Unintended Consequences of a Piece Rate?

Evidence from a Field Experiment

John S. Heywood[a], Stanley Siebert[b] and Xiangdong Wei[c]

May 2013

Abstract

This field experiment examines "rational cheating" in a data input business, comparing piece rates vs. time with more or less monitoring. We find that increased monitoring under time rates simply demotivates, reducing quantity produced without helping errors. The application of piece rates then increases quantity, and we find that the resultant cheating can be controlled by generating fines for errors via monitoring. Thus payment systems can be designed to ensure workers consider both quantity and quality. but we find an empirical tradeoff, so it is not possible to have more of both. Shifting the trade-off outwards requires more committed workers.

Keywords: piece rate, monitoring, shirking, quantity and quality trade off, field experiment

JEL Code: D2, J3, L2, M5

[a]Department of Economics, University of Wisconsin-Milwaukee, P.O. Box 413, Milwaukee, WI 53201 USA
[b]Birmingham Business School, University of Birmingham, Birmingham, B15 2TT, UK
[c]Department of Economics, Lingnan University, 8 Castle Peak Road, Tuen Mun, Hong Kong

Acknowledgement: We are most grateful to the funding support by the joint ESRC-RGC grant (RES-000-22-3653), and also the excellent research support provided by our research assistants, Dr. Jia Wu and Ms. Yufei Zheng.

I. Introduction

In their seminal paper on multitask principal-agent analyses, Holmstrom and Milgrom (1991) emphasize that "Multidimensional tasks are ubiquitous in the world of business. As simple examples, production workers may be responsible for producing a high volume of good quality output, or they may be required both to produce output and to care for the machines they use. If volume of output is easy to measure but the quality is not, then a system of piece rates for output may lead agents to increase the volume of output at the expense of quality." As MacDonald and Marx (2001) put it, for the owners, quality and quantity are complements, but for their agents, the workforce, quality and quantity are more likely substitutes. Indeed Freeman and Kleiner (2005) detail how a shoe manufacturer by abandoning piece rates was able to increase quality and profitability even as productivity declined. This paper provides field experiment results for workers doing data input for an actual firm to examine the trade-off between quantity and quality.

Lazear (1995 p. 24) recognizes the tradeoff between quantity and quality with a piece rate but emphasizes that firms can overcome this problem. "Piece rates do not necessarily overweight quantity. The exact compensation formula determines the emphasis on the one versus the other. For example, the typist who is paid on the basis of the number of pages typed goes too quickly and makes too many errors. On the other hand, the typist who is penalized significantly for each error may end up typing too slowly and producing too few pages. Thus, there is always an appropriate compensation formula that will induce workers to put forth the right amount of effort towards quantity and quality." Thus, our interest in the field experiment is to observe the advent of a piece rate that rewards quantity while penalizing low quality. We vary the expected penalty by examining two different monitoring intensities. We show that with a sufficient expected penalty it is possible to have a piece rate that increases quantity

while maintaining and even improving quality.

Despite frequent discussion in the theoretical literature and the obvious relevance to business practitioners, there are few if any experiments studying the advent of a piece rate that rewards both quantity and quality. One main reason for the general lack of such empirical studies may be that quality and quantity dimensions are endogenous and reflect worker sorting, as emphasized by Shearer (2004). To overcome such endogneity, one either has to estimate a structural model or use randomized experimental data or both. Paarsch and Shearer (2000) estimate a structural model of the firm's choice of a piece rate versus time rate and they show that under piece rates workers can respond by increasing quantity at the expense of quality. However, this finding is largely a function of the particular structural model they use. Several recent studies use experimental designs to study the effect of piece rates. Shearer (2004) examines the productivity effects associated with piece rates by randomly assigning tree planters to work under either a time rate or piece rate. He finds that piece rates increase productivity about 20%. Recognizing that his result may depend on the experimental environment (e.g. planting conditions may affect incentives), Shearer confirms out-of-sample effects of roughly the same size using a structural model. Guiteras and Jack (2012) also carry out a field experiment on worker response to different piece rates and monitoring regimes, this time in Malawi. They aim to separate the productivity effect of piece rates into two channels, an incentive and a selection effect. They conclude that the incentive effect outweighs the selection effect, and more relevant to our studythey find that the monitoring of quality induces workers to improve quality at some cost in quantity. This ability to influence quality through monitoring and penalties is the point of Lazear (1995) in the quote above. Finally, Nagin, Rebitzer, Sanders and Taylor (2002) examine field experimental data collected from a call centre which solicited donations. They show that work quality depends on monitoring intensity. The more intense the monitoring (raising the expected fine), the

less workers "rationally cheat".

In an idealized piece rate scheme, quality can be easily detected and the rate is paid only for those pieces meeting a quality standard. Yet, when quality is more expensive to detect, imperfect monitoring and quality control techniques become critical. At its most essential, this means that the quality of only some pieces will be examined and management will respond with financial incentives that reward both quantity and (imperfectly) observed quality. The extent of the monitoring and the nature of the rewards become crucial in determining the workers' quantity and quality outcomes. Importantly, the simple substitution of quantity for quality should not be taken for granted.

