Peer Effects and Observability: Theory and Evidence from a Field Experiment

C. Simon Fan, Xiangdong Wei, and Junsen Zhang

December 2014

Abstract

Observability is crucial for the effectiveness of “peer effects”, particularly when workers engage in multiple tasks. This paper presents a theoretical and empirical investigation into the peer effects with multiple dimensions of outputs (quality and quantity), which have differential observability. The model predicts that workers will pivot their efforts towards the dimension with more observability due to peer pressure, and shows that under reasonable conditions, an increase in the observability of quality will lead workers to produce higher quality, while reducing the quantity of output. We test this theoretical implication with the data from a field experiment in China. Our empirical analysis shows that in team production, a worker’s effort devoted to quantity and quality production does respond to the changing observability of these two dimensions of a job. Such a finding demonstrates not only the existence of peer effects but also the strategic responses to differential peer pressures from co-workers. Our further experiment of switching workers from team-based to individual-based incentive pay yields two other findings. First, the “social norms” generated in the period of team work persisted into the period of piece rates, particularly its early period. Second, workers exhibited higher productivity under team work than piece rates, which implies that peer pressure is an effective tool to mitigate the free-rider problem of team production.

JEL classifications:

# Lingnan University, Hong Kong
§ Corresponding Author: Junsen Zhang, Department of Economics, the Chinese University of Hong Kong, Hong Kong. E-mail: jszhang@cuhk.edu.hk
1. Introduction

Motivated by the seminal contribution of Kandel and Lazear (1992), a large number of empirical studies examined the significance of “peer effects” at work, and most of them show that a positive peer effect mitigates the agency problem in firms.¹ For example, Hamilton, Nickerson, and Owan (2003) demonstrate that for a garment plant in the US, average productivity increased by 14 percent when the production shifted from individual piece rate to group piece rate production.²

Hamilton, Nickerson, and Owan (2003) emphasize that a key element to their finding is the ease of peer monitoring within teams in the firm that they studied. However, somewhat surprisingly, the issue of observability has seldom been examined in the literature of peer effects. This negligence is possibly due to that the existing literature has mostly focused on the study of only one dimension of output --- quantity, which is usually easy to observe by peers.

When we consider the nature of multiple tasks for most jobs, observability can be a serious problem of peer monitoring. For example, Holmstrom and Milgrom (1991) emphasize that quality is usually less observable than quantity of output. Moreover, while observability may be an issue in all aspects of managerial economics, it is particularly important for peer monitoring. A manager has an authority to request to observe workers in many aspects, but peers do not have such an authority. Thus, observability is crucial for the study of “peer effects”.³

This paper extends the literature of peer effects by considering that a job often has multiple tasks, which have differential observability. It then derives, based on the received literature, a straightforward implication of this consideration that peer effects differ across different tasks. While peer effect may exist in the job aspect that is observable to coworkers, it should be greater in the aspect with greater observability.

¹ A partial list of this literature includes Hamilton, Nickerson, and Owan (2003), Falk and Ichino (2006), Mas and Moretti (2009), Bellemare, Lepage, and Shearer (2010), and the survey of Camerer and Malmendier (2007).
² Under such a scenario, profit sharing is a good incentive mechanism for firms. Moreover, since it is egalitarian in nature, it may also foster every worker’s morale (Fang and Moscarini, 2005).
³ Peer effect is not considered in Holmstrom and Milgrom (1991).
Furthermore, changing observability can change peer pressure and therefore relative performance of the different job tasks.

Kandel and Lazear (1992) argue that the peer effect includes both guilt and shame: a team may impose “shame” on a worker who exerts low effort, and a worker who exerts low effort may feel guilty toward co-workers even when his low effort is not observed. Based on sociologists’ research, Kandel and Lazear (1992, Page 806) state: “Sociologists sometimes distinguish guilt from shame. Guilt is internal pressure, whereas shame is external pressure. In the context of the firm, the important issue is observability. A worker feels shame when others can observe his actions. Without observability, only guilt can be an effective form of pressure. ... If others do not take the time to watch, then a worker feels no shame.”⁴ Therefore, a direct corollary of the received literature is that different dimensions of output with heterogeneous observability will have different combinations of shame and guilt for team members, which results in differential peer effects across various tasks of a job.

This paper first develops a model of peer effects by extending Kandel and Lazear (1992) into a setting with multiple tasks. Specifically, we consider that an output has two dimensions: quality and quantity.⁵ We assume that the quantity that an individual produces is more observable than the quality, which means that the peer pressure differs significantly across quality and quantity. If a worker produces less quantity, he is shamed in public and the peer pressure is large. If a worker produces with a lower quality, no one may observe and he only feels guilty privately, which generates much less peer pressure.

The model analyzes a worker’s allocation of efforts across tasks. A worker’s efforts consist of two dimensions, which increase the quantity and the quality of output, respectively. However, both efforts reduce a worker’s utility, and workers may feel exhausted from exerting such efforts. Our model shows that when the “exhaustion” effect from effort in both producing quality and quantity is large, and when the complementary

---

⁴ Also, Li, Rosenzweig and Zhang (2010) give a distinction between guilt and shame: guilt is something with harm to someone you care about; shame is something you feel bad about yourself but no direct harm to someone you care about.

⁵ This example is highlighted in Holmstrom and Milgrom (1991). Recently, Hong, Hossain, List and Tanaka (2013) tested the theory of multitasking with Chinese data, and find strong evidence of quantity-quality trade off.
effect of these two efforts in generating earnings is small, then an increase in the
obersevability of quality will lead a worker to produce less quantity and more quality.

We test this theoretical implication with the data we collected from a field
experiment in China. Our field experiment, which lasted for 20 working days, involves
hiring university students (summer interns) to carry out real jobs. During the whole
period, the workers (interns) were offered a job that was regarded as entirely normal ---
inputting survey questionnaires that were collected by us through another project, and
they were completely unaware that the patterns of their working behaviors would be
analyzed in research.

We conducted our experiment in the field rather than in a laboratory. A field
experiment is essential for the purpose of this paper, because only the setting with a
strong sense of “realism” can generate sufficient peer effects. Also, in order to identify
clean-cut peer effects with differential observability, we need to conduct our empirical
analysis in a company with new recruits. To achieve these purposes in the most cost
effective way, we set up a new internship scheme and commissioned a company to run
such a scheme in a major city, Shenzhen, in Guangdong Province of China. The
internship involves data input job of real survey questionnaires collected by us for
another project. The company we commissioned is an IT company in Shenzhen, whose
main business was software design and data processing. As part of our agreement, the
company also allows us to send two research assistants to work as “managers” of the
company to run this internship during the period of our experiments.

Our experiments lasted for only about one month. However, we need workers who
both took the jobs seriously and were happy with the short-term jobs. Fortunately, such
workers were readily available --- the university students looking for summer internships.
In fact, many Chinese universities today make internship experience a requirement for
graduation. The internship opportunity provided by our project was well appreciated by
the Student Service Centre of Shenzhen University, which helped us advertised it in its
intranet. All the interns were officially hired by the company, which also promised to
issue internship certificates with rating of their performance (excellent, good and about
average) to all interns after they finish the job.
We received over 200 applications for the internship job, from which we randomly selected 40 candidates. Then, through random draws, we divide them into 2 groups, each of which consists of 20 workers. The first group work in the morning from 8:30 to 12:00, while the second in the afternoon from 1:30 to 5:00. They worked from Monday to Friday, and each intern worked for 3.5 hours per day. For the first 10 days (two weeks), a team-based performance pay was used and worker pay is determined by a formula that increases with total quantity and decreases with average quality. Every day, a 5-minute job review was done by two managers (our RAs) on previous day’s work. For the first three days, individual quantity of output was made public along with the group mean quality of output. The information on individual quality of output was only known by the person himself/herself. For the next 7 days, the same was done for the morning team (control group) but both individual quantity and quality information were made public for the afternoon team (treatment group). So the morning team has observability on quantity of output only while the afternoon team has full observability on both quantity and quality of output.

Our empirical analysis yields the following findings. First, for the first three days when both groups were identically treated, there was no significant difference between them in terms of either the average quantity or the average quality produced. It suggests that the average productivities of the workers in both groups are largely the same, and hence the result of randomization in dividing the workers into 2 groups: the control group and the treatment group, are largely successful. Second, The control group and the treatment group began to differ in incentives since Day 4, when the morning group continues to learn individual quantity of output but NOT quality, whereas the afternoon group is told the both. We find clear evidence that for the rest of the seven days, the team of the afternoon produced less quantity of output with higher quality. This finding is in support of our theoretical predictions that peer pressures are indeed associated with observability and workers will respond to peer pressure strategically.

