ETLA Working Papers

No. 19

24 September 2013

Takao Kato
Antti Kauhanen
Essi Kujansuu

THE PERFORMANCE EFFECTS OF INDIVIDUAL AND GROUP INCENTIVES

A Case Study

Suggested citation: Kato, Takao, Kauhanen, Antti & Kujansuu, Essi (24.9.2013). "The Performance Effects of Individual and Group Incentives: A Case Study".
ETLA Working Papers No 19. http://pub.etla.fi/ETLA-Working-Papers-19.pdf

ETLA ELINKEINOELÄMÄN TUTKIMUSLAITOS
THE RESEARCH INSTITUTE OF THE FINNISH ECONOMY

ISSN-L 2323–2420
ISSN 2323–2420 (print)
ISSN 2323–2439 (online)

The Performance Effects of Individual and Group Incentives: A Case Study

Takao Kato[a], Antti Kauhanen[*b], Essi Kujansuu[b]

[a] Department of Economics, 13 Oak Drive, Colgate University, Hamilton, NY 13346, USA

[b] Research Institute of the Finnish Economy (ETLA), Lönnrotinkatu 4B, 00120 Helsinki, Finland.

Abstract

We study a warehouse that changes its incentive plan from a group plan to an individual plan. We focus on the impact that the change had on productivity, allocation of time to different tasks, and helping across departments. Utilizing time series methods we find that average productivity increased about 15 % but that individuals reacted heterogeneously to the change. The change also affected the organization of helping across departments and the allocation of time to various tasks. In addition to the econometrics analysis we provide a description of the process of changing the incentive plan and discuss the cost effectiveness of the change of the plan.

[*] Corresponding author. Tel. + 358 9 609 90224, Fax + 358 9 601 753. Address: Lönnrotinkatu 4 B, 00120 Helsinki, Finland. E-mail addresses: tkato@colgate.edu (T. Kato), antti.kauhanen@etla.fi (A. Kauhanen), essi.kujansuu@gmail.com (E. Kujansuu). Funding from Tekes—the Finnish Funding Agency for Technology and Innovation—is gratefully acknowledged.

Introduction

Incentive plans affect not only the level of effort, but also the allocation of effort to different tasks. Incentive plans that do not properly measure all aspect of employee performance may lead to distorted actions. The important role of performance measurement for the success of incentive plans has led to a growing empirical literature on performance measurement (e.g. Gibbs, Merchant, Van Der Stede, & Vargus, 2009; Ittner & Larcker, 2002; Kauhanen & Napari, 2012).

The role of performance measurement has received less attention in the literature considering productivity effects of incentive plans. Many studies have considered the productivity impact of a shift from time rates to piece-rates (Lazear, 2000; Paarsch & Shearer, 2000; Paarsch & Shearer, 1999; Shearer, 2004) or to some other types of incentive plan (Banker, Lee, & Potter, 1996; Kauhanen, 2011; Knez & Simester, 2001).

Fewer papers have considered the productivity impact of changing from one type of incentive plan to another. Bandiera, Barankay, and Rasul (2005) study a shift from relative incentive plan to piece-rates. They find that individual performance is notably higher in the piece-rate regime. Levenson, Zoghi, Gibbs, and Benson (2011) show how changing from pure commission based pay to base wage and steeper commission rate increased productivity of store delivery workers. Owan and Tsuru (2011) study a car dealership that changes the incentive system from salary and commission system to guaranteed base wage and steeper commission. They find that productivity of senior worker increases but that junior workers performance is unaffected. They conjecture that this is due to strong career incentives.

In addition to productivity effects, the literature has studied how incentives shape the allocation of effort between tasks. Drago and Garvey (1998) find that piece-rates and profit-sharing do not affect self-reported amount of helping, but that promotion incentive do. Dumont, Fortin, Jacquemet, and Shearer (2008) study a shift from fee-for-service payment system for medical doctors to a partial fee-for-service. They find that lowering the intensity of incentives led to lower effort on the measured tasks and higher effort on the tasks that were not rewarded under the original incentive plan. Levenson et al. (2011) also find that time allocation is affected by changes in incentives.

We conduct a case study of a firm that alters its incentive plan. A notable feature of this change is that performance measurement changes from group level measurement to individual level on the main tasks of some employees. This has implications for both the level of effort and allocation of effort to activities that are rewarded individually and those subsidiary tasks where performance is measured at a group level.

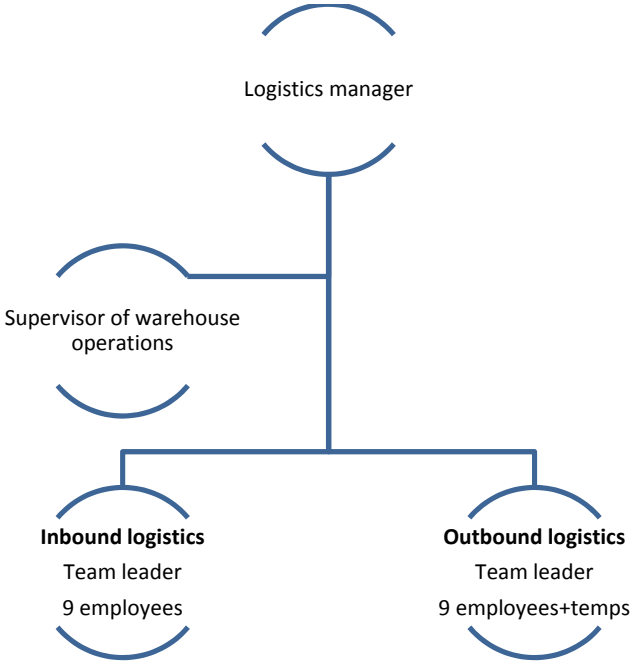
We use individual level time series data to study the impact that change in the performance measurement had on productivity and allocation of time to various activities. We find that the change increased average productivity by about 15 %. There is however, substantial individual heterogeneity in the impact. Some individuals did not respond to the plan at all and for some the change had a negative effect. We also find that the allocation of time to the main task *decreased* while the time spent on subsidiary tasks *increased*. This way the employees were able to maintain high average productivity in the main task. The change also affected the organization of helping in another department. Prior to the change most employees helped occasionally, but after the change this became the responsibility of a single individual.

In addition to the econometrics analysis we provide a description of the process of changing the incentive plan and discuss the cost effectiveness of the change of the plan.

