

IZA DP No. 7660

The Consequences of a Piece Rate on Quantity and Quality: Evidence from a Field Experiment

John S. Heywood
Stanley Siebert
Xiangdong Wei

September 2013

The Consequences of a Piece Rate on Quantity and Quality: Evidence from a Field Experiment

John S. Heywood

University of Wisconsin-Milwaukee

Stanley Siebert

*University of Birmingham
and IZA*

Xiangdong Wei

Lingnan University

Discussion Paper No. 7660
September 2013

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

Any opinions expressed here are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions. The IZA research network is committed to the IZA Guiding Principles of Research Integrity.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit organization supported by Deutsche Post Foundation. The center is associated with the University of Bonn and offers a stimulating research environment through its international network, workshops and conferences, data service, project support, research visits and doctoral program. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ABSTRACT

The Consequences of a Piece Rate on Quantity and Quality: Evidence from a Field Experiment*

This field experiment examines output quantity and quality for workers in a data input business. We observe two sets of workers that differ in monitoring intensity as they move from time to piece rates. The application of piece rates increases quantity, and we find that the resultant quality can be improved with sufficient monitoring. “Committed” workers also produce higher quantity and quality, showing the role of worker selection - which appears especially strong under time rates. Our results thus show how a firm can refine its worker selection and monitoring options together with the payment system to deliver its chosen quality-quantity combination.

JEL Classification: D2, J3, L2, M5

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

Corresponding author:

Stanley Siebert
Birmingham Business School
University of Birmingham
Birmingham, B15 2TT
United Kingdom
E-mail: w.s.siebert@bham.ac.uk

* We gratefully acknowledge funding support from the joint ESRC-RGC grant (RES-000-22-3653), and the excellent research assistance of Dr. Jia Wu and Ms. Yufei Zheng. We are also grateful for comments from Imran Rasul, Andrew Seltzer, and participants at the 7th Biennial 2012 Conference of the Hong Kong Economic Association, and participant's at the EU FP7 project e-Frame's Human Capital and Labour Markets 2013 Workshop.

NON-TECHNICAL SUMMARY

Our paper reports research from a field experiment into the effects of incentive pay, worker monitoring, and worker commitment on the quantity versus quality aspects of worker performance. The basis for the study is data input which poses well a common problem that the more that is produced, the lower the quality (errors) – the quality-quantity “tradeoff”. This tradeoff is a widespread fear in business, and is widely thought to be worsened by incentive payments, which typically, it is said, reward quantity but overlook quality. Our results do not accord with this argument at all.

In our experiment, quantity is measured in terms of 1000’s of words input from survey questionnaires per day, and quality is measured as the number of mistakes. (Mistakes are almost perfectly measured, because the questionnaire data had already been input and checked for another project, so computer comparison of the new with the checked data was possible.) 60 workers were tracked for a month, and separated randomly into high monitored (50% chance of being checked per day) and low monitored (10% chance) groups. Initially the groups were paid according to a time rate, then halfway through the period, the system was changed to a piece rate, with high returns for quantity and at the same time a fine for mistakes (the expected fine naturally being larger for the highly monitored group). At the same time the workers’ motivations for doing the work were carefully surveyed with the aim of measuring worker “commitment”, e.g., whether they found the task meaningful.

We find that worker monitoring, the piece rate’s high-powered incentives, and worker commitment all influence a worker’s quality-quantity performance. Hence, by choosing appropriate combinations of these variables, many different quality-quantity combinations are available to the firm. It is true that these different combinations have different costs, for example high monitoring requires supervisors to be paid, while high worker commitment implies costly worker selection. Nevertheless, our results show that a firm is far from being bound by a given quality-quantity tradeoff. It can refine its worker selection and monitoring options together with the payment system to deliver a chosen (optimum) quality-quantity. There is no reason to fear piece rates, provided that these are combined appropriately with worker monitoring. We also show that time rates, for their part, best combine with careful worker selection policies, and not with strict monitoring which seems simply to demotivate in this context. In all, a rich set of results that are important for personnel economics.

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.” We believe this last point deserves more emphasize in the literature and so we conduct field experiment which introduces 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 and quality.