Our experiment of team production is followed by a further experiment of switching workers from team-based (teamwork) to individual-based (piece rate) incentive pay. This happens from day 11, and both groups were switched to individual-based performance
pay or piece rate using the same pay formula. This follow up experiment allows us to compare worker performance under teamwork to that under individual piece rate. The latter is supposed to have relatively high individual incentive (less moral hazard problem) effect but less peer effect. We find that the afternoon group continued to produce better quality and less quantity of output than the morning group after both groups were put under the same piece rate scheme.

We provide the following interpretations. Team production formed a social norm, which was related to the fact that one’s output affected others’ earnings. Under piece rates, such an “economic fundamental” changed substantially. However, this old “social norm” did not disappear entirely even when the material basis that generated this culture in the first place disappeared. In fact, in the first two days under the piece rates, we find that workers in both groups exhibited the same patterns of production as in the team production: There is no significant change in quantity and quality produced during these two days. Quality of output starts to degenerate afterwards but quantity of output follows almost the same trend to increase. Thus, peer effects under piece rate were “path-dependent”: the group of workers in the afternoon had a tradition of emphasizing “quality”, which leads to a greater peer effect on quality in this new environment. Thus, they continue to produce higher-quality and lower quantity under piece rates, although doing so lead them to receive a lower income and a lower level of narrowly-defined utility than otherwise.

Moreover, in comparison with the outputs between the two periods, we find that workers produced more in the former than the latter period. The regression analysis shows that the introduction of individual piece rate led to the deterioration of quality produced for the morning group whilst for the afternoon group both quality and quantity deteriorated. This result implies that the peer effect under team rate was sufficiently strong that it overcame the free-rider problem under the payment of group rate, which in turn reinforces the results of the existence of peer effects of team production.

In what follows, we present our theoretical model in Section 2. Section 3 explains our experimental design and provides brief data description. Our detailed empirical analyses for the two experimental periods are contained in Section 4 and 5, respectively.
Section 6 offers some further suggestive explanations to the findings in Section 5. We conclude in Section 7.

2. A Simple Model of Peer Effects in Multiple Tasks

This section presents a theoretical analysis of the peer effects with differential observability when individuals engage in the production with multiple tasks. Our basic analytical framework is based on of Kandel and Lazear (1992), which is described in the next sub-section.

2.1. Kandel-Lazear Model

Kandel and Lazear (1992) consider a framework in which output is produced by a group of identical workers with a simple sharing rule that every worker receives an equal share of the output. Then, without peer pressure, a worker’s objective function is:

$$\frac{F(e_1, e_2, \ldots, e_N)}{N} - C(e_i)$$

where $e_i$ is the effort level of worker $i$; $F()$ is the production function (or earning function); $N$ is the number of workers; $C()$ is the cost of effort.

The Nash equilibrium is then determined by the first order condition of (1), namely

$$\frac{F_i(e_1, e_2, \ldots, e_N)}{N} = C'(e)$$

In this setup, the total surplus of all individuals takes the form
\[ F(e_1, e_2, \ldots, e_N) - \sum_{i=1}^{N} C(e_i) \]  

(3)

Hence, the social optimum of the whole team is determined by

\[ F_i(e_1, e_2, \ldots, e_N) = C'(e_i) \]  

(4)

Kandel and Lazear (1992) show that under reasonable conditions, the Nash equilibrium as defined by (2) is smaller than the social optimum as defined by (4).

Then, Kandel and Lazear (1992) introduce the notion of “peer effect”, which can modify a worker’s objective function into the following form:

\[ \frac{F(e_1, e_2, \ldots, e_N)}{N} - C(e_i) - P(\bar{e} - e_i) \]  

(5)

where \( \bar{e} \) is the average effort of other members.

Kandel and Lazear (1992) show that the free-rider problem in team production can be mitigated or even solved by the consideration of the “peer effect”. To see this point. Note that in this case, the Nash equilibrium is determined by

\[ \frac{F_i(e_1, e_2, \ldots, e_N)}{N} = C'(e) - P'(0) \]  

(6)

Kandel and Lazear (1992) show that the Nash equilibrium as determined by (6) is greater than the Nash equilibrium as determined by (2). In fact, under some conditions, the Nash equilibrium as determined by (6) can be equal to or can be even greater than the social optimum as determined by (4).

2.2. Peer Effects in Multiple Tasks
In Kandel and Lazear (1992), output (and effort) is of only one dimension. This section extends their model by considering the case where output has 2 dimensions: quality as well as quantity. If working alone, an individual’s earning function is defined as follows:

$$ G(m, q) $$

where $m$ and $q$ are the quantity and the quality of output, respectively. We assume that $G(\ , \ )$ is strictly increasing and strictly concave with respect to both of its variables.

The quantity and the quality of output are determined by a worker’s effort in producing quantity ($e^a$) and his effort in producing quality ($e^b$), respectively. Thus, we define

$$ m = f(e^a) \\
q = h(e^b) $$

We assume that $f()$ and $h()$ are strictly increasing with respect to their variables, respectively.

In team production, we assume that the total outputs are the summations of the outputs produced by all workers. Therefore, the earnings function of a team of $N$ members is as follows:

$$ G(e^a_1, e^b_1) + G(e^a_2, e^b_2) + \ldots + G(e^a_N, e^b_N) $$

where $e^a_i$ is the effort of Worker $i$ in producing quantity and $e^b_i$ is the effort of Worker $i$ in producing quality, respectively.

With equal sharing rule and without peer pressure, a worker’s objective function is:
where $C()$ is the cost of effort, and we assume

$$C_1 = \frac{\partial C(e^a_i, e^b_i)}{\partial e^a_i} > 0, C_{11} = \frac{\partial^2 C(e^a_i, e^b_i)}{\partial (e^a_i)^2} > 0$$

$$C_2 = \frac{\partial C(e^a_i, e^b_i)}{\partial e^b_i} > 0, C_{22} = \frac{\partial^2 C(e^a_i, e^b_i)}{\partial (e^b_i)^2} > 0$$

$$C_{12} = \frac{\partial^2 C(e^a_i, e^b_i)}{\partial e^a_i \partial e^b_i} > 0$$

The Nash equilibrium is determined by

$$\frac{G_i[f(e^a_i), h(e^b_i)]}{N} f'(e^a_i) = C_i(e^a_i, e^b_i)$$

$$i=1,2,\ldots N;$$

$$\frac{G_i[f(e^a_i), h(e^b_i)]}{N} g'(e^a_i) = C_i(e^a_i, e^b_i)$$

$$i=1,2,\ldots N.$$

Since individuals are all identical, we might as well focus on the equilibrium in which every individual plays the same strategy (the same effort).

Now we introduce peer effect into this framework in the spirit of Kandel and Lazear (1992). Kandel and Lazear (1992) argue that the peer effect includes both guilt and shame: the team may impose “shame” on a worker who exerts low effort, and a worker who exerts low effort may feel guilty toward co-workers even when his low effort is not observed.
Since quality is much less observable than quantity, the peer pressure differs significantly between them. If a worker produces less quantity, he is shamed in public and the peer pressure is large. If a worker produces with a lower quantity, no one may observe that and he only feels guilty privately, which generates much less peer pressure.

Specifically, we use the same notation in Kandel and Lazear (1992) to denote the peer pressure from quantity comparison, namely

\[ P(\bar{m} - m) \]  

where \( m \) and \( \bar{m} \) is the quantity produced by an individual worker and the average quantity produced by the group, respectively.

Note that there is a small difference between the formulation in (14) and the formulation of Kandel and Lazear (1992) in modeling the “peer effect”. In Kandel and Lazear (1992), the peer effect is captured by the difference between a worker’s effort and the average effort of the team; in (12) it is the difference between a worker’s output of quantity and the average quantity of the team. However, since there is a one-to-one mapping between a worker’s output of quantity and his effort in producing the quantity, the formulations are essentially identical. We choose such a formulation because it appears to be more intuitive that a worker observes his teammates’ outputs rather than efforts directly.

On the other hand, the peer pressure from (secret) quality comparison is defined as

\[ \lambda P(\bar{q} - q) \]  

where \( q \) and \( \bar{q} \) is the quality produced by an individual worker and the average quality produced by the group, respectively. \( \lambda \) is a non-negative coefficient. We assume that the quantity that an individual produces is much more observable than the quality, which
implies that the peer pressure on “quantity” results in both shame and guilt, while the peer pressure on “quality” results in guilt only. Therefore, we expect that

\[ \lambda < 1 \] (16)

In particular, if shame is much greater than guilt, then \( \lambda \) is much smaller than one.

