Performance Feedback with Team Incentive:

A Field Experiment in Chinese Factories*

Karen Meng Li and Sang-Hyun Kim

School of Economics, Yonsei University, Seoul 03722, Korea

Abstract

This paper employs a field experiment to investigate in which information environment team-based incentives work better. The experiment was conducted in two spinning factories in Henan, China. We focus on workers who were doing the same individualistic task but still were paid according to team performances. For about three months, we have given three different types of performance feedback, baseline, intra-team, and inter-team feedbacks. We find that workers' productivity was highest with the intra-team feedback and lowest with the baseline feedback, which suggests that peer pressure and group status concern are of importance in making team incentives work.

Keywords: Relative performance feedback, Peer pressure, Group identity, Field experiment

JEL Classification: C93, D91, M52

---

* We are grateful to Syngjoo Choi, Subhasish M. Chowdhury, Chulyoung Kim, Duk Gyoo Kim, SeEun Jung, Euncheol Shin, and the participants of the Experimental Economics workshop at Seoul National University 2019 for their valuable comments. Corresponding author: Sang-Hyun Kim, sang.kim@yonsei.ac.kr

1. Introduction

Many modern organizations rely on teamwork. Moreover, despite the potential inefficiency due to the free-riding problem (Holmström, 1982), team-based incentive schemes are widely used.[1] One explanation for why team incentives work is peer pressure (e.g., Kandel and Lazear, 1992). By creating direct monetary externalities, team incentives may strengthen "empathy, loyalty, and guilt" among teammates. Another explanation is group identity and status concern (e.g., Tajfel and Turner, 1979; Akerlof and Kranton, 2005). The theory predicts that once an individual identifies oneself as a member of a group, he/she may devote himself/herself to achieving the goal and/or enhancing the group's status.[2] Team-based monetary incentives may improve the productivity of individuals by strengthening their identity as a team member.

It has been well documented in the literature that the strengths of these factors depend on whether and by whom an individual is observed. Peer pressure will be stronger when an individual is observed by his/her peers than when observed by non-peers or not observed at all (e.g., Azmat and Iriberri, 2016). On the other hand, if one is observed by a member of a competing group, group identity may become more salient (e.g., Sausgruber, 2009). Then, the following questions arise naturally: Do observations by peers or by competing group members indeed influence workers' productivity in the field? If so, whose observation can improve productivity more effectively when the monetary incentive is based on team performance?

This paper employs a field experiment to investigate in which information environment team-based incentives work better. The experiment was conducted in two spinning factories in Henan, China, where a team-based piece-rate wage was paid. Spinning is the process of twisting strands of fiber to issue yarn. The entire process is divided into six steps, and we focus on tenders whose main task is to connect yarns broken in the process of spinning as fast as possible to minimize the idling time of the machines.[3] The task does not require much skill

---

[1] Ledford et al. (1995) report that 70% of large firms used some form of team incentives. Moreover, "the use of various forms of team production (including methods such as quality circles) have increased in recent decades." (Lazear and Gibbs, 2009, p.204) See Lazear and Shaw (2007), Babcock et al. (2015), and Bandiera et al. (2013) for more discussions on the use of team in modern organizations.

[2] According to Tajfel and Turner (1979), social identity has three components: categorization, identification, and comparison. Categorization is the process of putting people into categories. Identification is the process by which we associate ourselves with certain groups. Comparison is the process by which we compare our group with other groups.

[3] We focus on these workers because (i) there were a number of them in relatively small factories, (ii)

but does require attention and effort. In each factory, there were three teams of tenders, each of which had three or four members. The task was highly individualistic, and there was neither complementarity nor the necessity to coordinate among the team members. The three teams in each factory worked in shifts, morning, afternoon, and evening, so they did not face the other teams when working. Since each worker was responsible for the working of predesignated machines in predesignated time, we could measure individual workers' performances by weighing the yarn produced in each machine after the shift.

For about three months, we have given three different types of performance feedback, baseline, intra-team, and inter-team feedbacks, via a mobile messenger. We have observed how the workers have responded to them. The baseline feedback contained only one's own performance information.[4] In the intra-team feedback, not only one's own performance but also the teammates' performance information was provided. In contrast, on top of the baseline feedback, the inter-team feedback showed the team-level performances as well, i.e., the performances of their own team and the other teams. Arguably, peer pressure would be stronger with the intra-team feedback than with the other feedbacks, and group identity would be more salient. Thus group status concern is stronger with the inter-team feedback.[5]

We find that workers' productivity was highest with the intra-team feedback and lowest with the baseline feedback. First of all, this finding has a practical implication for managers of organizations: The observation by teammates who share the monetary payoffs improves productivity more than the observation by the members of other teams. Thus, managers who plan to give team-based monetary incentives to workers may want to consider introducing some form of intra-team performance feedback as well.

On the more theoretical side, we carefully avoid drawing the general conclusion that peer pressure is more important than group status concern because (i) we believe that both peer pressure and status concern are multi-layered factors that are potentially interacting with each other and with other factors, and (ii) the impacts of the feedbacks depend on the number of

---

individual performances were easy to measure, and (iii) they were paid according to team performances. About the reason why team incentives were given, see footnote 13.

[4] We could have considered no feedback treatment as well, but decided not to, because given limited time, we wanted to focus on the inter-team and intra-team feedbacks, and the baseline feedback is more comparable to the other feedbacks than no feedback.

[5] See for example Azmat and Iriberri (2016) and Sausgruber (2009).

workers in each team.[6] We instead interpret the finding as evidence that both peer pressure and group status concern are essential in making team incentives work.