Despite frequent discussion in the theoretical literature and the obvious relevance to business practitioners, there are few experimental studies of the advent of a piece rate that rewards both quantity and quality. Experiments with random assignment provide a critical method for overcoming the criticism that quality and quantity are endogenous and reflect worker sorting, as emphasized by Shearer (2004). Structural models provide one avenue as Paarsch and Shearer (2000) model the firm's choice of a piece rate versus time rate and show that under piece rates workers can respond by increasing quantity at the expense of quality. More directly, 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. Shi (2010) randomly assigns workers doing tree thinning to an hourly wage or a piece rate. Those on the piece rate produced greater output but did not decrease quality. The implicit penalty in this scheme was that sampled trees not meeting a quality standard were required to be redone but there is no variation in the monitoring intensity. 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 study they 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 be 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) and provides a series of robustness tests. A final section concludes.

II. An Illustrative Theoretical Model

We examine 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. The quantity of output the worker produces is N and the

measure of quality is E , the proportion of N with quality below a pre-defined standard¹. 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 output N_0 . The worker effort for producing N_0 is e_1^0 . The firm monitors quality with inspection rate m . However, under the time rate, no fine is charged for poor quality. Workers receive warnings from the manager and understand that producing poor quality will eventually lead to dismissal. A risk neutral worker's utility function takes the form:

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

$C(e_1, e_2)$ 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 ($P' > 0, P'' < 0$). This term allows individual heterogeneity 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 when the worker is reprimanded for low quality (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). This arises because even when the worker's output is not checked for quality, there remains a psychological cost for failure to exert the required effort.

The first-order conditions for utility maximization are:

¹ In our experiment it will be the proportion of words entered with errors or the error rate.

$$(2) \frac{\partial U}{\partial e_1} = -C'_1(e_1, e_2) + P'N' = 0 \text{ or } C'_1 = P'N'$$

$$(3) \frac{\partial U}{\partial e_2} = -C'_2(e_1, e_2) - mS'E' - G'E' = 0 \text{ or } C'_2 = -(mS' + G')E'$$

Since both C'_1 and C'_2 monotonically increase with e_1 and e_2 , these imply that e_1 (quantity of output) increases with P' and that e_2 (quality of output) increases with m , S' and G' other things equal. This allows the following:

Proposition 1: Under the time rate,

- (1) The quantity of output increases with the worker's sense of achievement or the intrinsic value of the job;
- (2) The quality of output increase is 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.

We now consider a piece rate in which the firm pays w_p per piece, and imposes a fine, $f w_p$, for every unit detected with low quality. Hence, the worker's expected fine is $w_p m f N E$ when the fault rate is E . The worker's utility function takes the form:

$$(4) \quad U = w_p N (1 - m f E) - C(e_1, e_2) - G(E)$$

The first-order conditions under the piece rate are:

$$(5) \frac{\partial U}{\partial e_1} = w_p (1 - m f E) N' - C'_1 = 0 \text{ or } C'_1 = w_p (1 - m f E) N'$$

$$(6) \frac{\partial U}{\partial e_2} = -w_p N m f E' - C'_2 - G'E' = 0 \text{ or } C'_2 = -(w_p N m f + G')E'$$

These imply that e_1 (quantity of output) increases with w_p and decreases with m and f while e_2 (quality of output) increases with m , f and G' other things being equal. So we have the following proposition.

Proposition 2: Under the piece rate

- (1) The quantity of output increases with the size of the piece rate but decreases with the monitoring level (or fine) for low quality;
- (2) The quality of output increases with the expected fine ($w_p mf$) and with the guilt a worker feels about producing poor quality.

Critically, we wish to compare quantity and quality under the time rate and piece rate.

Proposition 3: A worker produces 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 says that if a firm sets a high enough piece rate it will induce a worker to produce more output under piece rate than under time rate. It also implies that if a worker feels little of sense of achievement for producing more under the time rate ($P' = 0$), the piece rate results in greater output than the time rate other things equal.

Proposition 4. A worker devotes more effort to quality under the piece rate than under the time rate for a sufficiently large expected fine.

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 when the expected fine, $w_p mf$, is large.

Thus, we emphasize that the advent of the piece rate need not bring a trade-off between quantity and quality. Indeed, both can increase depending on the size of the piece rate and of the expected fine.² It is this point we think deserves greater emphasis. The trade-off that receives so much attention in the literature reflects the advent of the piece rate with loose monitoring and so a low expected fine may increase quantity at the expenses of quality, yet this need not be the only case.