In this case, a worker’s objective function is:

\[
\sum_{i=1}^{N} G(e_i^a, e_i^b) - C(e_i^a, e_i^b) - P(m-m) - \lambda P(\bar{q} - q)
\] (17)

The Nash equilibrium is determined by

\[
\frac{f'(e_i^a) G_1(m,q)}{N} = C_1(e_i^a, e_i^b) - \frac{f'(e_i^a) P^* (m-m) - \lambda P(\bar{q} - q)}{N} (18)
\]
i=1,2,…N; and

\[
\frac{q'(e_i^b) G_2(e_i^a, e_i^b)}{N} = C_2(e_i^a, e_i^b) - P(\bar{m} - m) - \lambda q'(e_i^b) P^* (\bar{q} - q)
\] (19)
i=1,2,…N;

We now discuss the solutions to Equations (18) and (19). Since individuals are all identical, we might as well focus on the equilibrium in which every individual plays the same strategy (the same effort).

To save algebra, we might as well assume

\[ P(0) = 0 \] (20)

From the discussions in the last section we can see that it is a reasonable assumption. Then, (18) and (19) are reduced to

\[
\frac{f'(e_i^a) G_1(m,q)}{N} = C_1(e_i^a, e_i^b) - f'(e_i^a) P^* (0)
\] (21)
and
\[
\frac{h'(e_i^b)G_2(m,q)}{N} = C_2(e_i^a,e_i^b) - \lambda h'(e_i^b)P'(0)
\]  

(22)

Note that as long as peer effect exists, we will have

\[P'(0) > 0\]  

(23)

In fact, as emphasized by Akerlof (1980), an individual’s any small deviation from a social norm will often result in that the individual will incur a major stigma, which implies that \(P'(0)\) is often large in magnitude. Then, we have the following proposition.

**Proposition 1.**

(1) \(\frac{de^a}{d\lambda} < 0\) if

\[C_{12}(e^a,e^b) > \frac{G_{12}f'}{N}\]  

(24)

(2) \(\frac{de^b}{d\lambda} > 0\)

**Proof:** See Appendix.

Note that Condition (24) will be satisfied if \(C_{12}(e^a,e^b)\) and \(N\) are sufficiently large, and if \(G_{12}(e^a,e^b)\) is sufficiently small. The intuitions can be explained as follows.

First, note that

\[C_{12}(e^a,e^b) = \frac{d[C_i(e^a,e^b)]}{de^b}\]

Thus, that \(C_{12}(e^a,e^b)\) is large means that an increase in the effort of producing quantity leads to a large increase in the marginal disutility of the effort of producing quality. In other words, the “exhaustion” effect from effort in both producing quality and quantity is large.
Second, that $N$ is sufficiently large means that the size of the team is large, which leads to little pecuniary gain from increasing one’s own output.

Third, $G_{12}$ represents the complementary effect of the quantity and the quality of outputs in generating earnings. Thus, that $G_{12}$ is small means that spending more effort in producing quantity does not increase much the marginal productivity of the effort in quality.

In sum, Proposition 1 shows that when the “exhaustion” effect from effort in both producing quality and quantity is large, and when the complementary effect of these two efforts in generating earnings is small, then an increase in the observability of quality will cause a worker to produce less quantity and more quality.

Moreover, recall that the quality produced increases with $e^b$, and the quantity produced increases with $e^a$. Then, from Proposition 1, obviously we have the following corollary.

**Corollary 1.** When $\lambda$ increases, the quality produced will increase, and the quantity produced will decrease if (22) is satisfied.

In Sections 3 and 4, we will use the data collected from a field experiment to test Corollary 1.

### 2.3. Peer Effect in Piece Rate with Multiple Tasks

In the existing literature, peer effects are generally discussed in the context of team production and the payment of team rate. In other words, an implication assumption is that peer effects do not exist when workers are under the contract of individual piece rates. While we agree that peer effects tend to be much stronger when the payment to each individual worker depends on the performance of all members of a group, peer pressures usually exist as long as there are human interactions, including the case of individual piece rates. For example, Andreoni and Bernheim (2009) postulate that many
people care much about their “social image”. According to this argument, for example, there may be peer pressure under piece rates to show off one’s ability. Also, peer effects, as any other social norms, are “path-dependent”. In particular relevance to the current paper, the peer pressure under individual piece rates may depend on the social norms (or peer effects) generated by the previous ways of production.

Specifically, we define a worker’s objective function in this case as

$$G(e^i_a, e^i_b) - C(e^o_a, e^o_b) - Q(m - m_i) - \eta Q(q - q_i)$$

(25)

Where \( Q(\cdot) \) is the function of peer pressure under the contract of piece rate, which is the counterpart of \( P(\cdot) \) is the function of peer pressure under the contract of team production. \( \eta \) is a non-negative coefficient, which measures the relative weight of peer pressure between “quantity” and “quality” of production.

As a prior, the peer pressure from producing low quality under the contract of piece rate should be lower than that under team production. In team production, if a worker produces low-quality, he will feel guilty even if the quality of individual workers is unobservable. Under piece rate, such a sense of guilt no longer exists since to the co-workers, the low-quality output that one produces does not affect his co-workers’ earnings. Of course, a worker may still feel guilty to the firm for producing low-quality. After all, while a worker is penalized financially for producing low quality, such a penalty is not sufficient to cover the cost to the firm. Thus, the peer pressure from producing low-quality goods may continue to exist under piece rate. In contrast, an individual may get a significant stigma from being regarded as low-ability due to a low amount of quantity produced, which motivates him to concern about the quantity of output. Therefore, \( \eta \) may be a relatively small number under individual piece rates.

Then, by a logic that is identical to the last subsection, we have the following proposition.6

---

6 A rigorous proof is provided in the working paper version of this paper.
Proposition 2. (1) \( \frac{de^a}{d\eta} < 0 \) if

\[ C_{12}(e^a, e^b) > G_{12}f'h' \]  \hspace{1cm} (26)

(2) \( \frac{de^b}{d\eta} > 0 \)

The intuition of the Condition (26) is similar to that of (24). As we will see, this proposition will enable us to explain why a group of workers who had concerned about quality before may continue to concern more about quality in the “piece rate” period. It is because that group had a greater concern of quality in its original social norm under team production, which persist partly into the period of individual piece rate. In other words, that group has a bigger \( \eta \) due to its history of emphasizing quality during “group rate”.

Moreover, in comparison with the outputs between team production versus the production under individual piece rate, we have the following proposition.

Proposition 3. If \( P'(0) - Q'(0) \) is sufficiently large, a worker’s outputs of both quantity and quality will be lower under individual piece rate under some parameter configurations of the model.

Proof: See Appendix.

From the proof of the appendix, we can see that an individual will devote more efforts to the production of both quality and quantity if the peer effects under team work are sufficiently larger than those under individual piece rate and if individuals concern greatly about the disutility from exerting efforts at work.
3. Experiment Design and the Data Descriptions

We used field experiment for our empirical test. A field experiment is essential for the purpose of this paper, because employee or firm level survey data are often contaminated by various endogeneity problems due to self-selection and sorting. Also, it is usually difficult to acquire information on peer pressure and quality of output in such datasets. On the other hand, lab experiment may not provide a strong sense of “realism” that can generate sufficient peer effects.

Like Gneezy and List (2006) and Bellemare, Lapage and Shearer (2010), our field experiment involves hiring students to carry out real jobs (instead of participating in lab experiment). We hired a firm in Shenzhen, China to implement a summer internship scheme by recruiting students in a local university there. The firm first advertised a data input job through the Student Union’s intranet for such summer internship opportunities. The target recruitment is 40, and it attracted over 200 applicants. All applicants were then invited to participate in a random draw for the 40 places. For the 40 successful candidates, another round of random draw was conducted to separate them into a morning and an afternoon group with 20 interns each. They were informed that the internship will last for a month or 20 working days and the working hours were 3.5 per day (8:30-12:00 for the morning and 1:30-5:00 for the afternoon group). The job duty for the internship was data entry for a survey collected by us for another project on rural education in China.

The survey actually had already been carefully inputted and checked by people before. So the quality of the input (the error rate) can be quickly checked by the computer using a custom-made programme. We emphasized at the beginning that both quantity (measured by the number of questionnaires) and quality (measured by the percentage error rate) were important. As for quality of work, they were told that it would be randomly checked by the manager on daily basis. For the first two weeks their pay was...