Case description

The case firm is a supplier of healthcare and laboratory equipment, supplies and services. We focus on the logistics operations of their central warehouse. The warehouse under study has a surface area of 12 500 m^2 and the number of pallet locations is 12500. The firm has 20000 products in total of which they stock 7000 continuously. The warehouse has two main departments: inbound and outbound logistics. The warehouse is led by a logistics manager who is helped by a supervisor of warehouse operations. Both departments have a team leader. Figure 1 shows the organization chart.

Figure 1 Organization chart



Inbound logistics deals with receiving of goods. The work process starts with unloading of pallets from trucks to the goods reception area. There are two persons who specialize in unloading of trucks. Other employees may help them as there are a few employees who can operate large forklifts. In the goods reception area the pallets are unpacked and the goods labeled. Once these operations are done, the goods are shelved either to long-term storage or short-term storage depending on the demand for the goods in question.

Inbound logistics is characterized by teamwork as all the employees are working on the same shipment. Task assignment is decided by a team leader and the supervisor of warehouse operations. Many of the employees working in the inbound logistics have previously worked in outbound logistics and are thus able to help there if needed. If there is need for help, the inbound and outbound logistics team leaders will coordinate help in consultation with the supervisor.

The inbound logistics department handles about 8500 purchase orders per year. Each order consists of about 5.8 order lines. An order line describes the number of units of a particular product. In physical quantities they handle about 12000 pallets and 130000 cartons per year. Due to team production performance indicators are only available at group level. The performance measures are

order lines received per hour and share of goods shelved in 48 hours. These measures are basically quantity measures as measurement of quality is difficult. An example of a quality problem would be shelving a product to an incorrect location. The goals for the quantity measures are 7.5 order lines per hour and 69 % of goods shelved in 48 hours[1].

Outbound logistics is about picking and shipping goods according to customer orders. This is individual work with little production externalities. The process starts with the team leader and supervisor creating order sheets. Each picker comes to the team leader's office and picks up an order sheet consisting of all order lines for one customer. The pickers cannot choose the order sheets themselves as some of them are easier than others to pick. Some orders have only one order line and some have over ten order lines (6.35 order lines per order on average).

The task of the pickers is to find each order line, pick up the right quantity of each line, and put them together for each customer. Some of the picking is done with forklifts and some is done manually using a cart. Most of the goods are in regular shelves, but some are in cold storage, some in long-term storage and then there are some special items whose picking is more involved. Once the items are boxed they are transferred to the dispatching area. Large quantity orders are wrapped in the dispatching area.

Outbound logistics picks about 55000 orders and 350000 order lines per year. The performance measures are number of orderliness picked per hour and quality. Quality is measured as correct order lines per total order lines. Incorrect order line can be due to wrong product being shipped or incorrect quantity (the most typical case). Another way faulty shipping can take place is that the whole order is shipped to a wrong place. The target for the quantity performance measure is 6 order lines per hour and for quality the target is 99.5 %. Both of these measures are available at the individual level.

The personnel policies were very traditional prior to July of 2010. The supervisors were in charge and employees had little opportunities to participate. Communication between management and floor level employees did not work very well. Starting in July 2010 more participatory practices were adopted. The employees were involved more in the day to day development of operations and new

[1] Hours worked are measured as the total working hours in the department in a given month.

weekly meetings were initiated. The meetings take place every Tuesday at noon and last about 15-20 minutes. In the meetings the logistics manager reviews last week's performance records and discusses the week's performance goals. In these meetings the employees are also informed of other developments that are pertinent to workers. At the same time monthly personal feedback sessions were initiated, where individual performance is discussed.

Change in the incentive plan

In the fall of 2010 the firm started to plan changing its incentive system. The compensation system for the employees consisted of two parts: hourly base wage that is stipulated by the collective agreement for trade industry and an incentive component that was unilaterally decided by the firm. The same compensation structure had been in place for more than five years.

The key features of the original incentive plan were as follows. The incentive payments depended on three performance measures: number of order lines picked per hour, quality (correct order lines per total order lines), and breakage (number of items damaged). All of these were measured at the warehouse level at monthly frequency and hours were measured as the total working hours.

The level of payments was determined in the following way. For each of the measures stepwise targets were set. First a range of acceptable level of performance was determined and then this range was divided to e.g. five equal bins. Thus, performance was rated on a scale of e.g. 1-5 for each measure. If performance was below acceptable range, the measure did not contribute to the bonus in that month. The scores for each measure were added together and this total score showed how large the bonus would be. It was expressed as a percentage of the hourly base wage. In a typical month, the bonus was around 5-6 % of income for the floor level employees.

The problem with the incentive plan was that the employees did not consider it motivating. A particular problem was the warehouse level performance measurement. The employees considered it risky, especially the pickers, since other employees actions affected the measure. This was particularly the case with the quality measure, since one shipment shipped to a wrong place would mean that they would not be rewarded for quality. The management also thought that the money spent on the plan could be used better.

To develop the incentive system, management decided that the employees should be involved in the design of the performance measures. They organized several meetings where three to four representatives of the employees participated in addition to the regular management. The aim was to make the plan more personalized.

The final strategy agreed on by the development team has different plans for outbound and inbound logistics. The performance measure for the inbound logistics is the quantity of order lines processed. Due to the team based nature of the work, performance is measured at the departmental level. In outbound logistics performance is measured by quantity and quality. Both of these are measured at the individual level. The quantity measure is a weighted sum of three types of picking: cart, forklift, and special (including cold storage, long-term storage, and special items). The weights are based on historical averages of order lines picked per hour. This weighting was adopted due to fairness reasons: the employees were concerned that some employees would gain more in the new system if they for example specialized in forklift picking.

The distribution of payments in the new plan was also different. Since the employees may work in both departments, an individual's bonus is calculated based on the hours worked in each department. In each month the hours worked and output produced in each department is calculated for the individuals. As the performance measurement in the inbound logistics is at the group level, the payments based on that measure are the same percentage of base pay for all employees. For the hours worked in the outbound logistics, the bonus percentage is based on individual performance. An individual's bonus percentage is thus a weighted average of the percentages of outbound logistics (common percentage to all) and inbounds logistics (individual percentage).

In the meetings several options for the plans were discussed before the final plan was accepted. There were six important points discussed. First point was the differences in different types of picking. Even though the job design of the pickers is very simple, fair performance measurement turned out to be quite difficult due to the heterogeneity in types of picking. Productivity in e.g. special picking is lower than in other forms of picking. This mattered because some individuals specialized in different forms of picking. Developing weights that reflect the productivity differences and were accepted by

everyone took considerable effort. This was done by calculating average productivity from data for the different activities.