This paper is related to the vast literature on relative performance feedback. Performance feedback has long been regarded as an essential element of organization design (e.g., Alvero et al., 2001; Ford, 1980).[7] However, most previous studies have focused on individual feedback where the monetary incentive is also based on individual performance or not given at all. And, most studies on group performance feedback rely on lab experiments. An exception is Delfgaauw et al. (2013), who ran a field experiment in a Dutch retail chain to examine the effects of relative performance feedback to groups (i.e., inter-team feedback). More specifically, they introduced sales competitions among stores and found that it had a large impact on sales growth. In contrast, we consider both intra- and inter-team feedbacks and compare the effectiveness of them.

Another related literature is that on peer effect (for a review, see Herbst and Mas, 2015). Babcock et al. (2015) examine the effect of a "weakest link" type team incentive according to which no one in the team gets paid if any one of the members fails to achieve the individual goal such as attending at a study room or a gym. They report that productivity was 9-17% higher with the team incentive than with the individual incentive. Cognet et al. (2015) investigate the effects of monitoring by peers in a public good game with real effort tasks. They show that even a weak form of peer monitoring improved productivity up to the level with individual incentives.

The most closely related study is Gjedrem and Kvaløy (2020). In a lab experiment, they consider both individual and team relative performance feedbacks (that is, intra-team and inter-team feedbacks) and both individual and team-based monetary incentives. They show that there is a complementarity between team incentive and team feedback. When an incentive is based on team performance, (inter-)team feedback is more effective than individual (i.e., intra-team) feedback. This is in contrast with our main finding. It is not clear exactly where the difference comes from, but it may have something to do with the fact that their subjects remained

---

[6] See the discussion at the end of Section 2.

[7] Evidence on the effects of performance feedback is mixed. For example, Blanes i Vidal and Nossol (2011), Kuhnen and Tymula (2012) and Charness et al. (2014) find significant positive impacts of relative performance feedback, while Guryan et al. (2009), Eriksson et al. (2009) and Bellemare et al. (2010) find no significant effects. Barankay (2012) find that removing relative performance feedback improve productivity.

anonymous throughout the experiment. In contrast, in our experiment, the workers faced their teammates every day.[8] Thus, peer pressure must have been stronger in our experiment. To our knowledge, this is the first field experiment designed to evaluate the effects of different types of performance feedback in an environment where monetary incentives are based on team performances.

The rest of the paper is organized as follows. Section 2 provides a theoretical framework in which peer pressure and group status concern are modeled as the second-order beliefs directly introduced in the utility function. We formally present the three types of feedback and analyze the effects of them. In Section 3, we describe the details of the factories and the experiment. Section 4 reports the results. Finally, we conclude in Section 5.

2. Theoretical Framework

In this section, we develop a simple theoretical model to present our experimental design and predictions clearly. In particular, our model is inspired by the studies in psychological game theory (e.g., Geanakoplos et al., 1989; Charness and Dufwenberg, 2006; Battigalli and Dufwenberg, 2009) according to which players' motivations are directly influenced by others' beliefs and expectations. For instance, when you have some economic surplus which can be shared with someone else (like in the dictator game), you may wish to look generous and fair-minded, meaning that others' beliefs on your generosity influence your subjective welfare and motivation (e.g., Andreoni and Bernheim, 2009). Or, you may dislike disappointing others, in which case the others expect certain behavior from you (the first-order belief), and you hold an expectation about the others' expectation (the second-order belief) which you wish to live up to (e.g., Battigalli and Dufwenberg, 2007).

In our context, the reasons why workers concern others' beliefs can be categorized into two, depending on who the 'others' are. First, the beliefs of one's own teammates would influence her motivation to work diligently. Since the wage depends on the average performance of the team, its members may blame the below-average performers (i.e., peer pressure). And, a worker would definitely not want to be regarded or believed as the one who drags everybody's wage

[8] Blazovich (2013) shows that teammates' familiarity with one another influences the effectiveness of the compensation scheme.

down. Also, a worker may want to be recognized by peers as an industrious and reliable person.

Second, the beliefs of workers in the other teams may also matter for one's motivation. Teams may implicitly compete with each other in their performance (and wage) to be recognized as a good team (i.e., status concern at group level). As often observed in team-based sports games, competition among teams can strongly motivate team members to do their best. Moreover, when a team breaks up, and its members are reallocated to other teams, an individual worker from a good team may be more welcome by the incumbent members than a worker from a bad team. This is at least partly because the fact that an individual is from a good (or bad) team may signal her ability.

The list is not exhaustive, and we do not intend to distinguish one reason from another. Instead, our model incorporates the idea that workers concern the beliefs of their teammates and of those in the other teams while being agnostic about the details within each. We, for the sake of the presentation, label the concern about the teammates' belief as "peer pressure" and that about the other team members' belief as "group status concern." Formally, the utility of worker i in team j is given by:

$$u_{ij}(x_{ij}) = \underbrace{E(\bar{q}_j|\{x_{ij}\})}_{\text{monetary payoff}} - \underbrace{c(x_{ij})}_{\text{effort cost}} + \underbrace{\alpha E[E(q_{ij}|I_j)|\{x_{ij}\}]}_{\text{peer pressure}} + \underbrace{\beta E[E(\bar{q}_j|I_{-j})|\{x_{ij}\}]}_{\text{group status concern}}$$

where x_{ij} is worker i's effort level, $\bar{q}_j \equiv \left(\sum_k^{n_j} q_{kj}\right)/n_j$ is the average performance of team j, n_j is the number of workers in team j, and $c(\cdot)$ is the effort cost, which is a twice-continuously differentiable, strictly increasing, and strictly convex function. Furthermore, assume that $c'(0) = 0$ to ensure the existence of an interior solution. A worker's individual performance is the sum of her *intended* effort and noise:

$$q_{ij} = x_{ij} + \varepsilon_{ij}$$

where q_{ij} is worker i's individual performance, and ε_{ij} is a normal random variable with zero mean and σ^2 variance, i.e., $\varepsilon_{ij} \sim N(0, \sigma^2)$.[9] The noises are independent of each other. We interpret x_{ij} as worker i's *intended* level of effort which is unobservable to others, and ε_{ij} as everything else that can influence worker i's *realized* performance (e.g., health

