (Years)	5.417	4.625	0.541	33	36
Grew Up Locally	0.727	0.750	0.860	33	36

Notes: Column 1 shows the mean value for post-intervention new entrants to treatment teams. Column 2 shows the mean value for post-intervention new entrants to control teams. Column 3 shows the p-value from a regression testing the statistical difference between the two groups.

Table A.11: Summary Statistics on Discussions about Intervention

	Mean	Std dev	N
Panel A: Workers			
Discussed with manager	.1357	.3427	685
Discussed with colleagues	.2272	.4193	682
Discussed with other teams	.0321	.1765	684
Hear other teams provide feedback (control)	.1101	.3132	690
Panel B: Managers			
Discussed with workers	.1818	.3901	44
Discussed with other managers	.2727	.4505	44
Discussed with their supervisors	.25	.4380	44

Table A.12: Heterogeneity by Treatment of Networks (Shares)

	Total KPI	Product KPI	Manage KPI	Life Satis.	Positive Mood
	(1)	(2)	(3)	(4)	(5)
Network Share Treated	0.190	-0.595	0.102	0.290*	0.247
	(0.396)	(0.378)	(0.198)	(0.166)	(0.607)
Treated X Network Share Treated	0.606	2.398***	0.176	-0.0421	0.0285
	(0.502)	(0.582)	(0.302)	(0.233)	(0.770)
Treated	1.255**	2.706***	0.693*	0.358**	0.643
	(0.586)	(0.813)	(0.403)	(0.138)	(0.417)
Lag DepVar	0.421***	0.748***	0.515***	0.293***	0.418***
	(0.0714)	(0.0695)	(0.0554)	(0.0346)	(0.0304)
Obs	1123	1098	1123	1126	1125

Notes: Regressions include controls for randomization strata, enumerators and second treatment. Standard errors clustered at the team level are shown in parentheses. ***, **, * denotes significance at the 1, 5 and 10% levels respectively.