Second, there was a lot of discussion about unincentivized tasks. There are some tasks that have to be performed but which are not recorded anywhere. An example would be the maintenance of the cleanliness of the picking and dispatch area. Cardboard boxes and other litter left on the floor affects the productivity of other employees. There was a worry that these types of tasks would receive less attention in an individual incentive plan.

Third, a lot of time was spent on discussing and researching the possibility of individual performance measurement in inbound logistics. There were ideas of reorganizing the work to make this possible. In the end it was decided that the nature of work is such that individual performance measurement does not make sense. Fourth, the supervisor tried to figure out ways to measure quality in inbound logistics. In the end, no feasible measures were found. Fifth, there were many practical problems with the IT-systems. Generating the reports that are needed for individual level measurement required combining reports from two distinct programs. It took many weeks to write a visual basic code that did the matching of the reports. Moreover, new time-stamping codes had to be developed to measure working time more accurately. Sixth, the division of payments generated debate. This partly reflected the fact that in the beginning, the possibility of individual measures in inbound logistics was entertained. However, even if that had been possible, the employees in that department strongly supported a collective plan.

For employees mainly working in the inbound logistics, the change in the incentive plan was modest in the sense that the plan was still collective. For the employees in outbound logistics, the change was much more substantial. Their plan changed from a collective plan to an individual plan. The performance measures were pretty much the same, but the level of performance measurement differed.

Conceptual framework

This section discusses the challenges of performance measurement from a theoretical perspective and develops predictions on the impact of the change of the incentive plan on the actions of the employees in outbound logistics.

Even simple jobs such as picking of goods in the warehouse involve multiple tasks. The picker's main task has two dimensions quantity (order lines picked per hour) and quality (number of mistakes). The performance measurement system has to account for both of these, with proper weights, in order not to generate distorted effort allocation (Baker, 2002; Holmstrom & Milgrom, 1991). It is important to note that proper weighting of these measures is essential. If for example the weight received by quality is too low compared to its value for the firm, the employees may emphasize quantity too much. The change in the plan of the current case firm does not lead to sharp predictions concerning quality. There are also tasks, which are not measured at all (like maintaining cleanliness of the warehouse), and such tasks are likely to get too little attention unless the supervisors monitor the effort put to these tasks.

Helping of other workers is also a part of the job design of pickers. Incentive pay also affects the marginal value of helping others (see e.g. Itoh, 1991). Individual incentives discourage helping, since they increase the opportunity cost of helping. Again, if the performance measurement system does not give adequate weight for helping, too little effort will be directed towards that task. In practice, it is difficult to calculate the optimal weights for different measures. Consider for example the formulas of optimal weights on measures in a multitask setting in Holmstrom and Milgrom (1991): they contain many variables that are difficult or impossible to estimate. Thus, it is likely that helping will decrease following the change in the incentive plan.

The performance measures may also cause distortions if they measure performance at an organizational level that is incompatible with the job design. For example, measuring performance at an individual level when team work is important will lead to distortions (Baker, 2002). However, at the same time measurement at the team level leads to a riskier performance measure, since more uncontrollable events (from the point of view of an individual) affect the measure. The optimal level of aggregation depends on the trade-off between risk and distortion (Baker, 2002).

Team level performance measurement tends to lower the incentive effects. This is due to the fact that each employee bears the full marginal cost of effort, but receives only a fraction of the marginal benefit. The employee does not internalize the positive externality that her effort generates. For this reason we expect that the change to individual level performance measurement should increase the

level of effort. However, this need not necessarily be the case since the free rider problem can be mitigated by the employee's engaging in mutual monitoring (Kandel & Lazear, 1992). This means that the team members monitor each other's level of effort and exert some kind of peer pressure on those who are deem to underperform. Such monitoring is likely to most effective when the team is relatively small and homogeneous (Kandel & Lazear, 1992). Che and Yoo (2001) show that in repeated interaction, paying for team output generates both means and incentives for mutual monitoring.

Any incentive plan may have heterogeneous effects. This is due to differences in ability (see e.g. Lazear, 2000). One way to interpret differences in ability is that the marginal cost of effort increases faster for lower ability individuals. For example, when facing a linear incentive plan, individuals with more convex cost of effort functions react less to the slope (i.e. the strength) of incentives (see e.g. Holmstrom & Milgrom, 1987 section 5). Thus we would expect the change in the plan to have heterogeneous effects.

To summarize, we would expect the change in the incentive plan to increase productivity but with heterogeneous effects. We would also expect to see more effort put to the tasks that are rewarded individually and thus to see the time allocated to helping reduced. With respect to quality the theoretical arguments are not clear.

Despite a large literature showing positive productivity effects of incentive pay[2] , only a few papers have considered their effect on profitability (Freeman & Kleiner, 2005; Kauhanen, 2011; Lazear, 2000). None of the papers have paid attention to the costs of running an incentive plan. Performance measurement is costly and maintaining and updating the incentive plan takes management's time. To pass cost-benefit analysis, the expected benefits should exceed the costs. The costs of implementation of management practices in general is one reason for their slow diffusion (Bloom & Van Reenen, 2007). Estimating the cost is often feasible, which makes it possible to figure out what the benefits of the plan should be to make it profitable.

[2] See e.g. Bandiera et al. (2005); Bandiera, Barankay, and Rasul (2007); Banker et al. (1996); Freeman and Kleiner (2005); Jones and Kato (1995); Lazear (2000); Paarsch and Shearer (2000); Paarsch and Shearer (1999); Shearer (2004).

Data and methods

The main data source is a task level dataset that is collected by the case firm and forms a basis for the pay for performance calculations. The dataset comprises the type, time, and duration of tasks done by 42 employees. The number of order lines completed during the interval (a measure of output) is also reported for certain, namely picking tasks. The dataset spans over 87 weeks starting from April 2011 and ending November 2012. The system was updated in the beginning of June 2012. Using this data we have assembled a weekly panel data of productivity and time allocation measures for the analysis. Aggregating the data for *the average day of the week* improves its time series properties by mitigating the influence of outliers and the dichotomical nature of some variables.

Data is collected under 20 different task classifications. We have grouped them into three categories (inbound logistics, outbound logistics and other tasks) in accordance to which department is primarily responsible for them. There are four tasks assigned to outbound logistics and three tasks assigned to inbound logistics. There rest are grouped under 'other tasks' and include assignments that are management related (1 task), regularly done by one or two persons only (5 tasks), or tasks that are performed occasionally by most of the employees such as training, meetings, and project work (7 tasks).