[9] With the noise being a normal random variable, q_{ij} can be negative, which is unrealistic. However, the purpose of this assumption is to derive a simple closed form solution from which we can obtain a clear intuition. We strongly believe that the intuition will remain the same even if we relax this assumption.

condition, randomly changing working environment, and so on). Some factors in ε_{ij} may be completely out of control of worker i, but others are partially controllable by worker i, and thus may be attributable to worker i (e.g., one must take care of her own health condition). We assume that for this reason (and also because the effort x_{ij} is not directly observed), other workers evaluate (i.e., praise or blame) worker i by her performance q_{ij}.

Since the wage is proportional to the average performance of the team $\bar{q}_j$ (i.e., a piece-rate wage based on team performance), the first term, $E(\bar{q}_j|\{x_{ij}\})$, stands for the monetary payoff which is uncertain to the worker when making the effort decision. The third term, $E[E(q_{ij}|I_j)|\{x_{ij}\}]$, is the second-order belief on the individual performance q_{ij}. In particular, I_j is the information set available to the members of team j, and thus it is the worker's belief on how her teammates would evaluate her individual performance, and α is the relative importance of this factor. So, this utility term captures the idea that worker i will feel good (bad) if her teammates believe that she did well (poorly).[10] On the other hand, the fourth term, $E[E(\bar{q}_j|I_{-j})|\{x_{ij}\}]$, is the second-order belief on the team performance $\bar{q}_j$. More specifically, it is the worker's belief on how the other teams would evaluate her team's performance, and β is the relative importance of this factor. I_{-j} is the information given to the other teams. Strictly positive β means that worker i in team j will feel good (bad) if workers in the other teams believe that her team did well (poorly).

The information sets I_j and I_{-j} differ across the treatments. Specifically, the members of team j learn their own team-performance $\bar{q}_j$ in all treatments, whereas worker i's individual performance q_{ij} is learned only in the intra-team feedback treatment (T1). On the other hand, those in the teams other than j learn $\bar{q}_j$ only in the inter-team feedback treatment (T2). Formally, the information sets are given as follows:

$$I_j = \begin{cases} \{\bar{q}_j\}, & \text{in C and T2} \\ \{\bar{q}_j, q_{ij}\}, & \text{in T1} \end{cases}$$

[10] It may be the case in reality that people care more about the ranks than the absolute values of q's. However, even if so, the intuition would be the same. Any increasing function of q would lead to qualitatively similar results.

$$I_{-j} = \begin{cases} \emptyset, & \text{in C and T1} \\ \{\bar{q}_j\}, & \text{in T2} \end{cases}$$

Note that for simplicity, here we present the information related to worker i and team j without listing up the information of other workers and teams. However, in the experiment, the environment was symmetric in the sense that in T1, all workers knew their team performance and teammates' individual performances, whereas in T2, all workers knew their team performance and the other teams' performances. Everybody knew their own individual performance and their team performance in all treatments.[11]

Given these assumptions and the assumption that the beliefs are consistent with the optimal behavior of workers as in sequential equilibrium,[12] below we derive the optimal effort with each feedback.

(1) Baseline feedback (C)

With the baseline feedback, the utility of worker i in team j is:

$$u_{ij}(x_{ij}) = E(\bar{q}_j|\{x_{ij}\}) - c(x_{ij}) + \alpha E\left[E\left(q_{ij}|\{\bar{q}_j\}\right)|\{x_{ij}\}\right] + \beta E\left[E\left(\bar{q}_j|\emptyset\right)|\{x_{ij}\}\right]$$

Note first that $E\left(q_{ij}|\{\bar{q}_j\}\right) = \bar{q}_j$.[13] Intuitively, it means the following: Suppose that you observe that the average performance of the team is 100. Then, in a symmetric environment, it is more reasonable to believe that everybody's performance is 100 than to believe that one worker's performance is 200 while the others' are 50. Note that $E\left(\bar{q}_j|\emptyset\right)$ is not a function of x_{ij}, meaning that the last term (group status concern) is regarded as constant when maximizing the utility. Because $\bar{q}_j = \left(\sum_k^{n_j} q_{kj}\right)/n_j$, the first-order condition is:

[11] Because in the experiment workers knew their own individual performance on top of the average performance of their team, the actual inference problem by workers was slightly more complicated than the one presented here. However, even if we take into account this, the qualitative results derived here do not change.

[12] In sequential equilibrium, the beliefs are consistent with the strategies, and the strategies are optimal given the beliefs (see e.g., Osborne and Rubinstein, 1994). Here, the belief does not have to be consistent with *the* strategy. All we need is the assumption that the other workers believe that $x_{ij} = x^*$ for some known value x^*.

[13] The derivation is presented in Appendix A.

$$\frac{1+\alpha}{n_j} = c'\left(x_{ij}^C\right)$$

The left-hand side is the marginal benefit of an additional unit of effort, and the right is the marginal cost. Note that the equilibrium effort increases in α. Intuitively, the more you care about your teammates' belief, the harder you will work.