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. Here, 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 for university students in Shenzhen, China. To provide essential workplace realism, we invited a real company to help us run the scheme³, which required workers to input data from questionnaires. The questionnaires used were from a real survey carried out for another research project. This

² 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.

³ This local Shenzhen IT company's main business was software design and data processing. 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.

method not only provided workplace realism, but also almost-perfect error checking since the data had already been input before, and could therefore be computer-checked⁴. 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 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. The experiment proper therefore started from Thursday in the first week and lasted for 17 working days. One student dropped out from the internship after three days due to family emergency. Since he left after the internship already started we did not find a replacement. So we had 59 students who completed the internship (29 in the morning and 30 in the afternoon shift). We also lost twelve observations due to sick leaves so this left us with an unbalanced panel of 991 observations (shifts x people).

Throughout out the entire period, the morning shift was strictly monitored for errors, with 50

⁴ Because the data had already been input by another firm for the other project, 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.

percent of workers randomly checked 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. As noted above, the errors were tabulated by a computer programmed with the correct information for each questionnaire.

The first two weeks (7 working days) were designed as the time rate period and all interns were paid a fixed wage of 50 yuan a day. Workers were given a minimum word input level as satisfactory output and were told all mistakes were to be avoided. If a worker was monitored (at random) and low output or mistakes were found, the worker was reprimanded for having “unsatisfactory” performance. If output and mistakes were satisfactory or they were not monitored, they were identified as having "satisfactory" performance. 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 daily earnings. The questionnaire entry errors were recorded for those monitored, and used to generate a quality index for each worker for each day: the 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, 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 diverges afterwards with the loosely monitored having a higher error rate.

The basic descriptive statistics are summarized 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 simply to do a better job (both fewer errors and more output) as indicated by our 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 thus 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, Panel A, and are similar to Table 2.

The next two columns in Panel A of 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 from Table 2. In sum, the OLS and FE results are practically identical, showing no signs of non-random allocation of workers into high or low monitored groups.

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.

Before moving to explore the role of differences in worker characteristics, we describe a robustness check for these initial results. It may be argued that the influence of monitoring is felt only by the individual worker being checked and that the proper way to capture the influence of monitoring is then to control for the influence on a specific worker being checked rather than for the monitoring characteristics of the entire experimental group. Thus, instead of controlling for the group monitoring intensity which reflects differences in expected fines, we could simply control for whether the given worker (under either degree of monitoring) has been checked. With the relatively small numbers of an experiment, these individual experiences may differ from those of the group average. To examine this we develop an alternative measure to high or low monitoring, which is the proportion of days (up to and including the current day of observation) in which an individual worker has been monitored. We call this the monitoring frequency and repeat our analysis using this actual measure in the place of the monitoring group identifier.

The lower panel of Table 3 presents estimates designed to mirror those already discussed. In most respects, the pooled results paint a similar picture to those using the group identifier. Under the time rate, those workers with a higher check frequency generate no fewer errors in the pooled estimate but produce significantly less. This continues to suggest that the time rate with high monitoring is unproductive. The piece rate increases both errors and output as in the earlier specification and, indeed, the size of both increases is actually larger when the monitoring frequency is higher. This is confirmed as the interaction of the piece rate with the check frequency is negative and highly significant in both the error and output estimations. This supports the previous result that only under the piece rate does higher monitoring intensity reduce errors and that it does so at the cost of reduced output.

A series of projections can be made using the estimates in the first two columns and recognizing that the average probability of the frequency check among those with low monitoring is .1 and that among those with high monitoring is .5. These projections are shown in Figure 3 and they largely mimic those shown in Figure 2. They make evident the reduction in both output and errors as one moves from low monitoring frequency with time rates to low monitoring frequency with piece rates and, finally, to high monitoring frequency with piece rates. High monitoring with time rates remains outside this apparent trade-off with both the lowest output but a still very high error rate.

For completeness we show the fixed effect variant of the estimates using the monitoring frequency measure. As the monitoring frequency actually varies within a worker's observations, we can estimate the influence of the variation in frequency on output and errors. The errors increase with monitoring frequency under the time rate while they decrease with monitoring frequency under the piece rate. The piece rate itself now appears to play no direct role. The output estimate continues to suggest that the piece rate is associated with

greater output although the usual decrease associated with higher monitoring frequency under the piece rate just misses statistical significance.