In addition to the main dataset we use smaller, monthly collected data on quality, where quality is measured as the percentage of correct order lines out of all order lines completed. The dataset covers the time period from January 2012 to November 2012. Quality data is given only for key outbound logistics tasks (1 - 2 tasks out of 4) and for those employees that worked at these tasks during the month. This dataset is not as detailed as the first dataset, but gives an opportunity to inspect how well potential (adverse) consequences of performance pay can be kept in check.

Included employees

We are mostly interested in employees that spend more than half of their time doing tasks in the outbound logistics. We call this group *the pickers* and it consists of 11 employees. The data available is most detailed for pickers as their performance is measured and rewarded individually. In the analysis we consider developments in their individual records as well as in the departmental average. To

investigate the effects of the new performance pay system and the change, the analysis is constrained to only those employees that work both before and after the system update.

Employees in managerial positions or performing special tasks lack comparable productivity measures or performance measures entirely and hence cannot be subject to the same kind of analysis. Likewise, goods receivers lack individual productivity measures for inbound logistics tasks. We have, however, detailed data on their time allocation between tasks. Several workers in both in- and outbound logistics are cross-trained and occasionally help out in the other department. Hence we can investigate if the renewed incentives have influenced helping behavior between the departments.

Productivity and quality

The main task of pickers is to collect the ordered items for dispatching. Hence productivity is simply a measure of lines collected in an hour and quality is percentage of correct lines out of all lines collected. Productivity data is available for four different picking tasks while there is quality data only for some. The analysis does not differentiate between the different picking tasks although there are considerable differences in the average efficiency levels. It is presumed that the mix of tasks remains stable on average[3]. The PRP system takes these differences into account with historically determined weights.

To start the analysis, Figure 2 shows average weekly productivity. Average productivity is the unweighted mean of individual productivities of pickers that worked the particular week, comprising at most 11 workers[4]. It is seen from Figure 2 that average productivity grew over the study period. The new incentive scheme was implemented the last day of week 22/2012. This point in time is represented by a vertical line in the graph below. The figure suggests that the new PRP plan has potentially had a positive impact on the average productivity of the outbound logistics department.

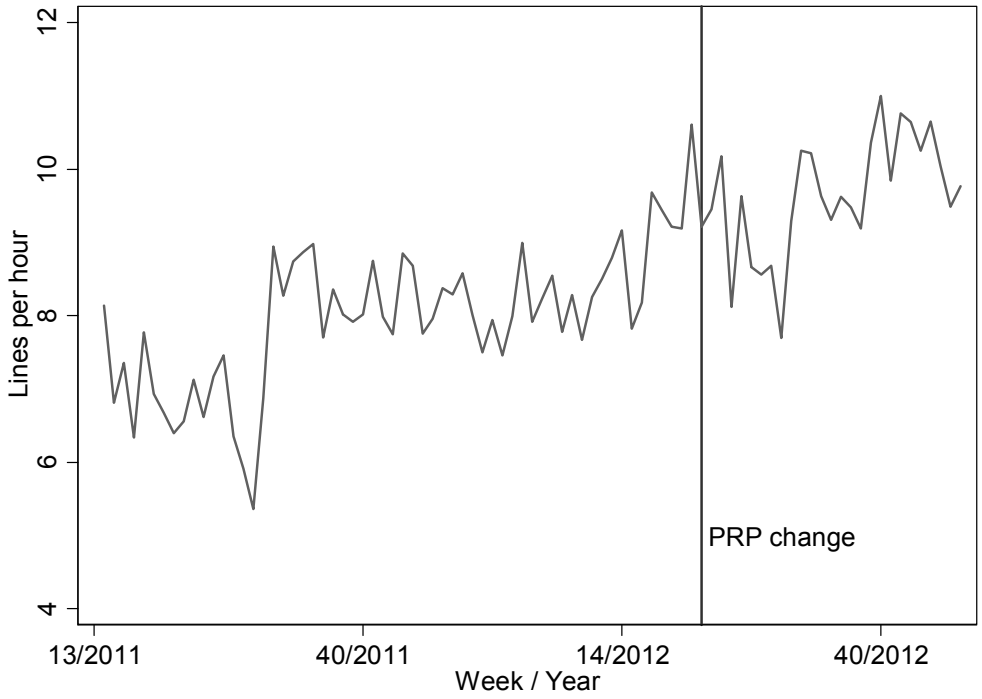
[3] Graphical inspection of mean productivity conditioned on the task mixture suggests that the assumption is reasonable. Task mixture does not seem to explain the observed growth in productivity as such but there is some indication of specialization or changing task mixtures, although they would indicate lower productivity rather than higher (suggesting gains from specialization).

[4] Note that the figures here a larger than the 6 lines per hour target. This discrepancy is explained by the fact that the target is measured by total lines in a month divided by total hours in a month. Here the hours are only a subset of the total hours whereas the lines picked are measured in a similar way to the target.

The figure shows to other features that are important for the analysis. First, there is another structural break in data at week 31/2011 due to the renewal of the warehouse floor layout. The development work took place in summer 2011. We take this into account by imposing a dummy variable for the old layout and the restructuring period, i.e. for the first 18 weeks.

Second, there are also two short term reductions in average productivity that both coincide with summer – approximately weeks 24/2011 – 33/2011 and 25/2012 – 34/2012. We take this potential negative impact of summer holidays and temporary summer substitutes into account by introducing a dummy variable for summer.