---

7 In the end, the actual working days are 19 as the last day’s work only lasted a little over one hour and remaining time was used to give out payment and internship certificate.
8 Bellemare, Lapage and Shearer (2010) and Kaur, Kremer, and Mullainathan (forthcoming) also use data entry as the job duty for their experimental studies.
group/team performance based (by two teams: morning vs. afternoon). The pay formula is as follows:

\[(1-1.5*\text{error rate})N \times \$1\]

where error rate is the average percentage of entries inputted with errors per questionnaire, and N is the number of questionnaires inputted by a group. Notice that error is fined 1.5 times. This was suggested by the company hired by us as their usual practice. In reality, since the error rate is very low (average about 2.6%) the fine is still quite small relative to the overall pay.

In the next two weeks, an individual piece rate system was adopted and the same pay formula applied except that the quantity and quality of output were measured at individual level. This second phase of the experiment enables us to further test the effects of changing peer pressure brought by individual piece rate.

There were two RAs hired by us and they also acted as managers sent by the firm. One RA/manager is in charge of monitoring workers and the other mainly helps to deal with any technical problems faced by interns. The latter is also in charge of compiling the daily quantity and quality figures. The former conducts a short daily meeting with interns in each group before the start of the work to give them feedback on their previous day’s work. During the first two weeks, the first three days were used as our baseline period and each individual’s output (quantity of work) was announced to the whole group together with the average error rate of the group for both groups. Starting from the fourth day, the format was changed for the afternoon (treatment) group. In addition to announce each individual’s daily quantity of output quality information of each individual were also given to the whole group\(^9\). By so doing, we expect that the peer pressures people feel on these two dimensions of production will be different for the two groups, and therefore effort devoted by the team members to perform the two tasks will be different between the two groups, too. This provides us with the experimental environment to test our theory.

---

\(^9\) These error rates were said to be based on our randomly check. In reality, we used the compute programme to check the error rates for every questionnaire inputted.
Before the experiment starts, we also conducted a short survey with all interns. The questions in the survey cover individual characteristics such as gender, age, major of study in the university, etc. (Are they not used in the regressions?) It also covers questions on how many people they consider as friends in the same work team and how important is the internship certificate for them. Related to the last question, we had informed all interns that we would issue an internship certificate after they completed the work and their performance would be rated by the firm in the certificate.

The descriptive statistics of quantity and quality of output as well as other main variables derived from the survey are shown in Table 1. In Table 1, the average error rate of input (our measure of quality, which is calculated as the daily percentage of incorrect entries per questionnaire) is about 2.6% whilst the average quantity of output is about 58 questionnaires. There are about 37% of female interns. Roughly half of the people consider that the internship certificate is important for them. There are also 11% of people who have known someone as friends in their work team previously.

The quality and quantity data for the entire experimental period (10 days) are plotted in Figure 1 and 2. This helps to provide a graphical overview of our empirical findings. Figure 1 plots the average of the error rates (our measure of quality) for the two groups during the course of the experiment, while Figure 2 presents the average of units of output within the two groups during the course of the experiment. The error rate is calculated as the daily percentage of incorrect entries per questionnaire.

From these 2 figures, we can see that for the first three days (our baseline period) the afternoon group (the experimental group) performs slightly better in terms of both quality (less errors) and quantity, and the differences are not statistically significant, namely that there is no difference between the afternoon and the morning groups in terms of either the average quantity or the average quality produced, suggesting that the average productivity of the workers in the both groups are largely the same. Therefore, the result of randomization is highly satisfactory in dividing the workers into 2 groups: the control group and the treatment group.

The control group and the treatment group began to differ in incentives since Day 4, when the morning group continues to get the individual quantity of output information.
only whereas for the afternoon group, they are told their individual quality in addition to quantity of output. This experiment lasted for the next seven days and workers were paid based on the group piece rate as described before.

In this “treatment period”, namely Day 4 to Day 10, we can observe from Figures 1 and 2 that the quality gap between the two groups became widened. This quality gap, namely the difference of the average error rate between the 2 groups, is particularly striking from Day 4 to Day 8. As for quantity, the gap between the 2 groups was reversed after day 4 as the treatment group’s output became consistently below that of the control group.

In summary of the above, a simple comparison in Figures 1 and 2 presents clear patterns of differential behaviors of the control group and the treatment group, and this difference precisely follows what is predicted in our theoretical analysis.

4. Empirical Results for Team Production

As mentioned before, during the experimental period while interns in the morning group receive only peer pressure on quantity due to “shame in public” those in the afternoon team face peer pressure from both quantity and quality. Our theory predicts that workers in the afternoon team shall put relatively more effort into quality and less effort into quantity production than workers in the morning team. Consequently, they shall produce higher level of quality and lower level of quantity of output. Our empirical analyses below aim to show whether this is indeed the case.

We estimate the difference-in-difference effect of treatment with OLS first. The results are shown in Table 2. We first look at the effects on quality\(^\text{10}\). We start with a parsimonious model (Model I in Column 2) that only contains a time trend, a dummy variable identifying the treatment period, a dummy for the treatment group and the interaction of the last two. The coefficient of this interaction variable is the difference-in-difference estimate. In Panel A, the estimated coefficient in Column 2 is -.1831 and

\(^{10}\) Since the error rate is a very small number deviates around 0.02, as typically done for variable like this we use the logodds transformation of the error rate in regression so as to enlarge the range of variation of this variable. Such a transformation should not change the main results here.
significant at 10% level. This translates into an effect of reducing the mean error rate from 2.6% to 1.9% or approximately 25% (about half a standard deviation).

Now, let’s turn to the quantity dimension of the production. The corresponding difference-in-difference estimate is shown in Panel B. The coefficient is -5.236 and significant at 5% level. The treatment effect is therefore about 10% of the mean or a third of the standard deviation. So these results together provides strong evidence to support our theory in which our model predicts effort spent on quantity should reduce and effort spent on quality should increase when both quantity and quality information are revealed as compared to only the quantity information is known to team members.

Now, let’s turn to the estimated coefficients for the treatment period dummy. It should indicate the quality and quantity differences between the two groups without treatment. We can see that the estimated coefficient is actually insignificant, which confirms that the performance of the two groups is not significantly different without treatment. In other words, our random assignment does make sure there are no systematic differences in performance between the groups prior to treatment.

For the other controls, the time trend variable turns out to be negative (positive) in the quality (quantity) and highly significant in the equation. It indicates there is a strong learning-by-doing effects for worker productivity.

Next, we expand our model by adding more controls. So Model II in Column 3 adds a gender dummy. This variable itself is not significant in either quality or quantity equation, and there are hardly any changes to the results after the inclusion of this variable. Model III in Column 4 includes another two dummies that captures whether a worker considers the internship certificate very important to him/her and also whether a worker has any friends in the team before the start of the work. The former shows how serious a worker treats his/her job, which should also be related to his/her commitment to the job. The latter may potentially affect the impacts of peer pressure that a worker feels. The first dummy turns out to be highly significant. The signs of the coefficients suggest that more committed workers perform better in both quality and quantity dimensions. For quality, by comparing this coefficient and that on the treatment effect, it indicates the effect of hiring a committed worker is equivalent to two thirds of the treatment effect on
quality. For quantity, the effect of hiring a committed worker is even larger than the treatment effect. Together, it shows that hiring a committed worker is almost equivalent to implement our treatment on the treated group. The signs of the coefficients for the dummy on whether one has friends in the team previously also suggest the same effects. So having friends in the same team does seem to make you exert more effort which leads to higher quality and quantity. This implies that people have friends working in the same team also feel more peer pressure. However, these effects are insignificant.

As we have a panel data on individual performance here. We can also estimate the treatment effect using a fixed-effect model. The fixed-effect model has the advantage of controlling for any remaining impacts due to individual heterogeneity. While our random selection of people into the two groups should largely rule out such impacts, the fixed-effect estimation can help us verify that. In Model I of Table 3, we include only the time trend, the treatment period dummy and the interaction of the treatment period dummy with the treatment group dummy. Once again, the latter should give us the treatment effect. The estimated coefficient on this interaction term is negative for error rate but positive for quantity. This is consistent with our basic difference-in-difference results in Table 2. Furthermore, the estimated coefficients are similar in size and significance level, too. So they confirm that our early results are robust with regard to individual heterogeneity.