The somewhat different pattern of the fixed effect estimate from the ordinary least squares may suggest the greater importance of individual heterogeneity. However, the difference may also reflect less precise or accurate estimates. The changes in monitoring frequency within worker are necessarily modest in size suggesting measurement error is important in the fixed effect estimates. More confounding is that the distribution of the changes in the monitoring frequency within worker are not random, they are disproportionately at the beginning of the panel. For example in the first period all frequencies are necessarily either zero or one and then begin to move toward their group means. Finally, we note that the original group indicators actually fit the underlying error and output data better as measured by the explained variation. This at least hints that there may be information associated with the group indicator. Workers may base their own expected chance of being monitored not only on their own past experience but on that of their coworkers. Hence, we continue to rely more heavily on the results using the group monitoring indicator.

Worker Heterogeneity

We now try to identify some of the individual characteristics that might affect quantity or quality irrespective of the payment system (the G function in equation 1). 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 monitoring⁵. We focus on the three of the variables that seemed important in the earlier estimates: whether the job is 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 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.

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

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. ADDITIONAL RESULTS

In this section we undertake a series of projections and explore the determinants of an exit survey given to all workers at the end of the internship. The projections are based on the estimates from the third and fourth columns of Table 4 and are designed to further isolate the consequences of the personal characteristics on quality and quantity. We make projections for each of the four regimes based on monitoring intensity and time vs. piece rates. Thus, the projections become points in the same space for which we plot the actual data in Figure 2. We make one set of four projections for those workers we identify as “committed.” These workers find the job meaningful, think it is difficult to find an internship and have anticipated earnings one standard deviation below the median. We make a second set of four projections for those workers we identify as “uncommitted.” These workers do not find the job meaningful, think it is easy to find an internship and have anticipated earnings one standard deviation above the median.

The wide locus to right in Figure 4 shows the projections for the uncommitted workers (white symbols). The two projections to the right of this locus show the extremely high error rates of those with low monitoring intensity. These low monitoring outcomes with uncommitted

workers can be contrasted with those with low monitoring among committed workers (black symbols) shown as the two projections to the right of the narrow locus on the left. Thus, the move from uncommitted to committed workers is associated with a large reduction in errors under either time or piece rates. Moreover, this change is also associated with increases in output. It seems clear that searching for committed workers is highly beneficial when monitoring is low or costly.

We now consider the projections associated with high monitoring. Here the projections associated with uncommitted workers are the two left most among the wide locus on the right. These are contrasted with the committed workers under high monitoring as shown by the two projections to the farthest left. As shown, the move from uncommitted to committed workers is associated with far more modest reductions in error rates and with small reductions in output as well. Here the benefit of searching for committed workers is modest at best, when monitoring is high, and may be nonexistent (as quantity reduces). We take this to suggest that hiring committed workers is extremely important when monitoring is loose but of far less value when monitoring is strict. Put differently, the results highlight the importance of the costs of hiring/selection vs. the cost of monitoring. As Figure 4 shows, monitoring and selection of committed workers are typically substitutes. Indeed, the quantity and quality produced by committed (time rate) workers with low monitoring is virtually identical to that produced by uncommitted (piece rate) workers with high monitoring. Thus, the investment in hiring or monitoring would reflect the relative cost.

A second implication of the projections is the importance of piece rates if either of the two dimensions of output thoroughly dominates the other in the objectives of the firm. Thus, the projection with the highest output is that with a piece rate and no monitoring combined with committed workers. The projection with the lowest error rate is the piece rate with high

monitoring combined with committed workers. This point will likely be the most costly for the firm to attain since management will have to incur both worker selection and monitoring costs. Interestingly, an intermediate quantity-quality position would be achieved either by time rates with careful worker selection and no monitoring (since monitoring demotivates time workers - Figure 4, black triangle), or piece rates with strict monitoring but no selection (Figure 4, white circle). This position might well be the cheapest for the firm, so we would expect the combination of piece rates plus monitoring, or time rates plus careful selection to be popular business options empirically.

