Goal Setting, Feedback and Bonus Pay in a Field Experiment.

Gerrit Erwin Zielhuis, 378221

Erasmus University Rotterdam

M.Sc. Economics of Management and Organisations.

Supervisor: Dr. J. Delfgaauw.

July 2017

Abstract

A Field Experiment is run at a large Dutch Company, specialized in stock taking activities. A total of 82 projects of the Company were used to test if different kinds of incentives would increase a workers' productivity. The projects were split up in 3 groups. Two groups were treatment groups, and the third group served as a control group. The workers in the treatment groups got performance related goals communicated just before projects started. In one of the treatment group, the workers were also told that they would receive feedback on their performance relative to the goals, and in the other treatment group workers were eligible for a bonus pay if they reached the goals. Workers in the feedback treatment increased their productivity with at least 7.3 percent and workers in the bonus pay treatment increased their productivity with at least 10.8 percent. It is shown that these results are mainly caused by the fact that workers put in more effort in their work activities, and therefore these results are not driven by multitasking or gaming.

I gratefully thank dr. J.Delfgaauw for the great support, in addition I thank employees from the Stock Take Company for their collaboration.

Table of Contents

INTRODUCTION 3
Section 2 5
2.1 Solutions to the agency problem 5
2.2 Intrinsic and extrinsic motivation 7
2.3 Feedback 8
2.4 Bonus Pay 10
Section 3 11
3.1 Experimental Context 11
3.1.1 Description of the Stock Take Company 12
3.2 Experimental design 13
3.2.1 Description of the Retail Firm used for the experiment. 14
3.2.2 Feedback and Bonus Pay implementation 15
3.2.3 Planning and Timing 18
Section 4 20
4.1 Data description 20
4.2 Empirical Strategy 25
Section 5 27
5.1 Results 27
Section 6 34
6.1 Robustness checks 34
Section 7 38
7.1 discussion 38
7.2 Further Research 40
Section 8 41
8.1 Appendix: Literature 41
8.2 Appendix: Additional Tables 44
8.2.1 One way ANOVA project level 44
8.2.2 one way ANOVA worker demographics 47
8.2.3 Chi^2 test Worker Demographics 48
8.2.4 Scans per Hour as Dep Variable. 49
8.3: Appendix: experimental design 50
8.3.1: Extra instruction project leaders bonus pay treatment 50
8.3.2 Extra instructions project leader feedback treatment 51
8.3.3 Official randomization of treatment groups 53
8.3.4. Actual randomization of treatment groups 54

INTRODUCTION

Formerly, fundamental economic models mostly considered a large perfect market. Assuming a large perfect market made it possible for firms to perfectly calculate optimal equilibria and to predict actions of other players in the market. However, Marschak (1955) and Arrow (1968) did already early recognize that these models did not hold in the real market. They started to look at members within a firm, and also recognized that these members did differ in risk attitudes and preferences. In addition, the models based upon a large perfect market did assume that all information was observable, and equal for all players, but in the real world, players have private information that other players in the market couldn't observe.

It is credible to state, that when members within a firm have (1) different preferences and risk attitudes, and (2) this preferences and attitudes are not observable for all players in the market, it is not possible anymore to perfectly predict the actions of other players, and the first best equilibria fall apart (Laffont & Martimort, 2001).

One of the most important relationships in the economics is the relationship of the manager of a firm and his workers. In the perfect market, the manager is able to create a perfect contract for the worker, where the worker takes the exact actions as the manager desires. If the manager and the worker do differ in preferences and risk attitudes, but all information is still given, the manager knows the preferences of the worker, it is still possible for the manager to create the perfect contract, because the manager can still predict the actions of the worker. But when also information is hidden, it not possible anymore for the manager to set up the perfect contract, because the manager cannot optimally predict the actions of his worker anymore. In this scenario, the manager has to choose another strategy to make the worker take the actions as the manager desires; giving the right incentives to the worker.