(2) Intra-team feedback (T1)

With the intra-team feedback, the utility of worker i is:

$$u_{ij}(x_{ij}) = E(\bar{q}_j|\{x_{ij}\}) - c(x_{ij}) + \alpha E[E(q_{ij}|\{\bar{q}_j, q_{ij}\})|\{x_{ij}\}] + \beta E[E(\bar{q}_j|\emptyset)|\{x_{ij}\}]$$

As in the baseline feedback case, the last term (i.e., group status concern) does not play any role here. Since $E(q_{ij}|\{\bar{q}_j, q_{ij}\}) = q_{ij}$, the first-order condition is:

$$\frac{1}{n_j} + \alpha = c'\left(x_{ij}^{T1}\right)$$

Note that the worker's motivation is stronger in T1 than in C (i.e., $x_{ij}^{T1} > x_{ij}^{C}$). To understand it intuitively, suppose that $\bar{q}_j$ increases by one unit. In such a case, an observer will not think that it is entirely due to one worker's additional contribution (i.e., due to that q_{ij} increases by n_j units) because it is equally likely that someone else contributes to the increase. Thus, when only the team performance $\bar{q}_j$ is observed as in C, only $1/n_j$ of worker i's effort will be attributed to worker i. In contrast, when q_{ij} is directly observed, worker i can show her contribution more clearly (i.e., any additional contribution will be fully appreciated), and thus the motivation to work is stronger in T1.

(3) Inter-team feedback (T2)

The last term in the utility function $E[E(\bar{q}_j|I_{-j})|\{x_{ij}\}]$ (i.e., group status concern) does not play any role in C and T1, which is not the case in T2. Specifically, with the inter-team feedback, $E[E(\bar{q}_j|I_{-j})|\{x_{ij}\}] = E[\bar{q}_j|\{x_{ij}\}]$. So, the utility can be written as

$$u_{ij}(x_{ij}) = E(\bar{q}_j|\{x_{ij}\}) - c(x_{ij}) + \alpha E[\bar{q}_j|\{x_{ij}\}] + \beta E[\bar{q}_j|\{x_{ij}\}].$$

And, the first-order condition is

$$\frac{1+\alpha+\beta}{n_j} = c'\left(x_{ij}^{T2}\right)$$

Not too surprisingly, this shows that the equilibrium effort is higher in T2 than in C (i.e., $x_{ij}^{T2} > x_{ij}^{C}$). In other words, when team j's performance is revealed to the other teams, the workers in team j will work harder than when not revealed.

However, it is not clear whether workers are more or less motivated in T1 than in T2. All we can say is that if α and/or n_j are sufficiently large in comparison with β, the workers will work harder in T1 than in T2. More precisely, the workers' motivations will be stronger in T1 than in T2 (i.e., $x_{ij}^{T1} > x_{ij}^{T2}$) if

$$\alpha\left(1-\frac{1}{n_j}\right) > \beta\frac{1}{n_j}.$$

An interesting and practically important implication of this result is that the effect of the intra-team feedback would not die out even if the team gets larger (i.e., n_j increases), whereas the effect of the inter-team feedback would.[14] Therefore, if the team is large enough, the intra-team feedback would be more effective than the inter-team feedback.[15]

3. Experiment

3.1. Context

The experiment was conducted in two spinning factories in Henan, China, which are about 30 kilometers apart. Spinning is the process of twisting strands of (cotton or polyester) fiber to issue yarn. The factories mainly produce 40s yarn and occasionally produce 32s yarn. 40s yarn is thinner than 32s yarn, and thus takes a longer time to produce the same weight of the product. It is the industry standard to convert one unit of 40s yarn into 0.8 units of 32s yarn. Thus, in

[14] We could not incorporate this idea to our experimental design for logistic reasons.

[15] It is conceivable that intra-team dynamics may be complicated by inter-team competition, which our model does not take into consideration. Many issues in team production including this hypothesis remain to be investigated.

recording performances and paying wages, this conversion rule was used.

The entire production process can be divided into six steps: opening, carding, drawing, roving, spinning, and packaging. Among these, spinning is the most critical process, determining the speed of the entire production process. Two types of workers are involved in the spinning process: tenders and doffers. For the study, we focus on tenders whose main task is to connect yarns broken in the process of spinning as fast as possible to minimize the idling time of the machines. The task does not require much skill but does require attention and effort.[16]

In each factory, there were three teams of tenders, each of which had three or four members, which means that nine or twelve tenders were working in a factory in a day. The three teams in each factory worked in shifts, morning, afternoon, and evening. In factory A, three tenders in a shift were responsible for 12 spinning machines, so each worker was responsible for four machines (1,920 spindles). On the other hand, in factory B, four tenders were responsible for 24 machines, so each was responsible for six machines (2,592 spindles). The task was highly individualistic, and there was neither complementarity nor the necessity to coordinate among the team members. But still, a team-based piece-rate wage was paid in both factories.[17] More precisely, in factory A, the team would get 1.17 Chinese yuan (CNY) per kilogram of product. Each team member got one-third of the total remuneration: 0.39 yuan (approximately 5 cents in USD). In factory B, on the other hand, the team got 1.33 yuan per kilogram of product, and each member got a quarter of the total remuneration, 0.33 yuan. Since each worker was responsible for the working of predesignated machines in predesignated time, we could measure individual workers' performances by weighing the yarn produced in each machine after the shift.[18]

---

[16] We focus on these workers because (i) there were a number of them in relatively small factories, (ii) their individual performances were easy to measure, and (iii) they were paid according to team performances.

[17] The reason is not entirely clear, but we think that the team-based incentive scheme is used because it is more convenient (or less cumbersome) to implement than an individualized incentive scheme. Also, the managers and the workers seemed to believe that each individual worker’s contribution to the team was more or less homogeneous, so they might believe that there was no reason to differentiate the compensation.

[18] The machines and time that a worker was responsible for were shifted regularly. In factory A, every four days evening team became afternoon team, and afternoon team became morning team. When changing the working time, they also changed the machines they were responsible for. Similarly, in factory B, the workers changed the machines and time every seven days.