The current research demonstrates the potential influence of monitoring intensity and fines when workers respond to a new piece rate. We provide a field experiment showing that when workers face increased monitoring and associated fines for poor quality, the well-known productivity increase (quantity increase) associated with piece rates (Lazear 2000) is muted. With hard to detect quality and without monitoring, the piece rate should generate greater output but lower quality as workers adversely specialize in only rewarded tasks (MacDonald and Marx 2001). As monitoring for quality increases (and so quality is implicitly rewarded), the extent of this specialization should be reduced. With sufficiently close monitoring, the expected rewards for quality grow and the advent of the piece rate can generate improved quality without an increase (or with only a small increase) in output. Again, borrowing language from Nagin et al. (2002), the closely monitored agent reduces the extent of "rational cheating" on quality with the advent of the piece rate. We emphasize, however, that reducing this cheating can come with a cost of reduced quantity, the magnitude of this cost depending upon worker characteristics.

This setting provides the opportunity to examine worker characteristics associated with variations in the quality-quantity trade-off. We show that workers who find the work intrinsically meaningful produce both greater quantity and higher quality. This pattern is mimicked by those who think it is difficult to find an alternative job. These workers also produce more and have higher quality. An alternative is provided by those workers who express that they had anticipated higher earnings than those associated with their current data entry job. These workers may well be the more able (or at least think they are) or they may those who value the job the least. They emerge as having a strong trade-off, producing more than otherwise comparable workers but with lower quality. Accounting for these, and other, worker differences does not change the critical result that the advent of the piece rate generates higher quality and more modest increases in output when the expected fine (dependent on the intensity of monitoring) is high.

The remainder of the paper is organized as follows. The next section presents a theoretical setting designed to isolate both the role of a piece rate with imperfect monitoring and the role worker heterogeneity. The third section describes the experimental design and the resulting data. The fourth section summarizes our estimations and isolates the critical role played by the extent of monitoring. It also isolates the differences associated with worker characteristics (heterogeneity). A final section concludes.

II. An Illustrative Theoretical Model

We construct a model examining a representative worker's quantity and quality choices under different payment and monitoring regimes. The worker makes an aggregate effort choice, e^*, and decides how much to devote to producing quantity e_1, and quality e_2. Thus, $e_1 + e_2 = e^* \leq e$, where e is a constraint given by ability. N is the quantity of

output the worker produces and *E* is the (inverse) quality measure, the proportion of *N* with quality below a pre-defined standard[1]. We assume $N = N(e_1), N' > 0, N'' < 0$ and $E = E(e_2), E' < 0, E'' < 0$. We first consider a time rate and then a piece rate each with imperfect monitoring.

Under the time rate, the firm pays a fixed wage w_0 and requires that the output level N_0. The worker effort for producing N_0 is e_1^0. The firm also monitors quality with a random inspection rate *m*. However, under the time rate, no fine is charged for poor quality. Instead, workers may receive warnings from the manager and understand that producing poor quality products will eventually lead to dismissal. So a risk neutral worker's utility function under the time rate takes the form:

$$(1) \qquad U = w_0 - C(e_1, e_2) + P(N - N_0) - mS(E) - G(E)$$

$C(e_1, e_2)$ is the worker's cost of effort function that measures the monetary value of the disutility associated with effort for quantity and quality, respectively. We assume $C_1' > 0, C_{11}'' > 0, C_2' > 0, C_{22}'' > 0$. $P(N - N_0)$ is the monetary value of the positive feedback (or sense of achievement) that a worker gets from producing output above the minimum required level ($P' \geq 0, P'' < 0$). This term allows individual heterogeneity, such as the intrinsic value of the job, to affect utility. So if $P(N - N_0) = 0$ *or* $N = N_0$, the worker derives no additional utility from producing above the minimum and simply adopts e_1^0. $S(E)$, (with $S'(E) \geq 0$) is the monetary value of disutility a worker feels when he/she is reprimanded by the manager for a fault rate of *E*. This can be interpreted as worker's fear of dismissal. $G(E)$, (with $G'(E) \geq 0$) is the monetary value of disutility associated with guilt for producing *E* proportion poor quality products (not checked and found by the manager).

[1] In our case, it will be the proportion of words entered with errors or the error rate.

This feeling arises as we assume that quality is not directly observable to outsiders. Hence, when the worker's output is not checked for quality only the worker knows how much effort was put into quality production. There remains a psychological cost due to guilt..

The first-order conditions for utility maximization are:

(2) $\frac{\partial U}{\partial e_1} = -C_1'(e_1, e_2) + P'N' = 0$

(3) $\frac{\partial U}{\partial e_2} = -C_2'(e_1, e_2) - mS'E' - G'E' = 0$

Or

(4) $C_1' = P'N'$

(5) $C_2' = -(mS' + G')E'$

Since both C_1' and C_2' are monotonically increasing with e_1 and e_2, respectively, (4) and (5) implies that e_1 (or quantity of output) will increase with P' and that e_2 (or quality of output) will increase with *m*, $S'$and G'other things equal. This allows the following:

Proposition 1. Under the time rate,

(1) The quantity of output is positively associated with the worker's sense of achievement or intrinsic value of the job;

(2) The quality of output is positively associated with the monitoring level provided a worker cares about reprimands from the manager. It also increases with the guilt a worker feels about producing poor quality.