This problem stresses the importance of incentives in today's world. Since people do differ so much in behaviour, preferences and attitudes, it is essential to give the right incentives for everything. People can be given incentives to work harder, incentives to produce good quality products, incentives to invest, incentives to save, incentives to study, even incentives to care for family. These incentives can come from a person's own intrinsic motivation, or from rewards or punishments. How to design institutions with good incentive systems is therefore a central questions in the economic environment (Laffont & Martimort, 2001). The theory of incentives, which is currently one of the most important and interesting theory's in the economics, will therefore play a central role in this thesis.

With the theory of incentives comes the principal agent theory, which is already briefly discussed above. If parties in a firm have different goals and the roles of these parties are divided in different divisions of labour, the agency problem could arise. The problem exists when a principal delegates work to the agent. The agent then has to perform the work. But when the goals or objectives of the agent differ with the objectives of the principal, the agent is not incentivized to behave and exactly take the actions the principal desires. This could be solved if the principal could fully observe the agents actions, but since this is not a realistic assumption, the agency problem exists (Jensen & Meckling, 1976; Ross, 1973). Due to the different goals, hidden information and disability to fully observe actions, the potential

surplus in these relationships is not fully utilized. Jensen & Meckling (1976) take this theory and try to describe this relationship using the metaphor of a contract.

By doing this, the Agency problem gets tangible for real life work relationships. By using agency theory, contracts can be formed where desired behaviour of the agent can be forced by giving the agent the right incentives. A widely used incentive is money, but intrinsic incentives can be at least as important. Here we are back at the importance of the right incentives. Agency theory offers managers and owners of a firm good insights into information system, outcome uncertainty, risk sharing and the important incentive theory, which also makes the theory a solid fundament for empirical research (Eisenhardt, 1989).

This thesis is therefore building on these theories to empirically test several solutions to the agency theory in trying to boost performance of workers by giving them incentives. In performing a field experiment in a real work setting two different kind of incentives are implemented, and the results of both incentive systems are interpreted and compared.

The context of the experiment is in the stock take environment, in which a large stock take company, counts the stock in stores of large external store chains. The experiment is done with 82 projects and almost 500 subjects. The projects were split up in 3 groups: Control Group, Feedback Treatment and Bonus Pay Treatment. In the feedback treatment, subjects got performance related target rates communicated at the start of the project. These targets are set by the stock take company, and are the expected levels of performance that the average worker should reach. In addition, workers in the feedback treatment, before they started working, were told that they would get feedback during the project. The feedback consisted of their personal performance during that specific project.

In the Bonus Pay Treatment, subjects also got performance related target rates, but now, subjects were told that they would get a bonus if they reached the targets. There were two targets in total with a Bonus Pay of 6€ on top of the normal pay out if target A would get reached and a Bonus Pay of €8,50 on top of the normal pay out if target B would get reached, where target B> target A . The targets consist of a quality constraint, to deal with possible multi task problems. The Feedback Treatment and Bonus Pay Treatment are almost identical to each other, except for the fact that in the Bonus Pay Treatment an extra monetary reward is implemented. A more accurate description of both treatments will be discussed, and the differences and similarities will be discussed in section 3.

Using a field experiment to test if these incentives have an effect on productivity is believed to be a reliable method to test the theory. The subjects were told that they were in a pilot performed by the company, so they couldn't know they were part of an experiment, which increases the robustness of the results. This thesis stands out because, surprisingly, field experiments with real companies performed to test monetary incentives are rather scarce and empirically testing the effect of more performance related feedback is also rather unique. Also the comparison between testing two different incentives in the same work setting attributes to the literature on incentives and agency theory, and gives supporting insights in the discussion on the differences between intrinsic and extrinsic incentives.

The field experiment resulted in that, when giving workers performance related targets in combination with feedback on these targets during the work activities, the workers would increase their productivity with circa 9.1 percent compared to the control group. Another significant positive reaction is found by workers who were eligible for the bonus pay, these workers increased their productivity with at least 10.8 percent compared to the control

group, all other factors kept equal. Addition tests proved that these results were only due to the fact that workers increased their effort input, and not because of shifting their effort from quality performance to quantity performance or by gaming the system and design.