It may also be worth mentioning that the performance measure that factory A used to determine the wage was noisier than that of factory B. Precisely, in factory B, the performance was measured by the weight of the product. In contrast, it was measured by how long the machine had worked in factory A. Thus, in factory A, even if the machine did not produce anything because the yarn was broken, as long as the spindles were spinning, it counted.[19] This means that the workers' incentive to connect the broken yarn as fast as possible might be weaker in factory A than in factory B. In other words, the monetary incentive was only loosely linked to the teams' actual performance (i.e., a larger noise) in factory A, while it was strongly linked (i.e., a smaller noise) in factory B. However, for the experiment, we employed simple product weight as the performance measure in both factories and used it in the performance feedback.

3.2. Implementation

As described in Section 2, we consider three types of performance feedback, baseline (C), intra-team (T1), and inter-team (T2) feedbacks. Everyone got informed of their performance and their team performance in all treatments. On top of that, in the *intra-team* feedback, the teammates' individual performance information was provided. On the other hand, the *inter-team* feedback showed team-level performances. Performances were measured on a daily basis, and the feedbacks were given on the day after.[20]

Taking into account the constraints in the factories, we implemented the design as follows. First, we have let all the workers exposed to all three types of feedbacks in sequence (i.e., within-subject design). We did so mainly to increase the number of subjects in each treatment. Second, we had given one type of feedback for about a month and then moved on to the next feedback. In factory A, the order of the treatments was: C, T1, and lastly T2. In factory B, the order was: C, T2, and T1. Instead, we could have shortened the time interval and randomized the treatments (for example, every Sunday, we could have randomly determined the feedback of the week). We, however, decided not to randomize the treatment every week (or everyday) for the following reasons. The factories' managers did not like the idea of altering the working

---

[19] Thus, the "unproductive" time can be divided into two: the time when the spindles are spinning and the time when they stop. The workers in factory A had the incentives to minimize the latter (i.e., the machines' idling time), instead of the total unproductive time.

[20] See Appendix B for a sample feedback.

environment too frequently since it might disturb the workers. We also were concerned about the possibility that the workers would have been too conscious of the experiment if we changed the treatment too frequently. Moreover, if the treatment effects lasted for days, then the effects couldn't have been clearly identified with frequently changing treatments.[21]

Table 1. Specifics by factory

	Factory A	Factory B
# of workers	3 per team	4 per team
# of machines	12 of FA506 (480 spindles)	8 of FA506 (456 spindles) and 16 of FA502 (420 spindles)
Piece rate	CNY 0.39 per kg	CNY 0.33 per kg
Performance measure (for a wage)	The hours that the machines have worked	The weight of the product
Timeline (days of work)	Jan. 19 - 29: C (9 days) Jan. 30 - Mar. 14: No experiment Mar. 15 - 31: C (19 days) Apr. 1 - 24: T1 (24 days) Apr. 25 -May 24: T2 (27 days)	Mar. 23 - Apr. 25: C (35 days) Apr. 26 -May 27: T2 (32 days) May 28 - Jun. 28: T1 (30 days)

Before the experiment got started, the participants were told that (i) the performance feedback would be privately given on the day after via WeChat, one of the most popular messaging, social media, and mobile payment apps in China, (ii) the wage scheme would remain the same as before throughout the experiment, and (iii) they could opt out from the experiment any time they wanted. Although there was no direct monetary incentive to participate in the experiment, the workers were encouraged to do so by the managers who wished to figure out what the most effective performance feedback is.

The experiment began in factory A with the baseline feedback on Jan. 19, 2018. But we had to stop the experiment because the factories were closed for the Chinese New Year holidays

[21] We are aware of the weakness of our implementation, namely that the sequence of the treatments might matter. However, because the same type of feedback has been given daily for a relatively long time (about a month), we believe that the sequence effect would be minor. And, the results from the factories with different sequences of the treatments (see Figure 1 for instance) seem to suggest that the sequence effect was indeed minor.

and started again in mid-March. The final treatment was over onMay 24. The experiment in factory B began on Mar. 23 and ended on Jun. 28. The detailed timeline is presented in Table 1.

During the experiment, there have been turnovers. In total, 16 workers in factory A and 26 workers in factory B participated in two or more treatments. In the following section, we analyze the data of these workers.