We now consider an exit questionnaire which asked workers at the end of their internship whether all things considered they preferred working under the piece rate system or the time rate system. The vast majority of workers preferred the piece rate system, 77% of all workers. Table 5 presents probit estimates of the determinants of preferring the piece rate system. The first column shows that those who prefer piece rates are males, those who find the work meaningful, those who thought it is difficult to find an internship and those with higher anticipated pay and higher actual pay. The latter two may well be measures of ability and it makes sense that the more able who benefit most from a piece rate will prefer it. It is less obvious why those who find the work meaningful and thought it hard to find an internship prefer the piece rate once their pay is held constant. It strikes us that these workers who tend to produce more at a higher quality may simply value the piece rate as it rewards their commitment and performance.

We expanded the specification by including the error rate for each worker and whether they were in the loosely or tightly monitored group. As shown in the second column, the error rate did not emerge as a statistically significant predictor but the monitoring intensity plays an enormous role. Those in the tightly monitored group are substantially less likely to prefer

the piece rate. The marginal effect shows that their likelihood of preferring the piece rate is over 13 percentage points lower, all else equal, than those in the loosely monitored group. As the financial consequences of errors are so much larger under the piece rate for the strictly monitored, they are presumably likely to associate both earnings risk and greater attention to accuracy with the piece rate and find it less desirable even with earnings held constant.

VI. 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. To generate high performance from workers on time rates (see Figure 4's white circle) it seems best to spend on selecting workers carefully, rather than monitoring strictly.

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 characteristics and attitudes bear importantly on the quality-quantity outcome. Measured worker attitudes emerged as significant determinants independent of both the payment scheme and strictness of monitoring. Workers who value the job are typically observed producing more and having higher quality. Understanding this is potentially important for management who must consider the costs of screening for such workers when hiring. The highest quality is likely to be most expensive, and in our results comes from requiring a piece rate system and both careful worker selection and monitoring. Nearly as good is either a time rate system, combined with careful worker selection, or piece rates coupled with strict monitoring. As these combinations are cheaper, we expect them to be popular options with firms. Put another way, our results show how a firm can refine worker selection and monitoring options, together with the payment system to deliver its chosen quantity-quality combinations, and we highlight the rough ranking of the associated costs.

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
Monitoring frequency	0.30	0.24
Prefer piece rate (1=yes)	0.77	0.42
Daily wage	57.90	12.26

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.

Monitoring frequency – measures the days for which a worker has been checked for quality as a proportion of the days worked to that date

Prefer piece rate – dummy=1 for those workers who answered that they preferred piece rate to time rate after the internship.

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*	D. in D.
	High	3.54	3.37	-0.17	
	Diff (High -Low)	0.00	-0.32**	-0.32**	

B. Output (000 words per day) - averages

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

Table 3: Errors and Output under Different Regimes

	Error Rate (Pooled)	Output (Pooled)	Error Rate FE	Output FE
Panel A. Monitoring intensity measured by the High Monitor dummy				
Constant	.0354 (47.4)**	14240.8 (51.67)**	.0354 (96.2)**	12783.5 (82.7)**
High Monitor	.0003 (0.28)	-2946.0 (8.01)**	time invariant	time invariant
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
Panel B. Monitoring intensity measured by the Monitoring frequency				
Constant	.0346 (42.48)**	14132.4 (49.57)**	.0355 (39.61)**	12760.5 (33.81)**
Monitoring frequency	.0030 (1.39)	-4653.9 (7.04)**	.0026 (3.21)**	78.1584 (1.50)
Piece Rate	.0022 (2.16)**	2403.6 (5.64)**	-.0005 (.19)	1641.2 (4.87)**
Monitor frequency x Piece Rate	-.0076 (2.78)**	-3643.5 (3.55)**	-.0080 (3.74)**	-1347.4 (1.51)
R-squared	.010	.165	.051	.041
N	991	991	991	991

***-significant at 10 percent **- significant at 5 percent 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 interacted with monitoring)	Output (characteristics interacted with 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.

Table 5. Probit Estimates of Preference over Piece Rate

	Model I (coefficients)	Model I (Marginal effects)	Model II (coefficients)	Model II (marginal effects)
Constant	-6.1999 (1.24)		-5.7015 (4.71)***	
Job is Meaningful	.6884 (10.25)***	0.1590	.7265 (10.40)***	.1631
Difficult to find Internship	.2943 (2.50)**	.0651	.0907 (0.83)	.0201
Anticipate Pay	.0376 (5.12)***	.0087	.0368 (4.97)***	.0083
Daily Wage	.0507 (7.73)***	.0116	.0471 (6.88)***	.0106
Male	.4090 (3.56)***	.0928	.4501 (3.62)***	.0991
Age	.0442 (0.85)	.0102	.0508 (0.97)	.0114
High monitoring			-.5838 (5.81)***	-.1307
Error rate			-5.7448 (0.98)	-1.2896
Pseudo R ²	.233	.233	.259	.259
N	983	983	983	983

*-significant at 10 percent **- significant at 5 percent ***-significant at 1 percent

Robust errors clustered on individuals are calculated for all regressions.