Under the piece rate, the firm pays w_p per piece, and imposes a fine, fw_p, for every unit of output detected with low quality. Hence, the worker's expected fine is $w_p mfNE$ when the fault rate is *E*. The the worker's utility function takes the form:

(6) $\mathbf{U} = w_p N(1 - mfE) - C(e_1, e_2) - G(E)$

The first-order conditions under the piece rate are:

(7) $\frac{\partial \mathbf{U}}{\partial e_1} = w_p(1 - mfE)N' - C_1' = 0$

(8) $\frac{\partial \mathbf{U}}{\partial e_2} = -w_p N m f E' - C_2' - G'E' = 0$

Or

(9) $C_1' = w_p(1 - mfE)N'$

(10) $C_2' = -(w_p N m f + G')E'$

From (9) and (10), e_1 (or quantity of output) increases with w_f and decreases with *m* and *f* while e_2 (or quality of output) increases with *m*, *f* and G' other things being equal. So we have the following proposition.

Proposition 2.

(1) The quantity of output increases with the size of the piece rate but decreases with the monitoring level (or fine) for (poor) quality;

(2) The quality of output increases with the expected fine ($w_p mf$) as well as with the guilt a worker feels about producing poor quality.

By comparing the optimal decision a worker makes under different payment and monitoring regimes, we also derive the following propositions.

Proposition 3. A worker will produce more output under the piece rate if

$w_p(1-mfE) > P'$.

Proof. Assume the optimal effort level for quantity under time rate is e_1^t and that under piece rate is e_1^p. Since $C_1' > 0, C_{11}'' > 0$ and $N' > 0, N'' < 0$, if $e_1^p > e_1^t$ then $\frac{C_1'(e_1^p)}{N'(e_1^p)} > \frac{C_1'(e_1^t)}{N'(e_1^t)}$.

However, $\frac{C_1'(e_1^p)}{N'(e_1^p)} > \frac{C_1'(e_1^t)}{N'(e_1^t)}$ implies $w_p(1-mfE) > P'$.

Proposition 3 means that if a worker feels little of sense of achievement for producing more under the time rate ($P' = 0$), then he will produce more under piece rates than time rates other things being equal.

Proposition 4. A worker devotes more effort to quality under the piece rate than under the time rate if the expected fine for poor quality is high.

Proof. Assume the optimal effort level for quality under time rate is e_2^t and that under piece rate is e_2^p. Since $C_2' > 0, C_{22}'' > 0$ and $E' < 0, E'' < 0$, if $e_2^p > e_2^t$ then $\frac{C_2'(e_2^p)}{E'(e_2^p)} < \frac{C_2'(e_2^t)}{E'(e_2^t)}$.

However, $\frac{C_2'(e_2^p)}{E'(e_2^p)} < \frac{C_2'(e_2^t)}{E'(e_2^t)}$ implies $w_p mfN + G'(e_2^p) > mS'(e_2^t) + G'(e_2^t)$.

The above inequality is more likely to hold if the expected fine, $w_p mf$, is high.

Thus, we emphasize that the advent of the piece rate may bring dramatically different consequences depending on the size of the piece rate and of the expected fine (the monitoring rate times the actual fine). While it may be impossible for a worker to increase both quantity and quality when facing a binding constraint on effort, it is unlikely that this constraint is binding under the time rate. Thus, the advent of the piece rate with loose monitoring and so a low expected fine may increase quantity (proposition 3) while decreasing quality. Yet, the

advent of a piece rate with close monitoring can increase quality (proposition 4).

We now design an experiment with differences in monitoring rates to draw out these differing implications for quality and quantity. We also explore whether the heterogeneity in worker tastes for which we allow is reflected in the data. We test whether those who value the job and hold themselves to a high standard indeed have higher quality and quantity across regimes (proposition 1).

III. Experiment Design & Data

Experiment Design

We designed our experiment as an internship scheme offered to university students in Shenzhen, China. To provide essential workplace realism, we invited a real company to help us run the scheme[2], which required workers to input data input from questionnaires. The questionnaires used were from a real survey carried out for another research project. This method not only provided workplace realism, but also advantages in error checking since the data had already been input before. To provide additional motivation, students were told about the importance of the underlying social research project based on the questionnaires.

We first announced this internship opportunity through the University intranet with the help of the Student Service Centre. We advertised for 60 positions and received over 200

[2] This local Shenzhen IT company's main business was software design. We paid the company to operate the scheme for us. All the interns were nominally hired and paid by this company. It also issued the internship certificates to the students, which are important to fulfill a graduation requirement for internship experience. The actual job design and day-to-day operation was run by two research assistants who were titled managers. One of them was in charge of the supervision and monitoring of workers (line manager), and the other was a technical manager who mainly helped students deal with technical problems with data input. He was also responsible for calculating the error rates of input.

applications. All these applicants were then invited for two random draws. The first draw determined whether they were hired. The second then determined whether they were in the morning or afternoon shift. They remained in the shift for the entire period. Each shift lasted for 3.5 hours (from 8:30 to 12:00 in the morning and from 1:30 to 5:00 in the afternoon). To preserve realism, students participating in this internship were not informed of the experimental changes in conditions.

The internship was designed to run for 4 weeks, and with five days a week from Monday to Friday. In order to make sure that all interns started at same level, we reserved the first three days for training[3]. The experiment proper therefore started from Thursday in the first week and lasted for 17 working days.