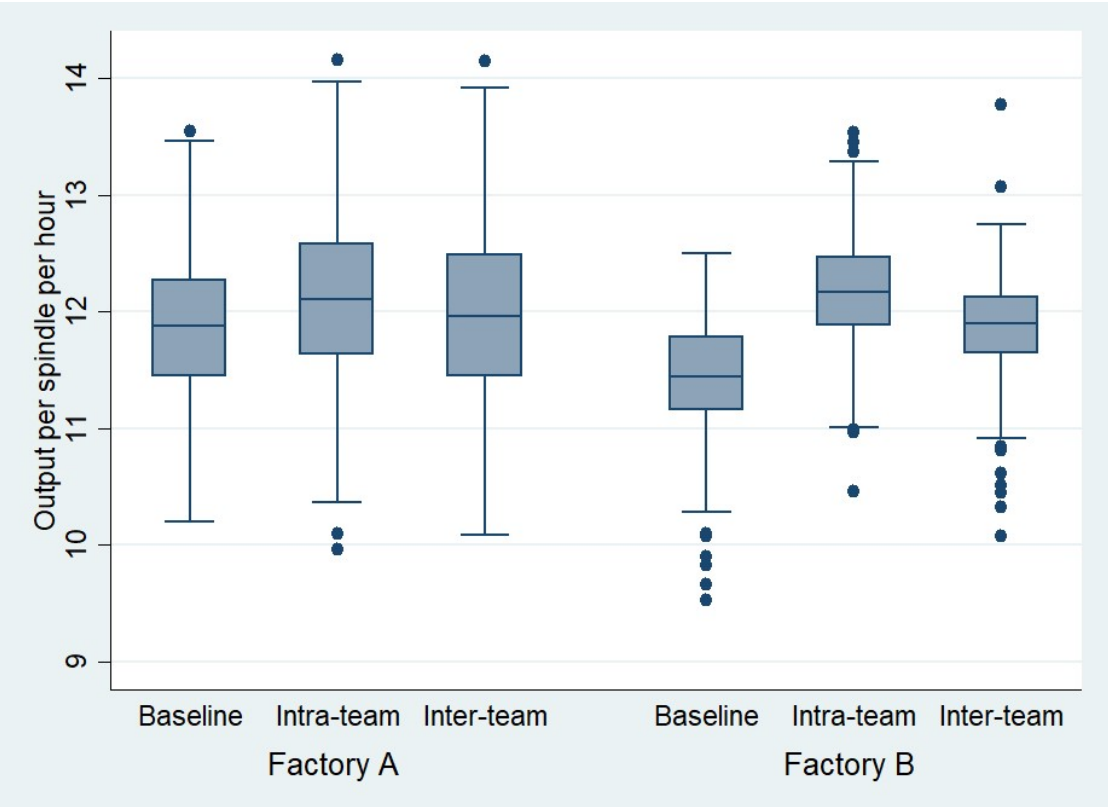

Figure 1. Output per spindle per hour

4. Result

Figure 1 summarizes the key results. The output is the weight (unit: gram) of the good produced in an hour in a spindle. We cannot directly compare the average productivities of the two factories because the number of workers per machine and the types of machines were different. But we can still compare the productivity differences within each factory. Observe first of all that the feedback had impacts on the productivity of workers. As predicted in Section 2, the average productivity was lowest with the baseline feedback. It was theoretically not clear whether T1 or T2 is more effective, and it turns out that T1 is. Note also that these patterns manifest in both factories despite the different sequences of the treatments,[22] which suggests that the sequence of the treatments would not overturn our results.

One can also see that the differences in productivity across treatments were more considerable in factory B than in factory A, and one may wonder whether the differences are significant in both or only one of the factories. Table 2 reports the results of two group t-test with equal variance. It turns out that the differences are all statistically significant in factory B, while in factory A, only the baseline feedback treatment was statistically different from the others.

Table 2. Comparison across treatments

		C	T1	T2	C vs. T1	C vs. T2	T1 vs. T2
	Mean	11.604	12.146	11.917	0.542***	0.313***	0.229***
All	Std. Dev.	(0.588)	(0.579)	(0.591)	(0.034)	(0.033)	(0.034)
	# of Obs.	632	564	620	1,196	1,252	1,184
	Mean	11.866	12.093	11.995	0.227***	0.129**	0.098
Factory A	Std. Dev.	(0.644)	(0.747)	(0.778)	(0.066)	(0.065)	(0.073)
	# of Obs.	245	203	239	448	484	442
	Mean	11.438	12.176	11.867	0.737***	0.429***	0.308***
Factory B	Std. Dev.	(0.482)	(0.458)	(0.429)	(0.034)	(0.033)	(0.033)
	# of Obs.	387	361	381	748	768	742

Since the ability of workers and other working environments might have affected their performances, we now move on to regression analysis in which we control for such variables. Table 3 reports the results of the regression analysis. The dependent variable is the weight of the good produced in an hour in a spindle. The key explanatory variables are the dummy variables for the feedbacks (*Intra-team* and *Inter-team*). The numbers in parenthesis are the t-statistics with the standard errors clustered at individual worker level.

In the first three columns, we control for the individual and date fixed effects. In comparison to the baseline feedback, the workers in factory A produced 0.197g more per hour per spindle (3.025kg more by a worker in a day) with the intra-team feedback, whereas those in factory B

[22] Recall that the sequence of the treatments was C, T1 and then T2 in factory A, whereas C, T2, and T1 in factory B.

produced 0.766g more (about 16kg more by a worker in a day).[23] With the inter-team feedback, factory A workers produced 0.11g more (1.69kg more by a worker in a day) than with the baseline feedback, and those in factory B produced 0.451g more (about 9.4kg more by a worker in a day). Comparing the results in Table 2 and Table 3, we can see that the treatment effects in factory B remain statistically significant while those in factory A become insignificant or only marginally significant. However, the sizes of the effects do not change much.

Table 3. Regression results

	All	Factory A	Factory B	All	Factory A	Factory B
Intra-team	0.532***	0.197*	0.766***	0.485***	0.169*	0.714***
	(6.68)	(1.99)	(22.58)	(6.07)	(1.79)	(21.20)
Inter-team	0.306***	0.11	0.451***	0.266***	0.11	0.398***
	(5.49)	(1.34)	(14.16)	(5.02)	(1.41)	(15.84)
Evening				0.118***	0.170***	0.08
				(3.19)	(3.62)	(1.64)
Afternoon				0.05	0.03	0.05
				(1.51)	(0.58)	(1.29)
# of spindles				0	0	-0.000**
				(-1.52)	(0.79)	(-2.55)
Refurbishment				0	0	0
				(-1.25)	(-0.54)	(-1.51)
# of workers				0.385***	0.29	0.284***
				(5.27)	(1.24)	(4.35)
32s yarn				0.06	0.365**	-0.03
				(0.86)	(2.52)	(-0.51)
Machine FE	No	No	No	Yes	Yes	Yes
Individual FE	Yes	Yes	Yes	Yes	Yes	Yes
Date FE	Yes	Yes	Yes	Yes	Yes	Yes
# of Obs.	1810	684	1126	1799	684	1115
R-squared	0.12	0.01	0.29	0.22	0.07	0.45

Note: The dependent variable is the weight of the good produced in an hour in a spindle. The numbers in parenthesis are the t-statistics with the standard errors clustered at individual level.