The thesis proceeds as follows, in the next section the literature on solutions to the agency problem is discussed, and the two forms of potential solutions that are implemented in the experiment are reviewed to help predict the outcomes, also the difference between intrinsic and extrinsic motivation is discussed. In section 3 the experimental context and the experimental design are described and section 4 contains the empirical strategy to produce the results in section 5. Section 6 tests the robustness of the results and Section 7 discusses and concludes.

Section 2

2.1 Solutions to the agency problem

A multitude of literature exists how to solve the principal agency problem. When we use the metaphor of a contract, it is the principal's task to design the right contract for the agent. The goal of the contract is to incentivize the agent to take the exact actions the principal wants.

One of the most used solutions to this problem is pay for performance. This means that the agent get payed for his performance, rather than getting a fixed wage. One of the greatest examples of this is described in a paper by Lazear (2000). A large auto glass company switched from paying a fixed wage to paying a piece rate and results were striking. The average productivity increased between 20% and 36%. Performance pay can be a strong instrument to increase the productivity of the agent. However performance pay has also disadvantages. As mentioned earlier, not all actions of the agent are observable, this is also called the monitoring problem. Monitoring all actions of the agent is first of all expensive, but it is also impossible in real life and not always ethical right. Due to this monitoring problem it is hard to perfectly contract on the firm's objective. The design of a performance pay contract containing of an undistorted performance measure is rather impossible (Oosterbeek, Sloof & Sonnemans, 2006). Which means that if the performance measurement is inaccurate and not identical to the principal's objective, the agent only has an incentive to put effort in the performance that is measured, rather than the intended objective of the principal; the so called multitask problem (Holmstrom & Milgrom, 1991; Baker,1992). For evidence of the multitask problem see Brickley & Zimmerman (2001), who show that the shift in rewarding teaching instead of research in a top tier business school results in a significant and immediate boost in teaching rating and a decrease in research ratings. Another experiment that tries to analyse behaviour in multitask settings, finds evidence for all the basic theories in the multitask and agency literature. It shows that the agent's effort allocation will be distorted when not all actions of the agent can be monitored. However this article also stresses that human behaviour doesn't follow the multitask theory in the extremes. This is because of factors as fairness and reciprocity (Fehr and Schmidt, 2004). This discussion on the multitask problem will be useful later on in this thesis, when we have to deal with this problem by designing the incentives.

Instead of rewarding a single agent for his performance, a team reward could be used as a solution to the agency problem. So, instead of rewarding personal performance, the reward is given if a team of several members reaches a performance target. This can have positive peer effects in the sense that team members keep an eye on each other and stimulate each other to put in more effort. This can solve the monitor problem in a way that the team members will monitor each other and in addition, the team members are able to learn from each other (Hamilton & Nickerson, 2003; Lazear & Gibbs, 2010). But if a member in a team believes that other team members will insert sufficient effort to reach the desired performance target, they might use others effort to put in less effort and the agent is free riding on others performance (Albanese & Van Vleet, 1985).

Another known method to get rid of the agency problem, and especially the monitoring problem causing it, is by using promotions as a reward. This is called a tournament model, where the one agent with the best overall performance of several agents in a firm gets the price of the tournament; the promotion. This often comes with a higher salary and a better function. Since not all performance of each single agent needs to be measured, is it easier and cheaper to monitor and agents will also put in effort in performance in line with the firm's objective (DeVaro, 2006; Prendergast, 1993). However, in this setting there is the threat that the competition among the agents gets too rough, and that some agents try to sabotage one another. Instead of focussing on own performance the agent focusses on sabotaging the others performance; yet another form of multitasking (Milgrom & Roberts, 1988).

Only a few solutions in trying to deal with the Agency problem have been discussed up to now. Obviously there are lots of other solutions to this problem, among others: Giving the agent asset ownership (Gibbons, 2005) or make the agents get a broader and more flexible responsibility than the original job task demands (Berlin, 2014).