5. A Follow-up Experiment: Differential Performances under Piece Rates

Our (main) initial experiment lasted for 10 days. Our interns were informed beforehand that starting from the 11th day of their work, the payment system would be switched to individual piece rate for the remaining 10 days (only 9 days counted in this paper due to the reasons stated in footnote 5). By so doing, we intend to compare the peer effect under our teamwork to individual incentive effect under piece rate.
Most of the existing literature presumes that under the contract of piece rates, workers are only motivated by pecuniary concerns. Such a prior reasoning appears to be reasonable at first thought. Indeed, under piece rates, a worker’s labor input does not affect co-workers’ earnings, one might be tempted to argue that an individual worker minds her/his own business and does not concern about others, which implies that there would be little or no peer pressure in this case. Based on this reasoning, a theoretical prediction is that under piece rates, individuals in our two teams will have similar performances in terms of both the quantity and the quality of production after switching to piece rate.

A quick look at our Figure 1 and 2 suggests that the above hypothesis may not hold. From Figures 1 and 2, the differential behaviors of the control group and the treatment group continued in the period of piece rate: the treatment group consistently produces lower quantity, but with higher quality, than the control group. In what follows, we will see if this is indeed the case by our regression analyses.

Since both groups of workers were switched from teamwork to piece rate we no longer have the experimental setting of one control and one experiment group here. Nevertheless, our sample can still be viewed as free from self-selection and sorting as we have everyone switched to piece rate and no one dropped out after the switch. We test the effects of introducing the piece rate scheme on work quantity and quality by comparing output and quality before and after the advent of the piece rate scheme.

We first estimate the effects using OLS. The results are presented in Table 4. In Table 4 Column 2 and 3, we show results for the separate regressions of morning and afternoon group on quality, respectively. Column 4 then checks if the estimated coefficients from the two separate regressions are significantly different. These results come from a fully stacked regression with interaction terms of the afternoon group dummy and all other control variables. Similarly, Column 5 and 6 list the results for the separate regressions of morning and afternoon group on quantity, respectively. Column 7 shows the differences and significant level of the coefficients in Column 5 and 6 using the same methods as in Column 4.
The effect of introducing the piece rate scheme is reflected by the estimated coefficient on the piece rate dummy. We can see across board, the introduction of the piece rate does not improve either quality or quantity. In fact, for the morning group the quality deteriorated whilst for the afternoon group both quality and quantity deteriorated. These results are precisely consistent with the theoretical predictions of Proposition 3.

Intuitively, the explanation can be sought from the fact that under our piece rate scheme individual output in terms of both quantity and quality are no longer public information to other group members. This significantly reduces any peer pressure from the feeling of “shame”. Furthermore, under piece rate, there no longer exists the feeling of “guilt”. This further weakens the potential peer pressure. In other words, our results here can be driven by both the change of peer pressures. Indeed, a careful examination suggests that this finding could be related to our experimental design, i.e. under piece rate individual output of quantity and quality is no longer public information to all. While a worker might still be able to observe the quantities of the output of her/his co-workers by observing the piles of questionnaires that the co-workers her/his co-workers finished inputting, the information was after all not publicly announced. This might have weakened potential impacts of peer effects by under individual piece rate.

In fact, these results are also consistent with the early field results from Hamilton, Nickerson, and Owan (2003), where they also show that team-based performance pay outperformed individual performance pay for a garment factory. They claim that peer pressure and complementarities of skills among team members can both mitigate the agency problem and improve productivity. Furthermore, Kaur, Kremer and Mullainathan (2010) argue that individuals working alone on their own may face serious self-control problems and closer interactions with other workers can help ameliorate such a problem. In our study, workers continued to work in a team environment even under individual piece rate. However, it might be more interactions among workers when they were under group piece rate than under group piece rate under individual piece rate.

Next, we examine if there are any systematic differences between the two groups after switching to piece rate. Here, we first find that there is a large and significant
difference in the constant term for the quality regression. This shows that afternoon group continues to outperform the morning group with regard to quality of output. Furthermore, for the quantity, there is also a huge and significant difference in the estimated coefficient on piece rate dummy, indicating that the morning group does better in terms of quantity of output than the afternoon group. In other words, despite the fact both groups are now under the same piece rate scheme, their differential performance pattern developed from the early teamwork treatment remains. We discuss in detail the implication of this interesting finding in the next section.

Some of the other personal characteristics variables included in the model are also significant. So just like the case under the payment system of income sharing, workers who value more the internship certificate produce better quality and quantity at the same time. This result once again indicates that it pays for the firm to screen for more committed workers. The time trend in the quantity regressions remains to be positive and significant. Intuitively, there are two interpretations of the time trend: (1) a worker’s productivity increases over time due to learning-by-doing; (2) the impacts of the old social norm changed over time, which affected workers’ output. It should be noted that both of these factors may change over time, which are captured by the variable of the time trend.

As for quantity, the sign of the time trend is positive, possibly indicating the learning-by-doing effect is still quite strong for its production. However, from Figure 2 we can see that the error rates were close to zero at the end of the period of group piece rate, which suggests that the effect of learning-by-doing should be small in the period of individual piece rate. Consistent with this reasoning, we find that the coefficient on the time trend for the quality regressions now reverses sign and becomes positive and significant, which suggests that the old social norm continued to affect the workers’ productions of quality, and the impact eroded over time.

The empirical evidence shown in the proceeding section indicates that the two groups seem to carry their quality and quantity differences formed during the teamwork period of income sharing into the individual piece rate period, i.e. the morning group continues to produce more quantity and less quality than the afternoon group despite of the fact that the underlying incentive mechanism has changed with the introduction of the piece rate contract. In particular, from Figures 1 and 2 we can see clearly that after the switching to piece rate workers in both groups continues its pattern of quality production for two more days in spite of the fact that the individual piece rate incentive should induce workers to exert more effort on quantity of production.

This finding suggests that social norms are “stubborn” and persistent. A large economics literature argues that social customs were usually determined by economic fundamentals. Moreover, once a social norm is formed, it will not disappear immediately after its underlying economic fundamental changes. For example, Akerlof (1980) shows that “social customs which are disadvantageous to the individual may nevertheless persist without erosion” (p. 749). As noted by Akerlof (1980), this phenomenon applied to many social norms, particularly in the short run.

In our experimental study, in the first ten days, the workers worked as teams, and might have formed social norms for their team productions. Starting from Day 11, they switched to the contract of piece rates, which substantially changed the “economic fundamentals” in this setting. Since every individual faced a different earning function from then on, the social norm of working will change accordingly no matter whether other kinds of peer pressures will be present or not under this new contract.

However, if every worker believed that the social norm was still the old one, and her/his co-workers would not change their behaviors, he/she would not deviate from the old norm.\footnote{Fischbacher and Gachter (2010) present an experimental study showing that beliefs significantly govern economic behaviors.} For simplicity, we do not write down a formal model of worker behaviors under this old belief, but it will be easy to check that if the item \(Q'(0)\) is large enough and if everyone believes that her/his coworkers will not deviate from the old norm, then
he/she will continue to follow the old norm in the production of both quantity and quality. The above reasoning explains workers’ behaviors on Day 11 and Day 12.

However, the change of social norm did appear to have occurred on Day 13, the third day of piece rates. Such a norm change is reasonable because workers operated in a different contract system, which was associated with a different social norm. Thus, sooner or later, the social norm is expected to change.

But why did the norm change occur in the third day of piece rate? Economic theory cannot answer this question, and we did not manage to get the relevant information in the experiment to answer this question. Indeed, economic theory suggests that the change of norm can happen any time with the advent of an exogenous shock to people’s “beliefs”. However, a plausible conjecture is that on Day 12, workers observed that their co-workers obviously increased the quantity of output, which signaled to them that all others started reducing quality, which induced another self-fulfilling Nash equilibrium to occur.

Moreover, once social norm changed, people’s behaviors may change drastically. For example, Akerlof (1980) emphasizes that people’s deviations from old social norm are usually “non-trivial”. What happened in our experiment was not an exception: the error rates increased more than 50 percent in comparison between Day 13 and Day 12. Since Day 13, however, the changes of error rates became small again.

The next question is: why did the “treatment” group (in team production) continue to produce less quantity and better quality in the new environment of piece rate? We suggest three related answers to this question: (1) Peer pressures might have continued to exist under piece rate. (2) As in the case of team production, peer pressures under piece rates also derive from a combination of shames and guilt. (3) This combination of shames and guilt is “path-dependent”, which means it is related to the history of team production.

Since the “treatment” group and the “control group” had different history of social norms in team production, their social norms in the new environment, which partly inherited their “old traditions”, were different in the period of piece rate. Thus, interns in the afternoon group behaved differently from those in the morning group. As demonstrated early in the section, the morning team produced more quantity on average, while the afternoon team emphasized more in the production of quality. With reference to
Equation (23), we hypothesize that the afternoon group is associated with a greater value of the parameter, $\eta$, which implies that the afternoon concerns less about quantity and more about quality even in the period of piece rate. Thus, this “cultural” difference continues when the workers engaged in the production of piece rates.