Throughout out the entire period, the morning shift was strictly monitored for work errors, with 50 percent of workers randomly checked to see if errors were made on daily basis, and reprimanded accordingly. The afternoon shift was monitored loosely with only 10 percent of workers checked. Individual workers did not know that group monitoring rates differed. The units of observation are the number of words entered for each day and the number of errors. The errors were tabulated by computer programmed with the correct information for each questionnaire prior to the experiments[4].

The first two weeks (7 working days) were designed as the time rate period and all interns

[3] We originally planned to use three days, Monday to Wednesday, for training. However, there was unexpected power maintenance in the building on Wednesday and so no work could be undertaken on that day. Thursday was then used as another training day.

[4] These questionnaires had already been input by another firm for the other project. So we had those inputs (checked twice before) as our "corrected" input in the computer data base. A programme was then compiled to compute errors by comparing the new inputs with the existing "corrected" inputs.

were paid a fixed wage of 50 yuan a day. A minimum word input level was identified to the work groups as a satisfactory output, and no mistakes were meant to be permitted. If a worker was monitored (at random) and low output or mistakes were found, the worker was reprimanded for having performed in an "unsatisfactory" fashion. If neither the output nor mistake criteria was violated or they were not monitored, they were identified as having performed in a "satisfactory" fashion. As a further penalty, an indication was made to the workers that these "marks" would be retained. However, their compensation did not vary with the results of the monitoring.

During the second two weeks, each group was paid a piece rate (w_p) of 1.2 yuan per completed questionnaire. The monitoring rates (m) remained as before. The number of completed questionnaires (pieces) was recorded for each worker for each day, *N*, and formed the basis for the initial earnings for the day. The questionnaire entry errors were recorded for those monitored, and used to generate a quality index for each worker for each day: the average share of entries per questionnaire that are errors. This error rate, *E*, typically less than .05, served as the basis for the fine (*f*) for low quality, and we set *f=1.5*. Therefore, in the piece rate period, pay was calculated as: $w_p N(1 - mfE)$ where w_p = *1.2* yuan, *m=0.1* or 0.5, and *f=1.5*.

This formula sets the reward structure for quantity and quality. Thus, if a worker completed 50 questionnaires and had an error rate of .04, as monitored, her standardized output would be 47 for which she would receive 56.4 yuan. A worker who escaped monitoring would be paid for all completed entries as indicated by the number of questionnaires regardless of accuracy.

The consequence of the piece rate is that the relative reward of quality vs. quantity becomes

larger for the group more intensely monitored. The anticipated cost for an entry error is five times larger for the closely monitored (50% vs. 10%). In terms of our theoretical presentation, there is a little penalty for substandard output for those monitored less often and they should respond by increasing output. In essence, it becomes rational for the loosely monitored to engage in more opportunistic "cheating" (Nagin et al. 2002) by reducing quality in an effort to increase quantity. On the other hand, there should be less trading-off of quality for quantity (and potentially none) for the closely monitored.

Data

Figure 1a shows the average output per worker over the course of the internship. The early periods are prior to the advent of the piece rate. The output of the more loosely monitored group is routinely larger and the gap appears to grow with the advent of the piece rate. Figure 1b shows the error rate of the two monitoring groups. This is the actual error rate as derived by the computer and is independent of monitoring. The overall error rate indicates that between 3 and 4 percent of all entries are incorrect. Yet, the pattern on the error rate is less obvious. The rates seem similar prior to the advent of the piece rate but diverge afterwards with the loosely monitored having a higher error rate.

The basic descriptive statistics are summarized in the means presented in Table 1. In addition to the critical information on quantity and quality, we have data on workers' attitudes toward the internship including whether they found the job meaningful, and how difficult it was to find an internship. We see that most tend to find the job meaningful and about one-third claim it is difficult to find an internship, both of which attitudes we expect to link to less opportunism. We also asked for anticipated pay, and will use this as a measure of their outside options.

Following from the work of Nagin et al. (2002), we wish to examine the role of monitoring and the attitudes elicited from the questionnaire. We examine whether those workers who value the job tend to simply do a better job (both fewer errors and more output) as indicated by out theoretical illustration. We also examine how these attitudes interact with monitoring by exploring whether monitoring is more effective for workers who value the job.

IV. RESULTS

Before presenting the regression estimates, we simply average the pooled data by treating the repeated daily results of each worker as units of observations. We divide the data between the loosely and closely monitored and into the periods before and after the advent of the piece rates. This gives us four means as shown in the Table 2 and allows a basic difference in difference presentation. The upper panel, for error rates, shows that error rates are not affected by stricter monitoring in the time rate regime, but decline significantly more steeply (-0.32) in the piece rate period. The bottom panel measures output in total entries and shows that monitoring causes a large decline in output in the time rate regime, and even more in the piece rate, though the difference is not quite significant (-0.8).

Put another way, our initial results suggest that stricter monitoring given time rates is counter-productive: it reduces quantity but with no improvement in quality. On the other hand, stricter monitoring given piece rates is at least potentially productive: it reduces quantity, but improves quality. We summarize these means in Figure 2, which identifies the four combinations of quantity and quality associated with the different monitoring and payment regimes. The three regimes across the top indicate combinations that are not dominated - we see that a choice of high monitoring and time rates would never be profitable. However, depending on the true value to the firm of quantity and quality, any of these three other

alternatives might logically be adopted. The dashed line also shows the practical trade-off between quality and quantity.