Figure 2 Average productivity, pickers



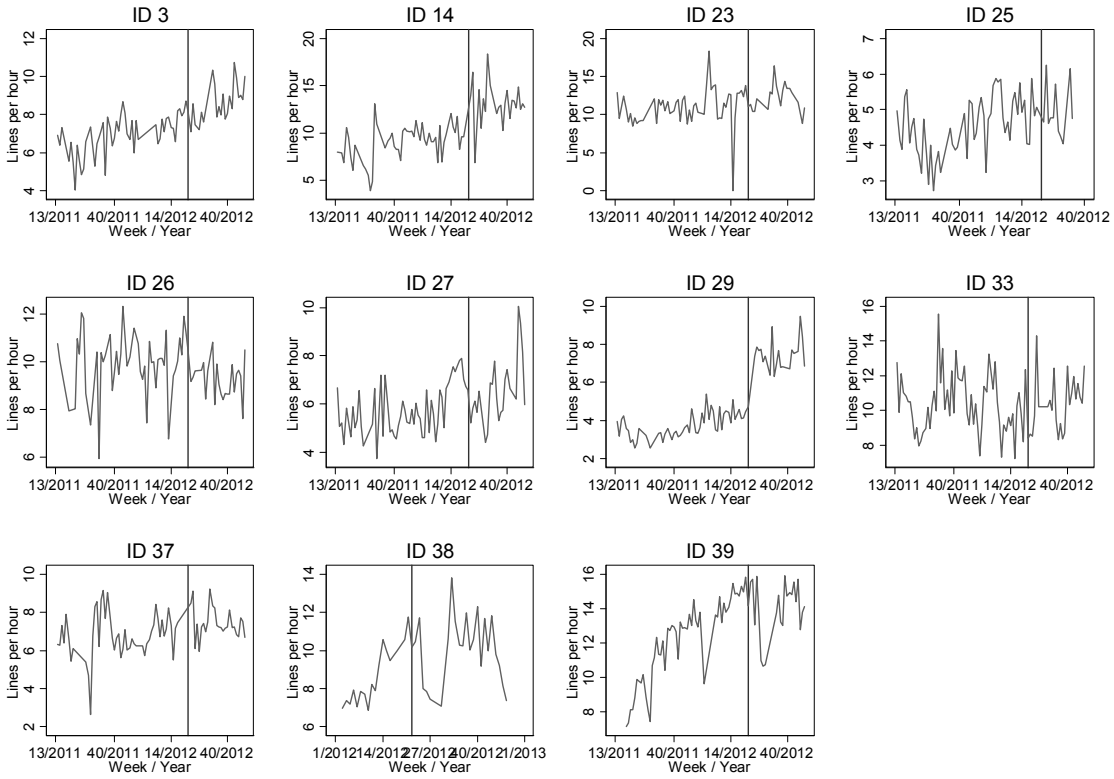
Turning to the individual level analysis, Figure 3 tracks the developments in the individual productivities of the 11 pickers. Again, the horizontal line at week 22/2012 notifies the change in the compensation scheme. Looking at the graphs, it is evident that there are considerable differences both in the responses to the new PRP scheme and in the averages and volatilities of weekly

productivities. Some workers do not seem to react at all, while some, like worker 29, respond quite notably to the new incentive scheme.

As with the average productivity, again, we can notice that the productivities of some workers are growing throughout the observation period. Besides the layout and summer effects, the newer employees might still be learning and gaining experience on the job. In the case firm, most employees are first recruited to the outbound logistics department. Within our sample, we have 2 new recruits, ID 38 and ID 39. For these workers we exclude the first week of observations from the analysis.

Next we move to the econometric analyses of average and individual productivity.

Figure 3 Productivity for individual pickers



Time series analysis

Framework and variables

We are interested in modeling how different workers respond to the updated incentives and how average productivity develops in the outbound logistics department. As the data has a strong time series dimension, it enables us to evaluate time series models separately for each individual, thus providing valuable information on the range of effects an incentive plan can have on productivity.

We use a simple auto-regressive and moving average process (ARMA) as the basic framework. The models we wish to estimate are represented by the following equations.

Unweighted average productivity:

$$ave\_prod_t = \delta + \sum_{j=1}^{p}(\beta_j ave\_prod_{t-j}) + \theta PRP_t + \phi summer_t + \eta L_t + \lambda t + \varepsilon_t + \sum_{k=1}^{q}\gamma_k \varepsilon_{t-k}$$

The individual productivities for each picker *i*:

$$prod_{i,t} = \delta_i + \sum_{j=1}^{p} \left(\beta_{j_i} prod_{i,t-j}\right) + \theta_i PRP_t + \phi_i summer_t + \eta_i L_t + \lambda_i t + \varepsilon_{i,t} + \sum_{k=1}^{q} \gamma_{k_i} \varepsilon_{i,t-k}$$

PRP, summer, and *L* are dummy variables constructed to capture the impact of the new PRP system, the effects of summer holidays and the old warehouse layout respectively while ε is the error term.

We consider both autoregressive (AR) terms and moving average terms as holidays and substituting are likely to have significant but short-lived impacts on productivity. However, as these effects on productivity are probably negative rather than neutral, they are more likely to be captured by the summer dummy rather than the moving average (MA) terms.[5] A time trend is also considered but only with average productivity and the new employees.

The variable *PRP* gets the value one after the implementation of the new incentive plan (from week 23/2012 onwards) and zero otherwise. The coefficient is expected to be positive as having more personalized performance measures is expected to encourage higher personal productivity since work in the outbound logistics is largely independent. Variable *L* gets the value 1 until week 31/2011 and zero thereafter. The coefficient of *L*, the layout, is expected to be negative as the new improved floor plan is anticipated to be improving productivity.

The summer period was chosen to reflect periods where there are several temporary employees working in the outbound logistics; the dummy gets the value 1 for ten weeks from mid-June to late August and zero otherwise. The effect of summer is expected to be more pronounced with the average productivity than with the individual productivities as the set of workers in the average changes from week to week.

The time trend captures other, less well-defined phenomena such as continuous learning, building up experience and (other) changes made in management. Thus it is also problematic to interpret as it can capture also the effects of the PRP regime switch or the new layout. To mitigate this problem, time trend is considered only in the regressions of the two new employees and the department average.

[5] We actually find that the inclusion of the summer dummy makes the AR(1) model more appropriate than ARMA(1,1) which is preferred otherwise, thus clearly indicating that the MA terms and the summer dummy are capturing the same phenomenon.

Model selection methodology

We utilize the Box-Jenkins methodology in selecting the most suitable econometric model. The inspection of autocorrelation and partial autocorrelation functions suggests a low-ordered ARMA or AR process for average productivity and to the majority of the individual productivity models. We also estimate Bayesian information criteria (BIC) for AR processes up to 9 lags and ARMA processes up to 2 AR and 2 MA terms. The information criteria also suggest 1 autoregressive lag for the majority of the pickers.

We use the same model and lag structure for all individuals in order to ensure straightforward comparison of the regression results. There are some individuals whose variation in productivity is not well described by the AR(1) model. This should be taken as another indication of the great heterogeneity amongst the employees. Nevertheless, we find that the results regarding the PRP regime switch or summer are quite robust over the choice of lag length.

Finally, a Box-Ljung Q-test is performed to test whether the residuals are white noise against the hypothesis that they are serially correlated. We find no indication that the residuals are serially correlated.

Results

Average productivity

The estimation results for equation (1) are reported in Table 1. Using the Bayesian information criteria (BIC), AR(1) process without a time trend is found to be the most suitable ARMA(p,q) process in describing average productivity. All estimates are found to be significant and the signs are as expected. It is also seen from the table that various alternative specifications lead to qualitatively similar results concerning the key variables.