In fact, under certain parameter configurations, our model can yield this outcome. The simulation exercise in Appendix 2 shows such a possibility.

The analysis of this section shares the essence of a recent study by Gneezy, Leibbrandt, and List (2014), who conduct an empirical study of social norms by comparing the behaviors of two types of Brazilian fishermen. They note that fishermen work in groups on the sea but work alone on the lakes. Based on this phenomenon, they hypothesize that the society of the fishermen located by the sea is more cooperative than that of the fishermen located by the lakes. Their empirical study confirms this hypothesis. Specifically, they find that “Sea fishermen trust significantly more and are also significantly more cooperative, return more money in the trust game, propose more equal offers in the ultimatum game, contribute more in the public goods game, and donate more to a charity outside their own society.”

In another related study, Frey and Meier (2004) examine how social comparisons affected students’ charity donation in a field experiment at the University of Zurich. They find that when a student was (privately) informed that a higher percentage of the student population made a donation, the more likely the student would contribute as well.

Gneezy, Leibbrandt, and List (2014) and Frey and Meier (2004) show that individual behavior is strongly affected by her/his reference group. In other words, a “sub-culture” is often formed as long as there are human interactions. This “sub-culture” may be developed over a long period of time as studied by Gneezy, Leibbrandt, and List (2014), but may also be formed instantly as studied by Frey and Meier (2004). Therefore, in line with the spirit of Frey and Meier (2004) and Gneezy, Leibbrandt, and List (2014), we expect that the morning group and the afternoon group under different team production arrangement might have developed different cultures, which will continue to exist (to some extent) in the new environment of piece rates. This provides explanations for the empirical findings of the follow-up experiment of piece rates.
7. Conclusion

Recently, there is a large literature that investigates the impacts of interpersonal comparisons on economic behaviors. For example, Card, Mas, Moretti, and Saez (2012) study workers’ reactions to the information on peers' wages in the University of California. They find that the workers with salaries below the median of their peers reported job dissatisfactions and were more likely to look for new jobs. Also, as described earlier, Frey and Meier (2004) examine find that social comparisons are also important for charitable donations.

Interpersonal comparisons also matter greatly to work effort through the channel of “peer effects”. However, when it comes to work effort, interpersonal comparisons often become more complicated than in other cases. First, workers’ effort is often of multiple dimensions, such as the efforts in the production of quantity versus quality. In contrast, a worker’s comparison with her peers’ wage is more straightforward. Second, workers’ effort often has differential observability in various aspects. In peer monitoring, observability is crucial because peers usually do not have such an authority of requesting to observe co-workers if certain aspects of work efforts are not easily observed.

The existing literature has largely focused on the identification of the existence of peer effect at work and the estimation of the magnitude of peer effects. However, it has largely ignored the feature of multiple dimensions of peer effects. This paper attempts to help fill this gap.

We first develop a model, in which the quantity that an individual produces is more observable than the quality. It shows that when the “exhaustion” effect from effort in both producing quality and quantity is large, and when the complementary effect of these two efforts in generating earnings is small, then an increase in the observability of quality will lead a worker to produce less quantity and more quality.

We test this theoretical implication with the data we collected from a field experiment in China. Our field experiment involves hiring students (interns) to carry out real jobs (instead of participating lab experiment). We randomly divide workers into two groups: morning and afternoon. In the morning team, the manager give the information
on the quantity of output of every team member together with the average quality of output (error rate) for the entire team of the previous day for both the morning and afternoon teams. In contrast, the afternoon team receives information on both quantity and quality of each team member. Our empirical analysis shows that the team of the afternoon produced less quantity of output with higher quality, which is in support of our theoretical prediction.

Our experiment of team production is followed by another experiment of switching the team-based incentive (teamwork) to individual-based incentive (piece rate) contract. There are several interesting observations. First, the workers seem to carry their quality and quantity production patterns formed during the teamwork period into the piece rate period, i.e. the morning group continues to produce more quantity and less quality than the afternoon group. Furthermore, after the switching to piece rate workers in both groups continues its pattern of quality production for two more days despite of the fact that the individual piece rate incentive should induce workers to exert more effort on quantity of production. We explain this finding as that it usually takes time for one social norm to change into another social norm even after the “material foundation” of a society changes. In the first day of the piece rate, every individual was concerned that her/his team members would continue to stick to old social norm. Consequently, this belief was self-fulfilling. The same happened on the second of the piece rate. On the third day, however, all individuals believed that the social norm would change to a new one, which was indeed established through another self-fulfilling prophecy. However, even in this case, both groups were affected by their differential history, which led to differential “sub-cultures” between them and in turn dictating differential equilibria between the two groups.

Moreover, our theoretical analysis shows that an individual will devote more efforts to the production of both quality and quantity if the peer effects under team work are sufficiently larger than those under individual piece rate and if individuals concern greatly about the disutility from exerting efforts at work. The empirical finding of our experimental study is consistent with this theoretical prediction: we find that both groups
produce less quantity of output under piece rate than under teamwork after the adjustment for the time trend.

Table 1. Means and Standard Deviation of the Main Variables for Regressions

<table>
<thead>
<tr>
<th>Variables</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quality (Error Rate: %)</td>
<td>0.026</td>
<td>0.012</td>
</tr>
<tr>
<td>Quantity</td>
<td>58.22</td>
<td>17.25</td>
</tr>
<tr>
<td>Female (1=yes)</td>
<td>0.37</td>
<td>0.48</td>
</tr>
<tr>
<td>Difficult to find internship (1=yes)</td>
<td>0.55</td>
<td>0.50</td>
</tr>
<tr>
<td>Internship certificate important (1=yes)</td>
<td>0.54</td>
<td>0.50</td>
</tr>
<tr>
<td>Have friends in the team (1=yes)</td>
<td>0.11</td>
<td>0.31</td>
</tr>
</tbody>
</table>

**Definitions:** Error rate – the daily percentage of incorrect entries per questionnaire  
Quantity – total number of questionnaires inputted a day (N)
Table 2: Difference-in-Difference Results for Teamwork: OLS Estimates

### Panel A: Quality (Logodds of Error rate)

<table>
<thead>
<tr>
<th>Control variables</th>
<th>Model I</th>
<th>Model II</th>
<th>Model III</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-3.1650</td>
<td>-3.2036</td>
<td>-3.0984</td>
</tr>
<tr>
<td></td>
<td>(59.47)***</td>
<td>(38.73)***</td>
<td>(34.36)***</td>
</tr>
<tr>
<td>Time trend</td>
<td>-.1322</td>
<td>-.1322</td>
<td>-.1322</td>
</tr>
<tr>
<td></td>
<td>(10.01)***</td>
<td>(10.00)***</td>
<td>(9.97)***</td>
</tr>
<tr>
<td>Period (=1 for treatment period)</td>
<td>.1417 (1.68)*</td>
<td>.1436 (1.71)*</td>
<td>.1447 (1.71)*</td>
</tr>
<tr>
<td>Treated (=1 for treated group: afternoon group)</td>
<td>-.1113 (1.16)</td>
<td>-.0947 (.93)</td>
<td>-.0733 (.67)</td>
</tr>
<tr>
<td>Period x Treated</td>
<td>-.1831 (1.76)*</td>
<td>-.1851 (1.77)*</td>
<td>-.1862 (1.78)*</td>
</tr>
<tr>
<td>Female (1=yes)</td>
<td>.0734 (.85)</td>
<td>.0492 (.57)</td>
<td></td>
</tr>
<tr>
<td>Internship certificate important (1=yes)</td>
<td></td>
<td>-.1279</td>
<td>(1.87)*</td>
</tr>
<tr>
<td>Have friends in the team (1=yes)</td>
<td></td>
<td>-.0937 (.79)</td>
<td></td>
</tr>
<tr>
<td>R²</td>
<td>.3779</td>
<td>.3808</td>
<td>.3890</td>
</tr>
<tr>
<td>Sample size</td>
<td>397</td>
<td>397</td>
<td>397</td>
</tr>
</tbody>
</table>