Continuing to use the repeated daily results of each worker as units of observation, we estimate the daily error rate and the daily output as functions of the difference in monitoring rates (high or low) and the piece rate (before and after). This exercise is similar to Table 2, except we now cluster the errors by worker to recognize that we are observing repeated draws that could bias the standard errors. The results are shown in the first two columns of Table 3, and are similar to Table 2.

The next two columns in Table 3 recognize that although individual workers were allocated randomly, they may have different abilities and preferences that determined their error rate and output. To account for this we estimate worker fixed effect estimates of the error rate and output. Such estimates hold constant all worker specific influences including the monitoring rate to which the worker was allocated. Thus, these estimates provide the influence of the piece rate separately for those in the two monitoring groups. The error rate estimate is very similar to the pooled estimate, and to Table 2. It continues to demonstrate that the advent of the piece rate causes the error rate of the low monitoring rate group to increase and the error rate of high monitoring rate group to decrease. As for quantity, again the pattern conforms to the pooled results, and the simple picture of Table 2.

Several basic points emerge from these estimates. First, monitoring not tied to explicit rewards or penalties has little influence on quality but negatively influences productivity. This could be consistent with the view that workers find supervision not tied to rewards simply demotivating. If they view monitoring as disproportionately about quality, they may maintain it, but reflect the lack of motivation in lower quantity. Second, the advent of the

piece rate system generates different responses by group since monitoring is now explicitly linked with rewards. While the loosely monitored group responded with the classic quantity rather than quality, the closely monitored group held quantity and actually increased quality. This fits with our theoretical illustration which emphasizes that when the expected penalty is large enough, quality can increase with the advent of a piece rates.

We now try to identify some of the individual characteristics that might have been subsumed in the fixed effect estimates. We control for sex, age and whether or not the students are business majors. While the first two are of general interest, the last could potentially be important as such majors value practical internship experience. We also control for three dimensions of past experience and/or personality. We identify whether the individuals consider it difficult to find an internship, whether they find the work meaningful and whether they expected to earn more than they currently earn (those who may see themselves as "overqualified"). We anticipate that those who value their current job more because it is meaningful or alternatives are hard to find, will perform better on both quantity and quality independent of monitoring and piece rates. We later test whether or not there are differences in these estimates by regime.

Table 4 presents the pooled results again clustering for errors by worker. The first estimate shows the estimate of the error rate as a function of regime and our new control variables. The first point is that the pattern isolated earlier remains. Imposing strict monitoring reduces the output but not the error rate. The advent of the piece rate remains associated with a significant increase in the error rate for those loosely monitored, as well as a significant increase in output. The second point is that our general hypothesis about the value of the job and the nature of personal characteristics seems to be reflected in the estimates. Thus, those who major in business, those who find the job intrinsically meaningful and those who

think it is difficult to find an internship all have significantly fewer errors than their otherwise equal counterparts. While we recognize that these indicators measure different dimensions of the worker, they share the characteristic that they reflect the importance of holding the job. Either it is important for the worker's education, for itself or because there are few other opportunities to earn money and/or fulfill the requirement. Thus, we see the greater accuracy of these workers as reflection of their valuation of the job and perhaps their desire to keep it and have it reflect well upon them.

Working the other way is the indicator of potential over-qualification - whether the worker anticipated earning more. Such workers might put less value on the job as they see their realistic alternatives as better. This indicator is associated with a significantly higher error rate. There are no differences by age but males tend to have a higher error rate.

We now turn to the estimate of output. The three measures we saw as associated with value of the job again emerge as important but in the opposite direction, so that both quantity and quality increase – there is no tradeoff. Thus, finding the job meaningful was associated with a reduced error rate and it is associated with nearly 700 additional words per day in output. Similarly, business majors and those who think it is difficult to find an internship each have higher quantity and quality than their counterparts. However, the indicator of over-qualification shows a tradeoff: the piece rate brings a higher error rate with higher output. Overall, the pattern from the earlier estimates regarding regimes remains in place.

We next turn to how these characteristics interact with differences in monitoring[5]. We focus on the three of the variables that seemed important in the earlier estimates: whether the job is

[5] None of the interactions with the advent of the piece rate were statistically significant in either the error rate or the output estimates

meaningful, whether it is difficult to get an internship and the anticipated pay. For each of these three variables we estimated the error rate and output with the specification in column 1 and 2 of Table 3 but adding interactions.

The estimates show the familiar patterns regarding the basic role of regimes but the interaction of anticipated pay (over-qualification) with high monitoring emerges as important in the error rate estimation. In the loosely monitored group, those who feel overqualified continue to show higher error rates (0.0001) than their otherwise equal counterparts, but this reaction reverses (-0.0004) significantly among the closely monitored group. Thus, among those being closely monitored, those who feel overqualified actually have a lower error rate. The importance of this characteristic carries over to the output estimates, which show that while those who feel overqualified more produce more than their counterparts when loosely monitored, their advantage in output increases when closely monitored. Thus, it appears that this group of workers is highly responsive to monitoring as it produces both more quantity and quality – the tradeoff moves outwards.

Neither of the other two attitude characteristics emerges as dramatically in the estimates but it is noted that those who think it is difficult to find an internship produce greater output but only under the closely monitored regime. This might be thought to fit with an efficiency wage notion. They recognize their alternatives are likely inferior and want to keep the current position and so respond when monitoring increases (see Drago and Heywood 1992). Finally, we note that men in basic pooled estimate produced insignificantly different output than did women. Yet, this hides a pattern by monitoring. Men produce more than women when loosely monitored but less than women when tightly monitored.