The results show that the changing to individual performance measurement increased productivity. Using the long run solution, the new PRP system is found to increase average productivity by 2.02 lines per hour[6]. Summer has a negative effect, decreasing productivity by 0.82 lines an hour. The new

[6] For an AR(p) model the impact on long-run mean is given by $\frac{\delta+\theta\overline{PRP_t}+\phi\overline{summer_t}+\eta\overline{L_t}}{\sum_{j=1}^{p}\beta_j}$ where the overbars denote averages.

warehouse layout improved productivity by 2.26. Thus the long run averages are 13.10 lines per hour with the old system and 15.13 lines with the new PRP system. This represents about 15 % increase.

Table 1 Time series models of average productivity

	With time trend			Without time trend		
Variable	**ARMA(0,1)**	**ARMA(1,0)**	**ARMA(1,1)**	**ARMA(0,1)**	**ARMA(1,0)**	**ARMA(1,1)**
Change in PRP	0.825**	0.816**	0.811**	1.364***	1.313***	1.279***
Summer	-0.470**	-0.459**	-0.453**	-0.563***	-0.535**	-0.508**
Warehouse layout	-1.012***	-1.061***	-1.071***	-1.426***	-1.466***	-1.496***
Time trend	0.015*	0.014*	0.014*			
Constant	7.893***	7.930***	7.935***	8.486***	8.505***	8.515***
ARMA						
AR		0.311**	0.348		0.351***	0.454*
MA	0.295**		-0.041	0.321***		-0.116
Estimated marginal effect of PRP switch	**0.83**	**1.18**	**1.24**	**1.36**	**2.02**	**2.34**
Statistics						
bic	194.629	193.934	198.387	193.173	191.796	196.144

Individual heterogeneity in productivity

Table 2 shows the AR(1) estimation results for equation (2). There is large variation in how individuals respond to the new PRP system. The new compensation system increased the long run mean productivity of four workers, but it decreased productivity for two workers at 10 % significance level. There are five individuals whose productivity was estimated to be unaffected by the new compensation scheme. Thus productivity dispersion increased with the change in the incentive system. This is consistent with the findings in Lazear (2000) and Bandiera et al. (2005).

Table 2 reports both the coefficient and the estimated change in the long run mean productivity caused by the change in the compensation system. Positive impacts ranged from 0.8 extra lines per hour to 4.7 additional lines. The estimated negative effects were smaller, -0.7 and -2.0 lines per hour.

Summer is significant only for two employees, meaning that summer impacts productivity through holidays and substitution rather than actually lowering individual productivities. The renewal of the warehouse layout, on the other hand, has increased individual productivities of five employees. Time trend is significant only for one of the two newer employees but the effect is positive.

The productivity analyses show that changing from collective performance measurement to individual level measurement lead to marked increase in productivity and the dispersion of productivity.

Table 2 AR(1) models for individual productivity

Old Employees	ID 3	ID 14	ID 23	ID 25	ID 26	ID 27	ID29
Productivity(t)							
Change in PRP system	1.41***	3.48***	0.88	0.76***	-0.82*	0.51	3.60***
Summer	-0.48*	-0.58	0.03	-0.84***	-0.03	-0.27	-0.17
Warehouse layout	-1.06***	-2.44***	-1.16	-0.34	-0.11	-0.64	-0.54*
Constant	7.33***	9.75***	11.22***	4.78***	10.03***	5.97***	3.90***
Productivity(t-1)	0.16	0.06	0.24	0.10	-0.11	0.40***	0.24*
Change in mean by PRP system	**1.68**	**3.70**	1.16	**0.84**	**-0.74**	0.85	**4.74**
N	72	76	76	65	73	78	76
BIC	212.59	322.24	352.4	157.35	253.4	248.26	170.91

Old Employees	ID 33	ID 37
Productivity(t)		
Change in PRP system	-0.15	0.52
Summer	0.44	-0.45
Warehouse layout	-0.74	-0.99**
Constant	10.49***	7.07***
Productivity(t-1)	0.27*	0.36***
Change in mean by PRP system	-0.21	0.81
N	83	76
BIC	336.78	225.88

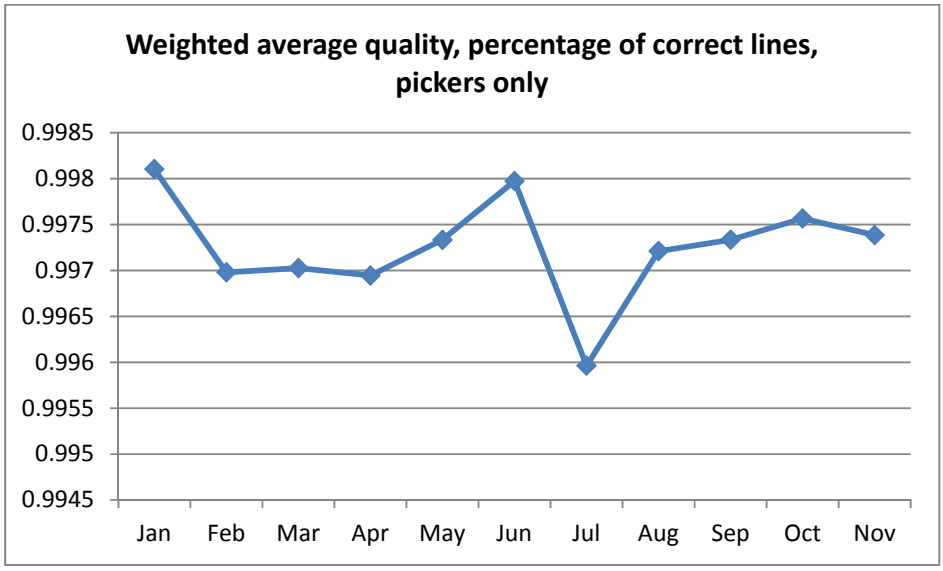
New Employees	ID 38	ID 39
Productivity(t)		
Change in PRP system	0.16	-1.27*
Summer	-0.87	-0.73
Warehouse layout		-2.21***
Time trend	0.05	0.06***
Constant	6.24	10.69***
Productivity(t-1)	0.56***	0.35***
Change in mean by PRP system	0.36	**-1.95**
N	42	74
BIC	166.39	255.94

Quality

The increase in productivity raises the question of what happened to quality. Theoretically the impact on quality is ambiguous. Figure 4 shows the monthly quality series for 2012. Here quality is measured as the percentage of correct lines out of all order lines. It is seen from the figure that no permanent changes in quality can be observed. Due to the short time span, we cannot confidently say whether the increased volatility in quality observed in June and July is due to the introduction of the new system or due to summer.