### Panel B: Quantity

<table>
<thead>
<tr>
<th>Control variables</th>
<th>Model I</th>
<th>Model II</th>
<th>Model III</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>28.3452</td>
<td>28.1326</td>
<td>22.4749</td>
</tr>
<tr>
<td></td>
<td>(15.17)***</td>
<td>(14.08)***</td>
<td>(8.48)***</td>
</tr>
<tr>
<td>Time trend</td>
<td>3.0818</td>
<td>3.0818</td>
<td>3.0818</td>
</tr>
<tr>
<td></td>
<td>(14.39)***</td>
<td>(14.37)***</td>
<td>(14.33)***</td>
</tr>
<tr>
<td>Period (=1 for treatment period)</td>
<td>5.5751</td>
<td>5.5857</td>
<td>5.5247</td>
</tr>
<tr>
<td></td>
<td>(3.28)***</td>
<td>(3.33)***</td>
<td>(3.23)***</td>
</tr>
<tr>
<td>Treated (=1 for treated group: afternoon group)</td>
<td>3.4079 (1.34)</td>
<td>3.4993 (1.51)</td>
<td>2.4546 (1.03)</td>
</tr>
<tr>
<td>Period x Treated</td>
<td>-5.2365</td>
<td>-5.2471</td>
<td>-5.1861</td>
</tr>
<tr>
<td></td>
<td>(2.53)**</td>
<td>(2.54)**</td>
<td>(2.48)**</td>
</tr>
<tr>
<td>Female (1=yes)</td>
<td>.4039 (.16)</td>
<td>1.7578 (.75)</td>
<td></td>
</tr>
<tr>
<td>Internship certificate important (1=yes)</td>
<td></td>
<td>6.8771</td>
<td>(2.77)***</td>
</tr>
<tr>
<td>Have friends in the team (1=yes)</td>
<td></td>
<td>4.5535</td>
<td>(1.38)</td>
</tr>
<tr>
<td>R²</td>
<td>.4512</td>
<td>.4513</td>
<td>.4928</td>
</tr>
<tr>
<td>Sample size</td>
<td>397</td>
<td>397</td>
<td>397</td>
</tr>
</tbody>
</table>

Note: 1. t-statistics are included in the brackets and clustered standard errors are calculated by individuals; 2. **-significant at 5% ***-significant at 1%.
Table 3: Difference-in-difference Results for Teamwork: Fixed-effect Estimates

<table>
<thead>
<tr>
<th>Control variables</th>
<th>Quality (Logodds of Error rate)</th>
<th>Quantity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Model I</td>
<td>Model I</td>
</tr>
<tr>
<td>Constant</td>
<td>-3.2163 (66.68)*****</td>
<td>29.7903 (38.46)*****</td>
</tr>
<tr>
<td>Time trend</td>
<td>-.1322 (10.37)*****</td>
<td>3.0818 (15.05)*****</td>
</tr>
<tr>
<td>Period (=1 for treatment period)</td>
<td>.1282 (1.35)</td>
<td>6.3454 (4.18)*****</td>
</tr>
<tr>
<td>Treated x Treatment period</td>
<td>-.1697 (1.74)*</td>
<td>-6.0068 (3.83)*****</td>
</tr>
<tr>
<td>R-squared</td>
<td>.3679</td>
<td>.4348</td>
</tr>
<tr>
<td>Sample Size</td>
<td>397</td>
<td>397</td>
</tr>
</tbody>
</table>

Note: 1. t-statistics are included in the brackets and clustered standard errors are calculated by individuals; 2. **-significant at 5%  ***-significant at 1%.
Table 4: Teamwork vs. Piece Rate: OLS Estimates

<table>
<thead>
<tr>
<th>Control variables</th>
<th>Quality (Logodds of Error rate)</th>
<th>Quantity</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Morning</td>
<td>Afternoon</td>
<td>Diff</td>
<td>Morning</td>
<td>Afternoon</td>
<td>Diff</td>
<td>Morning</td>
<td>Afternoon</td>
<td>Diff</td>
<td>Morning</td>
<td>Afternoon</td>
<td>Diff</td>
</tr>
<tr>
<td>Constant</td>
<td>-4.0019 (24.14)**</td>
<td>-4.3685  (40.07)***</td>
<td>-3.3665 (1.87)*</td>
<td>30.0602 (5.26)***</td>
<td>31.2575 (7.28)***</td>
<td>1.1973 (.17)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time trend</td>
<td>.0196 (1.55)</td>
<td>.0293 (3.40)***</td>
<td>.0096 (1.57)</td>
<td>2.4807 (13.09)***</td>
<td>2.4915 (15.19)***</td>
<td>.0108 (.04)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Piece Rate</td>
<td>.2566 (2.55)**</td>
<td>.3571 (3.87)***</td>
<td>.1005 (.75)</td>
<td>-1.4772 (.74)</td>
<td>-8.3239 (5.71)***</td>
<td>-6.8467 (2.81)***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1=yes)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>.0914 (.98)</td>
<td>.0024 (.02)</td>
<td>-.0890 (.66)</td>
<td>1.3290 (.28)</td>
<td>.5772 (.14)</td>
<td>-.7518 (.12)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1=yes)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Internship</td>
<td>-.1980 (2.61)**</td>
<td>-.0984 (1.28)</td>
<td>.0995 (.93)</td>
<td>10.2907 (2.18)**</td>
<td>4.6803 (1.14)</td>
<td>-5.6104 (.91)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>certificate</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>important</td>
<td>(1=yes)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Have friends</td>
<td>.1701 (2.30)**</td>
<td>-.02601 (.31)</td>
<td>-.1962 (1.77)*</td>
<td>3.9982 (1.25)</td>
<td>5.1322 (1.19)</td>
<td>1.1340 (.21)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>in the team</td>
<td>(1=yes)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R-squared</td>
<td>.1721</td>
<td>.2246</td>
<td>.2202</td>
<td>.5087</td>
<td>.3504</td>
<td>.4586</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sample size</td>
<td>320</td>
<td>320</td>
<td>640</td>
<td>320</td>
<td>320</td>
<td>640</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: 1. t-statistics are included in the brackets and clustered standard errors are calculated by individuals; 2. **-significant at 10% **-significant at 5% ***-significant at 1%.

Table 5: Teamwork vs. Piece Rate: Fixed-effect Estimates

<table>
<thead>
<tr>
<th>Control variables</th>
<th>Quality (Logodds of Error rate)</th>
<th>Quantity</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Morning</td>
<td>Afternoon</td>
<td>Diff</td>
<td>Morning</td>
<td>Afternoon</td>
<td>Diff</td>
<td>Morning</td>
<td>Afternoon</td>
<td>Diff</td>
<td>Morning</td>
<td>Afternoon</td>
<td>Diff</td>
</tr>
<tr>
<td>Constant</td>
<td>-4.0863 (42.00)***</td>
<td>-4.4481  (42.93)***</td>
<td>-3.3665 (1.87)*</td>
<td>38.1281 (26.00)***</td>
<td>36.2239 (27.95)***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time trend</td>
<td>.0196 (1.58)</td>
<td>.0293 (2.22)**</td>
<td>.0096 (.53)</td>
<td>2.4807 (13.27)***</td>
<td>2.4915 (15.08)***</td>
<td>.0108 (.04)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Piece Rate</td>
<td>.2566 (2.23)**</td>
<td>.3571 (2.91)***</td>
<td>.1005 (.60)</td>
<td>-1.4772 (.85)</td>
<td>-8.3239 (5.42)***</td>
<td>-6.8467 (2.95)***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1=yes)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R-squared</td>
<td>.1284</td>
<td>.2210</td>
<td>.1222</td>
<td>.4275</td>
<td>.3265</td>
<td>.4003</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sample size</td>
<td>320</td>
<td>320</td>
<td>640</td>
<td>320</td>
<td>320</td>
<td>640</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: 1. t-statistics are included in the brackets and clustered standard errors are calculated by individuals; 2. ***-significant at 1%.
Figure 1. Error Rate

Daily Average Error Rate

Date

teamwork  piece rate  Morning Group  Afternoon Group

y1  y2
Figure 2. Quantity of Output

Daily Average Quantity of Output

Date

teamwork
piece rate
Morning Group
Afternoon Group
y1
y2
Appendix. Mathematical Proofs

Proof of Proposition 1.