V. Conclusions

Our field experiment has generated several findings. First, monitoring that was not attached to monetary rewards (or fines) had no impact on quality but actually depressed quantity (Figure 2). This we take to be evidence that monitoring by itself may demotivate workers and potentially reduce intrinsic incentives to be productive.

Second, the advent of a piece rate brought with it more powerful rewards and penalties, which matter. The loosely monitored group, with a lower expected fine for poor quality responded with a large increase in output and a decrease in quality (a higher error rate). This we take to be the stereotypical response that has received widespread attention. The closely monitored group responded with both an increase in output and in quality (a lower error rate). This we take as evidence of Lazear's (1995) point that rewards for quality (or penalties for low quality) can be part of piece rates, and that as a consequence it should not be assumed that quality will be sacrificed for quantity. It is this point that our theoretical development illustrated. Importantly, our experimental application is one in which the expected penalty increases because of increased monitoring. Thus, the combination of a penalty and closer monitoring serves to encourage improved quality with the advent of the piece rates.

Finally, we emphasize that worker attitudes appear important independent of the payment scheme. Workers who value the job are typically observed producing more and having higher quality. Understanding this is potentially important for management who may wish to screen for such workers when hiring. At the same time, those who felt the internship should pay more (perhaps overqualified) emerged as more opportunistic in sacrificing quality for higher output. Perhaps obviously, it is also this type of person who needs close watching. Less obvious is the fact that this type, when closely monitored, actually increases both quantity and quality.

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

Variables	Mean	Standard Deviation
Error Rate (%)	3.55%	0.0091
Output (000 words inputted per day)	13.51	4.27
Male (1=yes)	0.45	0.50
Age	22.4	0.91
Business major*	0.90	0.30
Difficult to find internship (1=yes)	0.36	0.48
Job is meaningful (4=very meaningful; 0=not meaningful at all)	3.41	0.79
Anticipated Pay (yuan per hour)	18.8	10.2
Piece rate (1=yes)	0.58	0.49
High monitoring (1=yes)	0.50	0.50

Definitions:

Error rate – the daily number of incorrect entries per questionnaire as a share of all entries per questionnaire, *E*

Output – total number of questionnaires inputted a day (*N)* typically multiplied by the average number of words inputted per questionnaire

Business major - the interns are either from business school or from engineering school. So the omitted group is engineering major.

Anticipated Pay – in practice, pay turned out at between Y14 on time rates, and Y20 per hour (piece rates, low monitoring).

High monitoring – dummy=1 for workers with a 50% monitoring rate and =0 for workers with a 10 percent monitoring rate.

Table 2: Basic Difference-in-Difference Results for Monitoring and Piece Rate

A. Error rate per word typed (%)- averages

		Piece rate payment			
		No	Yes	Diff (Yes – No)	
Monitoring	Low	3.54	3.69	0.15*	
	High	3.54	3.37	-0.17	
	Diff (High -Low)	0.00	-0.32**	-0.32**	D. in D.

B. Output (000 words per day) - averages

		Piece rate payment			
		No	Yes	Diff (Yes – No)	
Monitoring	Low	14.2	15.9	1.7**	
	High	11.3	12.2	0.9	
	Diff (High - Low)	-2.9**	-3.7	-0.8	D. in D.

Table 3: Errors and Output under Different Regimes

	Error Rate (Pooled)	**Output (Pooled)**	**Error Rate FE**	**Output FE**
Constant	.0354 (47.4)	14240.8 (51.67)**	.0354 (96.2)**	12783.5 (82.7)**
High Monitor	.0003 (0.28)	-2946.0 (8.01)**		
Piece Rate	.0015 (1.70)*	1668.8 (4.40)**	.0015 (2.22)**	1634.4 (5.72)**
High Monitor x Piece Rate	-.0032 (2.57)**	-771.7 (1.56)	-.0027 (2.78)**	-792.5 (1.97)**
R-squared	.015	.183	.014	.081
N	991	991	991	991

***-significant at 10 percent **- significant at 5percent or better**

In all pooled regressions, robust clustered errors are calculated, with observations clustered by individual.

Table 4: Errors and Output – the Role of Characteristics

	Error Rate	Output	Error Rate (characteristics by monitoring)	Output (characteristics by monitoring)
Constant	.0393 (5.69)**	1594.1 (5.26)**	.0459 (6.22)**	8770.3 (2.88)**
High Monitor	.0003 (0.25)	-2818.3(7.71)**	.0030 (0.88)	-2057.7 (1.47)
Piece Rate	.0015 (1.80)*	1669.8 (4.49)**	.0015 (1.81)*	1672.4 (4.52)**
High Monitor x Piece Rate	-.0030 (2.50)**	-828.1 (1.71)*	-.0031 (2.610**	1396 (2.61)**
Business Major	-.0024 (2.37)**	1376.4 (3.86)**	-.0025 (2.40)**	1803.5 (4.49)**
Job is Meaningful	-.0018 (4.72)**	695.7 (4.34)**	-.0022 (4.06)**	592.5 (2.73)**
Job Meaningful x High Monitor			.0013 (1.56)	-416.9 (1.29)
Difficult to find Internship	-.0014 (2.43)**	437.1 (1.65)*	-.0010 (1.27)	-172.1 (0.44)
Difficult to find x High Monitor			-.0009 (0.71)	1396.5 (2.61)**
Anticipate Pay is Higher	.0001 (2.96)**	54.47 (4.92)**	.0001 (4.24)**	48.86 (3.93)**
Anticipate Pay x High Monitor			-.0004 (4.76)**	68.78 (2.23)**
Male	,0013 (2.11)**	371.9 (1.38)	.0015 (1.67)*	1431.4 (3.55)**
Male x High Monitor			.0004 (0.34)	-2424.4 (4.65)**
Age	.0001 (0.45)	-288.6 (2.19)**	-.0002 (0.48)	15.89 (0.12)
R-squared	.083	.222	.105	.243
N	991	991	991	991