Figure 1a: Output over the Internship

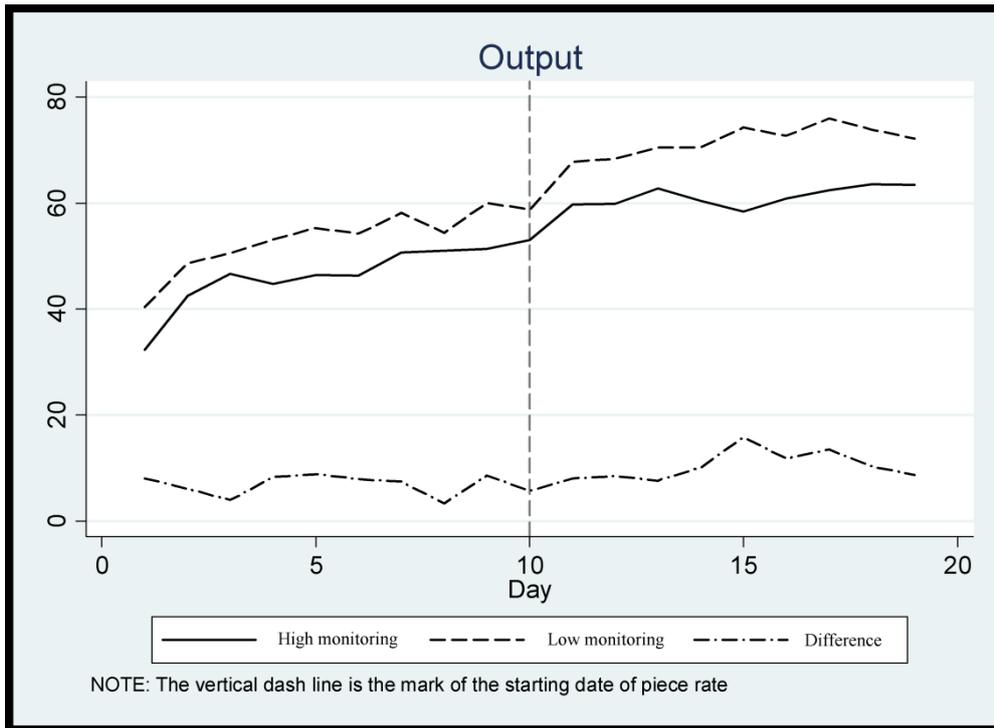


Figure 1b: Error Rate over the Internship