[23] In factory A, a worker was responsible for 1,920 spindles, whereas in factory B, a worker was in charge of about 2,600 spindles. And, a worker worked for 8 hours a day.

In the next three columns, we control for the shift (*Evening* and *Afternoon,* with morning being the baseline group), the number of spindles that the worker was in charge of (*# of spindles*), the length of time that the machines stopped for refurbishment (*Refurbishment*), the number of workers in the team (# of workers), whether 32s yarn was produced by the worker (*32s yarn*), and the machine fixed effects.[24]

Note first that the treatment effects remain significant, and their magnitudes are sizable as in the first three columns. It turns out that compared to the baseline feedback, the workers in factory A produced 0.169g more per hour per spindle (2.594kg more by a worker in a day) with the intra-team feedback, whereas those in factory B produced 0.714g more (about 15kg more by a worker in a day). With the inter-team feedback, the workers in factory A produced 0.11g more (1.69kg more by a worker in a day) than with the baseline feedback, and those in factory B produced 0.398g more (about 8.3kg more by a worker in a day).

Also, the average productivity was higher in the evening because there were fewer distractions, but the effect turns out to be significant only in factory A. The number of workers in a team turns out to be negatively related to productivity, which means that when there was an absent worker in a team, the others had to cover a greater number of machines, and the productivity from each spindle was lower.

5. Concluding Remarks

We experimentally investigate the impacts of three different kinds of performance feedbacks in two factories where a piece-rate wage based on team performance was paid. To our knowledge, this is the first field experiment that evaluates the impacts of the performance feedbacks when the monetary incentives are based on the performances of teams. The productivity of workers was highest when the feedback contained the information of individual performances of teammates and was lowest when it delivered only one's own individual performance information. Our result implies that managers who plan to give team-based monetary incentives to workers may want to consider introducing some form of intra-team

---

[24] The number of workers in a team was constant for most days, but occasionally there were absent workers due to illness or personal businesses. In such occasions, the number of spindles that a worker was in charge of was larger.

performance feedback as well. Also, it suggests that both peer pressure and group status concern play an important role in making team incentives work.

This study has a couple of limitations which invite further investigations. First, the implementation of the experiment was not ideal because of the constraints imposed in the field. For instance, we considered only two possible sequences of the treatments, but one may want to investigate whether or not the order of the treatments matters by considering other possible sequences. Or, one may conduct the experiment adopting between-subject design as opposed to within-subject design. Second, the theory predicts that the size of teams would alter the impacts of feedbacks, but we could not examine this prediction in our experiment. It would be interesting to explore this dimension in the lab or the field.

Lastly, the data seem to suggest that the impact of the additional information is especially strong when the monetary incentive is more strongly linked to the performance of the teams than when weakly linked. More specifically, the monetary incentive was only loosely linked to the true performance of the teams (i.e., a larger noise) in factory A, while it was strongly linked (i.e., a smaller noise) in factory B. We find that the impacts of the intra-team and the inter-team feedbacks were larger in the factory where the wage and the performance were more tightly linked. Further examinations on a grander scale would deepen our understanding of team production.

Reference

Akerlof, George A, and Rachel E Kranton. 2005. 'Identity and the Economics of Organizations', Journal of Economic Perspectives, 19: 9-32.

Alvero, Alicia M, Barbara R Bucklin, and John Austin. 2001. 'An objective review of the effectiveness and essential characteristics of performance feedback in organizational settings (1985-1998)', Journal of Organizational Behavior Management, 21: 3-29.

Andreoni, James, and B Douglas Bernheim. 2009. 'Social image and the 50–50 norm: A theoretical and experimental analysis of audience effects', Econometrica, 77: 1607-36.

Azmat, Ghazala, and Nagore Iriberri. 2016. 'The provision of relative performance feedback: An analysis of performance and satisfaction,' Journal of Economics & Management Strategy, 25: 77-110.

Babcock, Philip, Kelly Bedard, Gary Charness, John Hartman, and Heather Royer. 2015. 'Letting down the team? Social effects of team incentives', Journal of the European Economic Association, 13: 841-70.

Balcazar, Fabricio, Bill L Hopkins, and Yolanda Suarez. 1985. 'A critical, objective review of performance feedback,' Journal of Organizational Behavior Management, 7: 65-89.

Bandiera, Oriana, Iwan Barankay, and Imran Rasul. 2013. 'Team incentives: Evidence from a firm level experiment,' Journal of the European Economic Association, 11: 1079-114.

Barankay, Iwan. 2012. 'Rank incentives: Evidence from a randomized workplace experiment,' Working paper.

Battigalli, Pierpaolo, and Martin Dufwenberg. 2007. 'Guilt in games', American Economic Review, 97: 170-76.

———. 2009. 'Dynamic psychological games', Journal of Economic Theory, 144: 1-35.

Bellemare, Charles, Patrick Lepage, and Bruce Shearer. 2010. 'Peer pressure, incentives, and gender: An experimental analysis of motivation in the workplace,' Labour Economics, 17: 276-83.

Blanes i Vidal, Jordi, and Mareike Nossol. 2011. 'Tournaments without prizes: Evidence from personnel records,' Management Science, 57: 1721-36.

Blazovich, J. L. (2013). Team identity and performance-based compensation effects on performance. Team Performance Management: An International Journal.

Charness, Gary, and Martin Dufwenberg. 2006. 'Promises and Partnership, Econometrica, 74: 1579-1601.

Corgnet, Brice, Roberto Hernán-González, and Stephen Rassenti. 2013. 'Peer pressure and moral hazard in teams: Experimental evidence', Working paper.

Delfgaauw, Josse, Robert Dur, Joeri Sol, and Willem Verbeke. 2013. 'Tournament incentives in the field: Gender differences in the workplace,' Journal of Labor Economics, 31: 305-26.