However, quality is a complex concept, and given that the percentage of correct lines it is integrated into the compensation system, this measure might not be suitable to capture the unwanted effects of performance pay. Nevertheless, these results indicate that aspects of quality can be kept in check by monitoring them.

Figure 4



Allocation of time to different tasks

Incentives affect not only the level of effort but the allocation of effort to different tasks. We have constructed weekly variables from the task-based panel data also for the allocation of time. Similarly to the productivity data, the data is expressed for the average day of a week. The time allocation variables tell the percentage of work time allocated per each of the three task categories (inbound, outbound, other). For example, employee with ID 3 spent 91 % of his time in outbound logistics and 9

% doing other tasks when the old system was in place. The focus is on employees working either in inbound or outbound logistics, 9 and 11 employees respectively, both before and after the system update.

Table 3 reports the time allocation percentages for pickers separately for before and after the PRP update, and whether the difference between the two periods is significant by a t-test. For example, employee with ID 23 spent approximately 93 % of her time in outbound logistics, but after the change, only 81 %. The decrease is statistically significant. The last column of Table 3 reports the results for the unweighted department average.

The results indicate that time allocated towards picking has decreased significantly on average and the same is observed for seven out of the eleven individuals. The average time spent in goods reception has not changed but helping in the inbound logistics department has clearly been centralized to only one person with the introduction of the new compensation system although 2 -5 workers are cross-trained. The "extra" time available is allocated to other tasks that take on average 5 % more of the work time.

Table 3 Average time allocated to different tasks: Outbound logistics

ID		3	14	23	25	26	27	29	33	37	38	39	Average
Before	Outbound	91.00	81.46	92.84	96.89	91.56	84.58	98.30	92.50	99.54	97.58	98.77	92.69
	Inbound	0.00	0.00	0.44	0.05	5.15	10.71	0.18	0.00	0.00	0.00	0.00	1.67
	Other	9.00	18.54	6.73	3.06	3.29	4.71	1.52	7.50	0.46	2.42	1.23	7.62
After	Outbound	92.36	75.62	81.34	94.26	79.27	93.46	94.03	87.02	98.93	61.61	97.31	86.27
	Inbound	0.00	0.00	0.00	0.00	9.98	0.00	0.00	0.00	0.00	0.00	0.00	1.00
	Other	7.64	24.38	18.66	5.74	10.75	6.54	5.97	12.98	1.07	38.39	2.69	12.82
Diff.	Outbound	1.36	-5.83	**-11.49**	**-2.63**	**-12.29**	**8.88**	**-4.27**	**-5.48**	-0.61	**-35.97**	**-1.46**	**-6.42**
	Inbound	0.00	0.00	-0.44	-0.05	4.83	**-10.71**	-0.18	0.00	0.00	0.00	0.00	-0.67
	Other	-1.36	5.83	**11.93**	**2.68**	**7.46**	1.83	**4.45**	**5.48**	0.61	**35.97**	**1.46**	**5.21**

The figures are percentages of total working time. Bolded differences are significant at 10 % level of significance

Table 4 reports the same results but for the inbound logistics department. In the inbound logistics the department averages remain stable while there is some movement within. One person has

significantly changed time allocated to picking by increasing it. Three have decreased time spent in outbound logistics while one has increased. Regarding other tasks, we report that 3 individuals increased and 2 decreased the time allocated for the miscellaneous tasks.

Table 4: Time allocated between task groups, as percentage of work time, for personnel in inbound logistics department

ID		1	4	5	8	9	12	16	21	30	Average
Before	Outbound	0.00	1.73	3.55	8.52	4.67	0.00	0.24	11.32	6.36	3.89
	Inbound	83.06	96.43	92.56	78.50	91.51	97.40	93.64	71.56	92.24	88.61
	Other	16.94	1.84	3.89	12.98	3.81	2.60	6.12	17.12	1.39	9.28
After	Outbound	0.00	2.06	5.58	12.90	9.95	0.00	0.20	3.76	5.32	4.02
	Inbound	90.93	97.07	91.81	86.00	87.14	92.20	76.11	67.88	94.41	88.64
	Other	9.07	0.87	2.61	1.10	2.91	7.80	23.69	28.36	0.27	8.67
Diff.	Outbound	0.00	0.33	2.03	4.39	**5.28**	0.00	-0.05	-7.56	-1.04	0.13
	Inbound	**7.87**	0.64	-0.75	7.50	**-4.37**	**-5.20**	**-17.52**	-3.68	2.16	0.03
	Other	**-7.87**	**-0.97**	-1.28	**-11.88**	-0.90	**5.20**	**17.57**	11.24	-1.12	-0.61

The changes in the outbound logistics are much more pronounced and trended compared to those of the inbound logistics department. Hence, some of the changes observed in inbound logistics might be consequence of internal recruitment rather than responds to the changed incentives in the inbound logistics.

Discussion

Time stamping practices. As some employees work in both in- and outbound logistics they are subject to both team and individual effort based performance pay. PRP is paid through both systems in accordance to the number of hours that they spend in each department. When doing 'other tasks,' performance pay is based on the average productivity of the entire warehouse or of the department. The way productivity is calculated both within the PRP system and in this analysis is sensitive to time stamping practices. Certain thresholds must be satisfied in order to get performance pay and hence an employee might get better compensation by decreasing the time spent. Hence, employees have increased incentives to report their time usage more accurately.

We observe that in the outbound logistics department, the time spent on picking tasks has decreased while productivity has correspondingly improved. Hence some of the observed productivity improvement might be a consequence of changed time stamping practices rather than actually working more effectively. However, when investigating the developments in *executed lines per day* we find very similar trends compared to those of the *lines per hour average productivity*, indicating that time stamping alone does not necessarily explain the majority of the observed productivity improvements.

Specialization. Aligning the interests of the employee and the employer equips the employee to make better decisions also for the firm. This might happen for example in the form of individuals specializing in tasks that they have an advantage with. Given that the PRP system takes into account the difficulty differences between the different picking tasks, some workers might specialize over time in tasks that are harder in the sense that they have lower *lines per hour* figures. Such specialization might lower individual productivity as measured in this analysis (although the effort is fully compensated by the performance pay system), yet simultaneously improving the departmental average productivity.