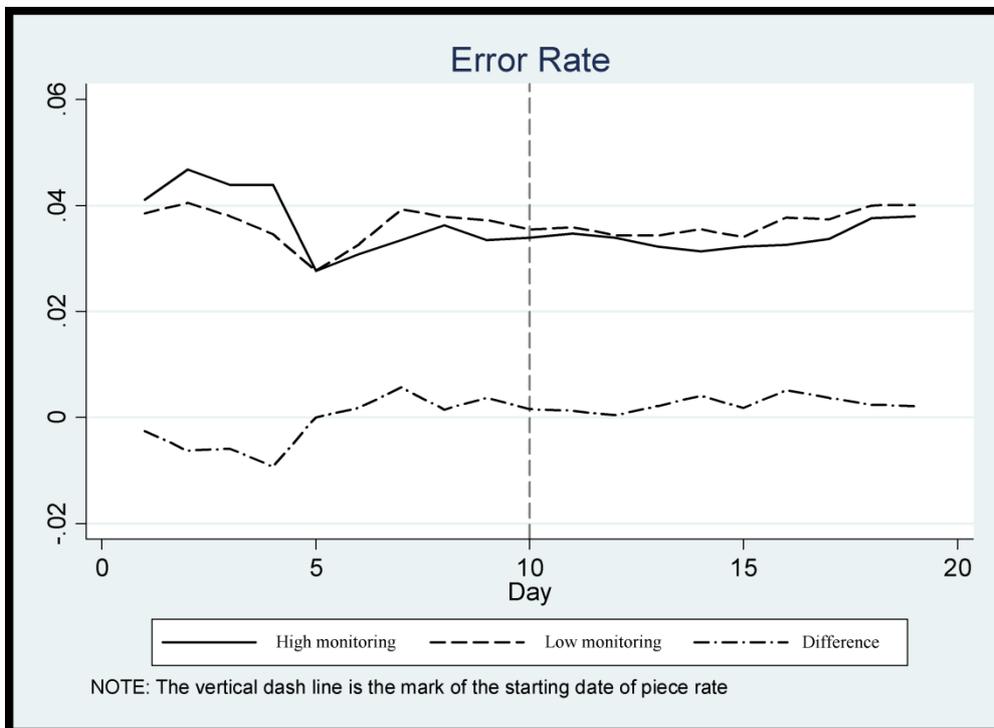
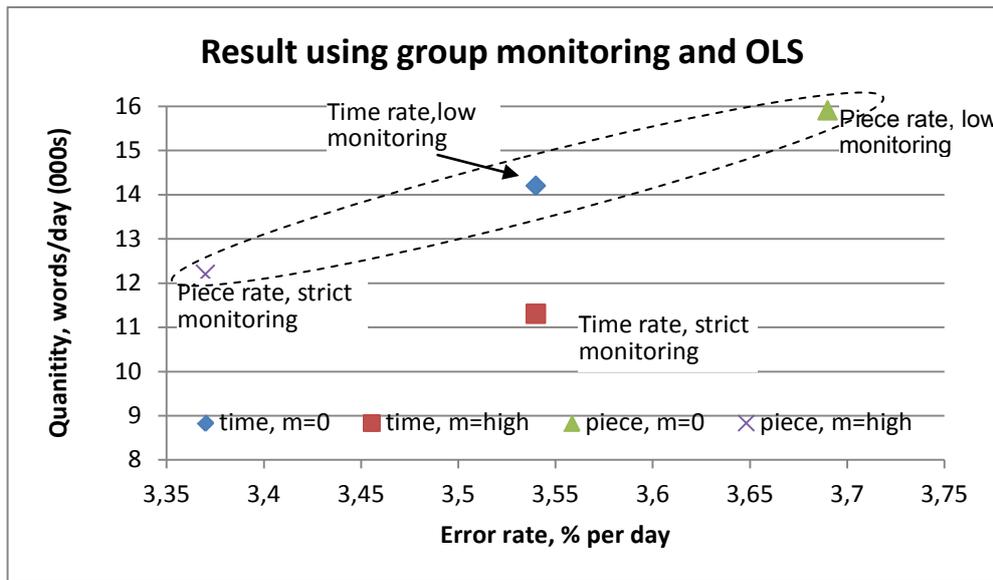
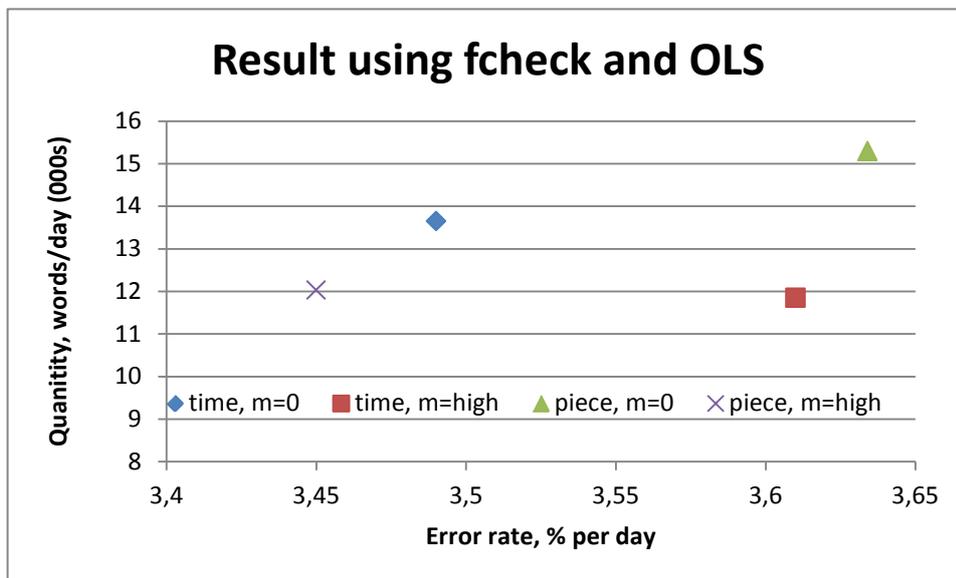