Totally differentiating (21) and (22) with respect to $e^a$, $e^b$ and $\lambda$, we get

\[
\begin{align*}
\left[ \frac{f^a(e^a)G_1(m,q)}{N} + \frac{f^a(e^a)G_1(m,q)}{N} \right] de^a + \frac{f^a(e^a)G_{12}(m,q)}{N} de^b & \\
- \left[ C_{11}(e^a,e^b) + f^a(e^a)P'(0) \right] de^a - C_{12}(e^a,e^b) de^b &= 0 \\
\text{(A1)}
\end{align*}
\]

and

\[
\begin{align*}
\left[ \frac{h^b(e^b)G_2(m,q)}{N} + \frac{h^b(e^b)G_2(m,q)}{N} \right] de^b + \frac{h^b(e^b)G_{21}(m,q)}{N} de^a & \\
- C_{21}(e^a,e^b) de^a - C_{22}(e^a,e^b) de^b + P'(0)h'(e^b)d\lambda + \lambda P'(0)h'(e^b) de^b &= 0 \\
\text{(A2)}
\end{align*}
\]

We can rewrite (A1) and (A2) in the following form:

\[
\begin{align*}
\left[ \frac{f^a(e^a)G_1(m,q)}{N} + \frac{f^a(e^a)G_1(m,q)}{N} - C_{11}(e^a,e^a) - f^a(e^a)P'(0) \right] de^a & \\
+ \frac{f^a(e^a)G_{12}(m,q)}{N} - C_{12}(e^a,e^a) & \\
\left[ \frac{h^b(e^b)G_2(m,q)}{N} - C_{22}(e^a,e^a) + P'(0)h'(e^b)d\lambda \right] de^b & \\
+ \frac{h^b(e^b)G_{21}(m,q)}{N} - C_{21}(e^a,e^a) + \lambda P'(0)h'(e^b) de^b &= 0 \\
\text{(A3)}
\end{align*}
\]

We define
\[ \Delta = \begin{vmatrix} \frac{f''(e^a)G_{11}(m,q)}{N} + \frac{f'(e^a)G_{11}(m,q)}{N} - C_{11}(e^a, e^b) - f''(e^a)P'(0) & \frac{f'(e^a)G_{12}(m,q)}{N} - C_{12}(e^a, e^b) \\ h'(e^b)G_{21}(m,q) - C_{12}(e^a, e^b) & \frac{h'(e^b)G_{22}(m,q)}{N} + \frac{h'(e^b)G_{22}(m,q)}{N} - C_{22}(e^a, e^b) + \lambda P'(0)h''(e^b) \end{vmatrix} \]

(A4)

Then, we have
\[
\frac{d e^a}{d \lambda} = \frac{1}{\Delta} \begin{vmatrix} 0 & \frac{f'(e^a)G_{12}(m,q)}{N} - C_{12}(e^a, e^b) \\ -P'(0)h'(e^b) & \left[ \frac{h'(e^b)G_{21}(m,q)}{N} + \frac{h'(e^b)G_{22}(m,q)}{N} - C_{22}(e^a, e^b) + \lambda P'(0)h''(e^b) \right] \end{vmatrix}
= \frac{P'(0)q(e^b)}{\Delta} [C_{12}(e^a, e^b) - \frac{f''(e^a)G_{12}(m,q)}{N}]
\]

(A5)

and
\[
\frac{d e^b}{d \lambda} = \frac{1}{\Delta} \begin{vmatrix} \frac{f''(e^a)G_{11}(m,q)}{N} + \frac{f'(e^a)G_{11}(m,q)}{N} - C_{11}(e^a, e^b) - f''(e^a)P'(0) & 0 \\ \frac{h'(e^b)G_{21}(m,q)}{N} - C_{12}(e^a, e^b) & -P'(0)h'(e^b) \end{vmatrix}
= \frac{P'(0)q(e^b)}{\Delta} \left[ -\frac{f''(e^a)G_{11}(m,q)}{N} - \frac{f'(e^a)G_{11}(m,q)}{N} + C_{11}(e^a, e^b) + f''(e^a)P'(0) \right]
\]

(A6)

Note that \( \Delta \) is just the Hessian determinant. Thus, at the optimum, we must have \( \Delta > 0 \). Thus, if Condition (24) holds, we have \( \frac{d e^a}{d \lambda} < 0 \).

Then, from (11) and the assumption that and \( G_{11} < 0 \)
\[
\frac{f''(e^a)G_{11}(m,q)}{N} - \frac{f'(e^a)G_{11}(m,q)}{N} + C_{11}(e^a, e^b) + f''(e^a)P'(0) > 0 \quad \text{(A7)}
\]
Then, from (A7) and (A6), we have \( \frac{de^b}{d\lambda} > 0 \). □

**Proof of Proposition 3.**

The equilibrium is

\[
\begin{align*}
G_1(e^a, e^b) &= \frac{C_1(e^a, e^b) - f'(e^a)P'(0)}{N} \\
G_2(e^a, e^b) &= \frac{C_2(e^a, e^b) - \lambda f''(e^b)P'(0)}{N} \\
G_1(\tilde{e}^a, \tilde{e}^b) &= C_1(\tilde{e}^a, \tilde{e}^b) - Q'(0) \\
G_2(\tilde{e}^a, \tilde{e}^b) &= C_2(\tilde{e}^a, \tilde{e}^b) - \eta h'(\tilde{e}^b)Q'(0)
\end{align*}
\]

The function forms can be set as follows:

\[
\begin{align*}
G(e^a, e^b) &= \Lambda mq \\
f(e^a) &= e^a \\
h(e^b) &= \tau_0 + \tau_1 e^b \\
C(e^a, e^b) &= (e^a + e^b)^2
\end{align*}
\]

Therefore, eq (A8) can be expressed as

\[
\begin{align*}
\frac{\Lambda}{N} \tau_0 + \frac{\Lambda}{N} \tau_1 e^b &= 2e^a + 2e^b - P'(0) \\
\frac{\Lambda}{N} \tau_1 e^a &= 2e^a + 2e^b - \lambda \tau_1 P'(0) \\
\Lambda \tau_0 + \Lambda \tau_1 \tilde{e}^b &= 2\tilde{e}^a + 2\tilde{e}^b - Q'(0) \\
\Lambda \tau_1 \tilde{e}^a &= 2\tilde{e}^a + 2\tilde{e}^b - \eta \tau_1 Q'(0)
\end{align*}
\]

By solving eq (A10), we get

\[
\begin{align*}
e^a &= \left( \frac{\Lambda}{N} \lambda \tau_1^2 - 2\lambda \tau_1 + 2 \right) P'(0) + 2 \frac{\Lambda}{N} \tau_0 \\
&\quad \frac{\Lambda}{N} \tau_1 \left( 4 - \frac{\Lambda}{N} \tau_1 \right) \\
e^b &= \left( \frac{\Lambda}{N} \right)^2 \tau_0 \tau_1 - 2 \frac{\Lambda}{N} \tau_1 \tau_0 + \left( \frac{\Lambda}{N} \tau_1 + 2\lambda \tau_1 - 2 \right) P'(0) \\
&\quad \frac{\Lambda}{N} \tau_1 \left( 4 - \frac{\Lambda}{N} \tau_1 \right) \quad \text{(A12)}
\end{align*}
\]
\[
\tilde{e}^a = \frac{\left( \lambda \eta \tau_1^2 - 2 \eta \tau_1 + 2 \right) Q'(0) + 2 \Lambda \tau_0}{\Lambda \tau_1 (4 - \Lambda \tau_1)} \tag{A13}
\]
\[
\tilde{e}^b = \frac{\Lambda^2 \tau_0 \tau_1 - 2 \Lambda \tau_0 + (\Lambda \tau_1 + 2 \eta \tau_1 - 2) Q'(0)}{\Lambda \tau_1 (4 - \Lambda \tau_1)} \tag{A14}
\]

Therefore, under certain condition, we can have \( e^a > \tilde{e}^a, e^b > \tilde{e}^b \) if

\[
P'(0) > \max \left\{ \begin{array}{c}
\frac{2 \Lambda^2 \tau_0 \tau_1 (N-1)}{(4 - \tau_1 \Lambda) N \left( \tau_1^2 \lambda \Lambda - 2 \tau_1 \lambda N + 2 N \right)} - \frac{\Lambda^2 \eta \tau_1^3 - 4 \Lambda \eta \tau_1^2 + 8 \Lambda \eta \tau_1 + 2 \Lambda \tau_1 - 8 N}{(4 - \tau_1 \Lambda) N \left( \tau_1^2 \lambda \Lambda - 2 \tau_1 \lambda N + 2 N \right)} Q'(0), \\
\frac{2 \Lambda^2 \tau_0 \tau_1 (N-1)}{(4 - \tau_1 \Lambda) N \left( 2 \tau_1 \lambda N + \Lambda \tau_1 - 2 N \right)} - \frac{\Lambda^2 \tau_1^2 - 4 \Lambda \eta \tau_1^2 - 8 \Lambda \eta \tau_1 + 8 N \eta \tau_1 - 2 \Lambda \tau_1 + 8 N}{(4 - \tau_1 \Lambda) N \left( \tau_1^2 \lambda \Lambda - 2 \tau_1 \lambda N + 2 N \right)} Q'(0)
\end{array} \right\} \tag{A12}
\]
References


42