Eriksson, Tor, Anders Poulsen, and Marie Claire Villeval. 2009. 'Feedback and incentives:

Experimental evidence,' Labour Economics, 16: 679-88.

Ford, John E. 1980. 'A classification system for feedback procedures,' Journal of Organizational Behavior Management, 2: 183-91.

Geanakoplos, John, David Pearce, and Ennio Stacchetti. 1989. 'Psychological games and sequential rationality,' Games and Economic Behavior, 1: 60-79.

Gjedrem, William, and Ola Kvaloy. 2020. 'Relative performance feedback to teams,' Labour Economics, 66: 101865.

Guryan, Jonathan, Kory Kroft, and Matthew J Notowidigdo. 2009. 'Peer effects in the workplace: Evidence from random groupings in professional golf tournaments,' American Economic Journal: Applied Economics, 1: 34-68.

Herbst, Daniel, and Alexandre Mas. 2015. 'Peer effects on worker output in the laboratory generalize to the field,' Science, 350: 545-49.

Holmstrom, Bengt. 1982. 'Moral hazard in teams,' The Bell Journal of Economics: 324-40.

Kandel, Eugene, and Edward P Lazear. 1992. 'Peer pressure and partnerships', Journal of Political Economy, 100: 801-17.

Kuhnen, Camelia M, and Agnieszka Tymula. 2012. 'Feedback, self-esteem, and performance in organizations', Management Science, 58: 94-113.

Lazear, Edward P, and Michael Gibbs. 2009. Personnel Economics in Practice, John Wiley &Sons, NJ.

Lazear, Edward P, and Kathryn L Shaw. 2007. 'Personnel economics: The economist's view of human resources', Journal of Economic Perspectives, 21: 91-114.

Ledford Jr, Gerald E. 1995. 'Paying for the skills, knowledge, and competencies of knowledge workers', Compensation & Benefits Review, 27: 55-62.

Osborne, Martin J, and Ariel Rubinstein. 1994. A course in game theory, MIT Press, MA.

Sausgruber, Rupert. 2009. 'A note on peer effects between teams,' Experimental Economics, 12: 193-201.

Tajfel, Henri, and John C. Turner. 1979. ‘An integrative theory of intergroup conflict,’ in W. G. Austin, and S. Worchel, (Eds) Organizational identity: A reader, 56-65.

Appendix A. Proof of $E\left(q_{ij}\middle|\{\bar{q}_j\}\right) = \bar{q}_j$.

Let us first assume that the belief is consistent with the equilibrium strategy x^*.[25] This implies that the others believe that $q_{ij} = x^* + \varepsilon_{ij}$ for all i and j. The inference problem boils down to forming the expectation of ε_{ij} conditional on the average performance $\bar{q}_j$. The average performance in equilibrium is:

$$\bar{q}_j = \frac{\sum_k (x_{kj} + \varepsilon_{kj})}{n_j} = x^* + \frac{\sum_k \varepsilon_{kj}}{n_j}.$$

Thus,

$$\varepsilon_{ij} + \sum_{k \neq i} \varepsilon_{kj} = \varepsilon_{ij} + v_{ij} = n_j(\bar{q}_j - x^*).$$

where $v_{ij} \equiv \sum_{k \neq i} \varepsilon_{kj}$, and $v_{ij} \sim N\left(0, (n_j - 1)\sigma^2\right)$. According to the above equation, the observation $n_j(\bar{q}_j - x^*)$ can be divided into two parts, the signal ε_{ij} and the noise v_{ij}. Because both ε_{ij} and v_{ij} are normal random variables, the conditional expectation of ε_{ij} is the weighted average of the unconditional mean (i.e., zero) and the new observation (i.e., $n_j(\bar{q}_j - x^*)$) where the weights are given by the precisions (i.e., the inverse of variance) of each information:

$$E\left(\varepsilon_{ij}\middle|\{\bar{q}_j\}\right) = \frac{\frac{1}{\sigma^2} \times 0 + \frac{1}{(n_j - 1)\sigma^2} \times n_j(\bar{q}_j - x^*)}{\frac{1}{\sigma^2} + \frac{1}{(n_j - 1)\sigma^2}} = \bar{q}_j - x^*$$

Therefore, $E\left(q_{ij}\middle|\{\bar{q}_j\}\right) = E\left(x^* + \varepsilon_{ij}\middle|\{\bar{q}_j\}\right) = \bar{q}_j$.

---

[25] Here the belief does not have to be consistent with *the* strategy. All we need is the assumption that the other workers believe that $x_{ij} = x^*$ for some known value x^*.

Appendix B. A Sample Feedback

[Common part]

Greetings!

Yesterday, machines 1 to 4 [the machines covered by the recipient of this feedback] produced ___ kg of 40s yarn and ___ kg of 32s yarn. The average output per machine was ___ kg, and the weight of the wasted material was ___ kg.

The total output produced by your team was ___ kg of 40s yarn and ___ kg of 32s yarn. The total output produced in the factory was ___ kg of 40s yarn and ___ kg of 32s yarn.

Our goal is to produce 12.3 tons per week in this factory, and the daily goal per team is 595 kg. To achieve this goal, one person should produce 198 kg per day and 6.188 kg per hour per machine. If we achieve the goal, each worker will get about 2,350 yuan in a month.

Contact information: XXXXXXX

WECHAT: YYYYYYY

[Intra-team feedback only]

Ranking	[Workers'] Name	Average Production Per Hour (kg)
1	AAA	X
2	BBB	Y
3	CCC	Z

[A bar diagram depicting the average production]

[Inter-team feedback only]

Ranking	Name	Team Production Per Day (kg)
1	Team A	XX
2	Team B	YY
3	Team C	ZZ

[A bar diagram depicting the team production]