Evidence of this kind of specialization can be found in that the departmental average has improved more than the average of the individual productivities over the course of time. This is however a natural phenomenon as the firm recruits workers internally and picking is usually the starting point. Performance pay can intensify this by promoting self-selection or specialization. That helping out in the inbound logistics has been concentrated to only on picker after the implementation of the new PRP plan is an example of how performance pay can catalyze this process.

Conclusion

We study a warehouse that changes its incentive plan from a group plan to individual plan. We use time series methods at the warehouse level and the individual level to study the quantitative impact that the change had on productivity. We find that this change increased productivity by around 15 %. We also find substantial variation between individuals: the change increased productivity for some, had no impact for others, and for two employees it had negative effects. These results are line with

previous literature showing the stronger incentives increase both the level and dispersion of productivity. Quality did not change although productivity increased.

We also find that the change affected allocation of time to different tasks. Under the collective plan employees in outbound logistics tended to sometimes help in the inbound logistics. After the change, these responsibilities shifted to a single individual. We also find that the employees spend less time in their primary task, e.g. picking of goods, and more time in other tasks. This may reflect change in the time-stamping practices, where the pickers are more careful to record their time accurately under the new plan.

The change in the incentive plan increased productivity but changing the plan was costly. First, the whole process took about 18 months during which many meetings were held. Second, developing the software routines that combine reports from different programs took time and effort of a specialist. Third, running the system takes about half a day of management time and half a day of a controller's time each month to calculate and pay the individual bonuses. These costs made the management to think about other, more profitable, ways to motivate their employees. In the economics literature status incentives have recently attracted attention (see e.g. Besley & Ghatak, 2008).

References

Baker, G. (2002). Distortion and Risk in Optimal Incentive Contracts. *Journal of Human Resources, 37*(4), 728-751.

Bandiera, O., Barankay, I., & Rasul, I. (2005). Social Preferences and the Response to Incentives: Evidence from Personnel Data. *Quarterly Journal of Economics, 120*(3), 917-962.

Bandiera, O., Barankay, I., & Rasul, I. (2007). Incentives for Managers and Inequality among Workers: Evidence from a Firm-Level Experiment. *Quarterly Journal of Economics, 122*(2), 729-773.

Banker, R. D., Lee, S.-Y., & Potter, G. (1996). A Field Study of the Impact of a Performance-Based Incentive Plan. *Journal of Accounting and Economics, 21*(2), 195-226.

Besley, T., & Ghatak, M. (2008). Status Incentives. *American Economic Review, 98*(2), 206-211. doi: http://www.aeaweb.org/aer/

Bloom, N., & Van Reenen, J. (2007). Measuring and Explaining Management Practices across Firms and Countries. *Quarterly Journal of Economics, 122*(4), 1351-1408. doi: http://www.mitpressjournals.org/loi/qjec

Che, Y.-K., & Yoo, S.-W. (2001). Optimal Incentives for Teams. *American Economic Review, 91*(3), 525-541.

Drago, R., & Garvey, G. T. (1998). Incentives for Helping on the Job: Theory and Evidence. *Journal of Labor Economics, 16*(1), 1-25. doi: http://www.journals.uchicago.edu/JOLE/home.html

Dumont, E., Fortin, B., Jacquemet, N., & Shearer, B. (2008). Physicians' Multitasking and Incentives: Empirical Evidence from a Natural Experiment. *Journal of health economics, 27*(6), 1436-1450. doi: http://www.elsevier.com/wps/find/journaldescription.cws_home/505560/description#description

Freeman, R. B., & Kleiner, M. M. (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), 307-330.

Gibbs, M. J., Merchant, K. A., Van Der Stede, W. A., & Vargus, M. E. (2009). Performance Measure Properties and Incentive System Design. *Industrial Relations, 48*(2), 237-264.

Holmstrom, B., & Milgrom, P. (1987). Aggregation and Linearity in the Provision of Intertemporal Incentives. *Econometrica, 55*(2), 303-328. doi: http://www.econometricsociety.org

Holmstrom, B., & Milgrom, P. (1991). Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design. *Journal of Law, Economics, and Organization, 7*, 24-52.

Itoh, H. (1991). Incentives to Help in Multi-agent Situations. *Econometrica, 59*(3), 611-636. doi: http://www.econometricsociety.org

Ittner, C. D., & Larcker, D. F. (2002). Determinants of Performance Measure Choices in Worker Incentive Plans. *Journal of Labor Economics, 20*(2), S58-S90.

Jones, D. C., & Kato, T. (1995). The Productivity Effects of Employee Stock-Ownership Plans and Bonuses: Evidence from Japanese Panel Data. *American Economic Review, 85*(3), 391-414. doi: http://www.aeaweb.org/aer/

Kandel, E., & Lazear, E. P. (1992). Peer Pressure and Partnerships. *Journal of Political Economy, 100*(4), 801-817.

Kauhanen, A. (2011). The Perils of Altering Incentive Plans: A Case Study. *Managerial and Decision Economics, 32*(6), 371-384. doi: http://www3.interscience.wiley.com/cgi-bin/jhome/7976

Kauhanen, A., & Napari, S. (2012). Performance measurement and incentive plans. *Industrial relations: a journal of economy and society, 51*(3), 645-669.

Knez, M., & Simester, D. (2001). Firm-Wide Incentives and Mutual Monitoring at Continental Airlines. *Journal of Labor Economics, 19*(4), 743-772.

Lazear, E. P. (2000). Performance Pay and Productivity. *American Economic Review, 90*(5), 1346-1361.

Levenson, A., Zoghi, C., Gibbs, M., & Benson, G. S. (2011). Optimizing Incentive Plan Design: A Case Study: IZA Discussion paper series // Forschungsinstitut zur Zukunft der Arbeit, No. 5985.

Owan, H., & Tsuru, T. (2011). *Integrating High-Powered Performance Pay into a Seniority Wage System*. Institute of Social Science, University of Tokyo.

Paarsch, H. J., & Shearer, B. (2000). Piece Rates, Fixed Wages, and Incentive Effects: Statistical Evidence from Payroll Records. *International Economic Review, 41*(1), 59-92. doi: http://www.blackwell-synergy.com/loi/iere/

Paarsch, H. J., & Shearer, B. S. (1999). The Response of Worker Effort to Piece Rates: Evidence from the British Columbia Tree-Planting Industry. *Journal of human resources, 34*(4), 643-667. doi: http://jhr.uwpress.org/archive/

Shearer, B. (2004). Piece Rates, Fixed Wages and Incentives: Evidence from a Field Experiment. *Review of Economic Studies, 71*(2), 513-534.