*-significant at 10 percent **- significant at 5 percent or more

Robust errors clustered on individuals are calculated for all regressions.

Figure 1a: Output over the Internship

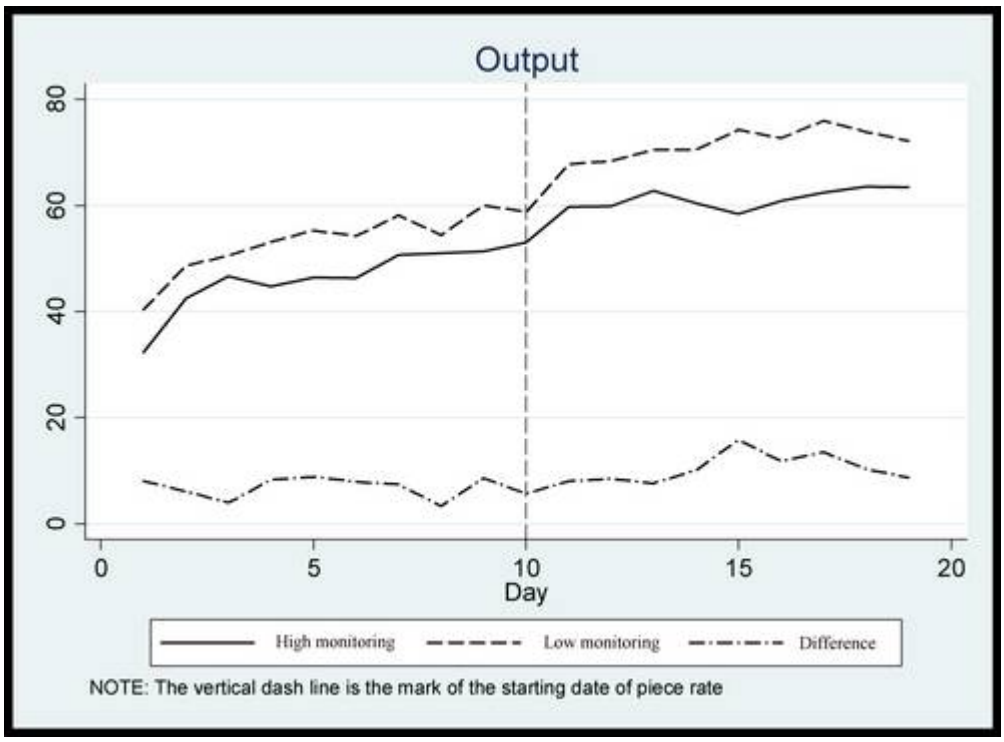


Figure 1b: Error Rate over the Internship

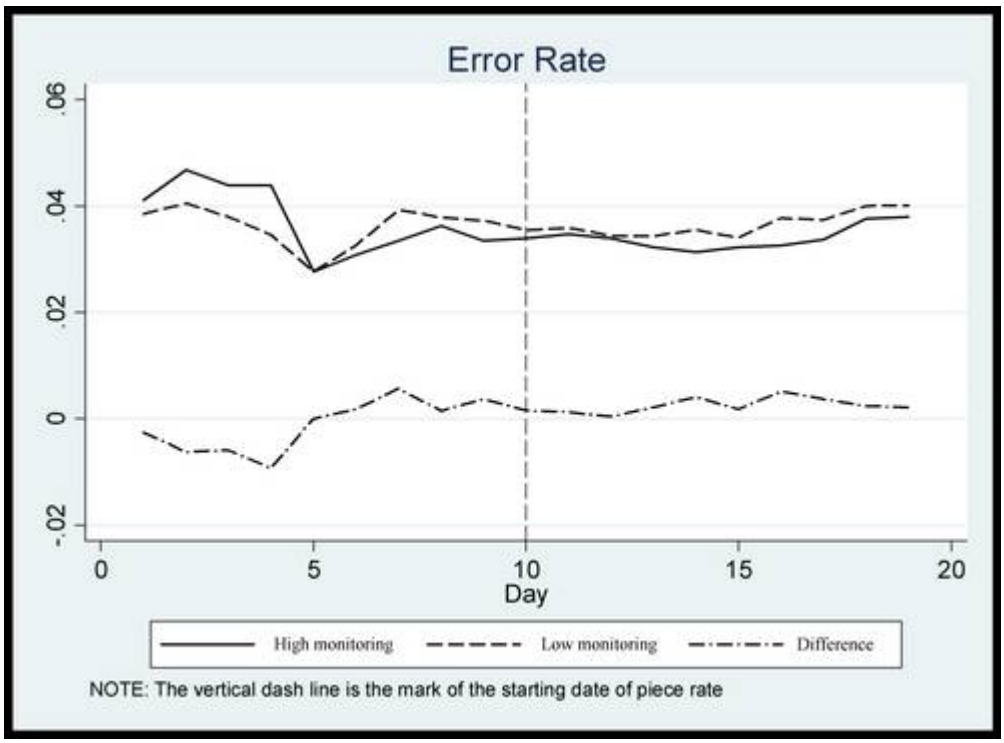
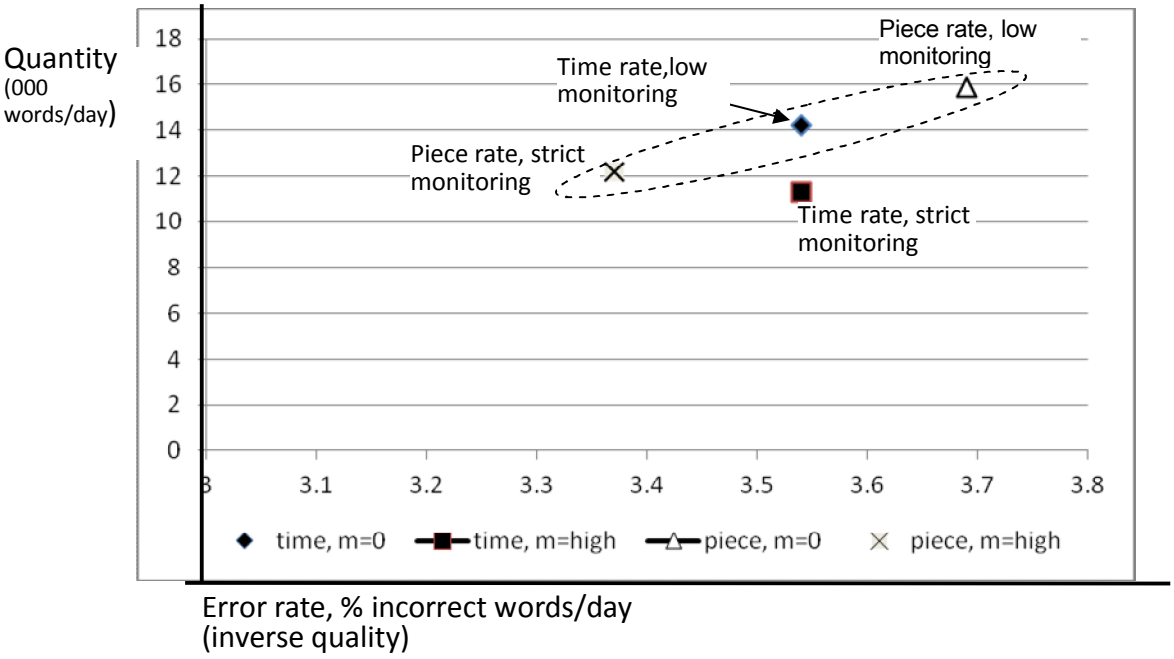


Figure 2: Quantity-Quality Results by Payment System and Monitoring



The four points are our observations. As can be seen, we have a trade-off given by the 3 top points. Clearly a time rate with high monitoring (bottom square) is useless. Hence, the firm will choose between the 3 top points shown according to the importance of quantity vs. quality in its profit function.

References:

Drago, Robert and John S. Heywood (1992). "Is Worker Behavior Consistent with Efficiency Wages?" *Scottish Journal of Political Economy* 39(2): 141 – 53.

Freeman, Richard B. and Morris M. Kleiner (2005). "The Last American Shoe Manufacturers: Decreasing Productivity and Increasing Profits in the Shift from Piece Rates to Continuous Flow Production," *Industrial Relations* 44(2): 307 – 33.

Guilteras, R.P. and B.K.Jack (2012). "Incentives, Productivity and Selection in Labor Markets: Evidence from Rural Malawi." Mimeo.

Holmstrom, Bengt and Paul Milgrom (1991). "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design" *Journal of Law, Economics, & Organization*, Vol. 7, Special Issue: [Papers from the Conference on the New Science of Organization, January 1991] (1991), pp. 24-52.

Lazear, Edward P. (1986). "Salaries and Piece Rates." *Journal of Business,* 59(July): 405–31.

Lazear, Edward P. (1995). *Personnel Economics*, MIT Press.

Lazear E.P. 2000. "Performance pay and productivity," *American Economic Review* 90(5):1346–1361

Paarsch, Harry and Bruce Shearer. 2000. "Piece Rates, Fixed Wages, and Incentive Effects: Statistical Evidence from Payroll Records," *International Economic Review* 41: 59 – 92.

MacDonald G, Marx L.M. 2001. "Adverse Specialization," *Journal of Political Economy* 109(4):864–899

Nagin, D.S., Rebitzer, J., Sanders, S., and Taylor, L. (2002)."Monitoring, Motivation and Management: The Determinants of Opportunistic Behavior in a Field Experiment." *American Economic Review* 92: 850:872.

Schearer, Bruce (2004). " Piece Rates, Fixed Wages and Incentives: Evidence from a Field Experiment**, "** *Review of Economic Studies* 71: 513 -34.