Figure 2: Quantity-Quality Results by Payment System and Monitoring



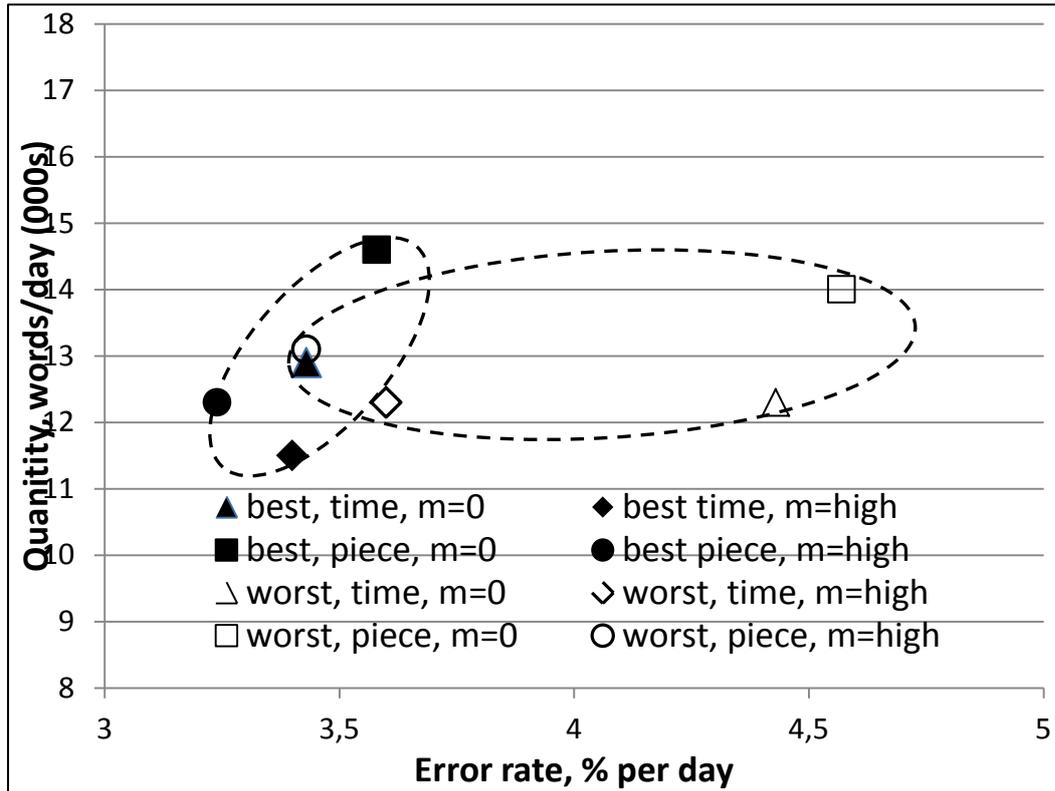
Notes: The four points are our observations. As can be seen, we have a trade-off given by the 3 top points. Clearly a time rate with high monitoring (bottom square) is simply dominated by that with low monitoring. Hence, the firm will choose between the 3 top points shown according to the importance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

Figure 3: Quantity-Quality Results using Monitoring Frequency and OLS



Notes: These points represent projections from the pooled OLS equations in Panel B of Table 3. The average monitoring frequency for the low monitoring group is taken as $f_{check} = 0.1$ and that for high monitoring as $f_{check} = 0.5$.

Figure 4 Projections for Committed and Uncommitted Workers



Notes:

These points represent projections from Table 4 equations with interactions. Committed workers are shown as “best” in the left locus with black symbols; uncommitted workers are shown as “worst” in the right locus with white symbols.

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.
- Shearer, Bruce (2004). " Piece Rates, Fixed Wages and Incentives: Evidence from a Field Experiment, " *Review of Economic Studies* 71: 513 -34.
- Shi, Lan (2010). "Incentive Effect of a Piece Rate Contract: Evidence from Two Small Field Experiments," *B.E. Journal of Economic Analysis and Policy* 10 (1).