Most of the suggestions discussed above are examples of extrinsic motivation. All of the incentives consist of a part in which a monetary reward is given to the agent, trying to incentivize him to put more effort in his job. Another way to improve the agents productivity is to incentivize his intrinsic motivation. We already see ways of intrinsic motivation at the Gibbons (2005) and (Berlin, 2014) examples. When the agent sees that he is responsible for his tasks, he will also get utility from just executing his job, because he might feel more satisfied when the job is done well. This has nothing to do with an extra monetary pay off that he gets when the job is performed well. The same way of reasoning can also be used with the promotion incentive; when the agent gets promoted, he might feel better because he gets more recognition for his job. And when agents are divided in teams, the agent can enjoy more social contact, yet another way of intrinsic motivation.

We see that every method has its own advantages and disadvantages, and that every method is not always just an extrinsic motivation. It is therefore important that enough research is conducted to help understand which mechanisms drive the agents actions. This thesis hopes to attribute and broaden the use of incentives to increase the agent's productivity and to help improve the principal agent's relationship.

2.2 Intrinsic and extrinsic motivation

This subsection discusses the differences between extrinsic motivation and intrinsic motivation and gives a brief analyse on positive and negative sides. This will help to interpret the results of the experiment , in which both extrinsic and intrinsic motivation strategies are implemented, and to get the right understanding of the mechanisms that drive the results.

Before starting to discuss intrinsic and extrinsic motivation, it is also important to understand what motivation itself means. If a person is motivated, this person is moved to do something. This person has intention and/or inspiration to perform a task or a job. Some persons are more motivated than others but beside of the amount of motivation, persons can also be motivated in different ways (Ryan & Deci, 2000). A useful theory to help understand the different kind of motivations is given by Deci & Ryan (1985). The so called Self-Determination Theory distinguishes between different kinds of motivation based on the different reasons or goals that give rise to an action. Here the distinction is made between intrinsic motivation, which refers to doing something because it is inherently interesting or enjoyable, and extrinsic motivation, which refers to doing something because it leads to a separable outcome. For example a student could be motivated to learn a new set of skills because he or she understands their potential value (intrinsic motivation) or because learning will yield a good grade and the benefits that come with good grades (extrinsic motivation).

In the principal agent setting, the principal can choose to motivate the agent with extrinsic incentives; e.g. with monetary rewards , or to motivate the agent with intrinsic incentives. With an intrinsic incentive, the agent will get more joy from his job and will gain utility from performing the job, which has nothing to do with the reward for the job. The principal may make the agent clear that his work is meaningful for the organization. Another way is to let the agent choose how to accomplish his tasks, which will give the agent more responsibility and joy. Confirmation and feedback that the agent is doing his work well and encouraging the agent by noticing that his effort is really accomplishing something are all forms of intrinsically motivating the agent (Thomas, 2009).

For example in an experiment with 300 employees working on a 3 hour task, a random sample of workers unexpectedly got recognition for their work after two hours. This increased the subsequent performance substantially (Bradler, Dur, Neckermann & Non, 2016).

If the comparison is made between intrinsic motivation and extrinsic motivation, one might notice that intrinsic motivation doesn't involve spending extra money, at least not as much as with extrinsic motivation. However it is relatively harder to implement intrinsic motivation and it varies in its scarcity. A principal cannot tell the agent every hour that he is doing a good job, because the agent will start to believe that the principal is only saying this because he wants the agent to work harder (Bradler, Dur, Neckermann & Non, 2016). There is a difference between saying and meaning, and when forms of intrinsic motivation are implemented too often it might lose its credibility (Stalnaker, 2006).

On the other hand, extrinsic incentives also have negative sides. One of the biggest problems with extrinsic incentives is that it could have the opposite effect. It could decrease the agents performance instead of increasing it. This has everything to do with the intrinsic motivation of the agent. Extrinsic incentives might crowd out the intrinsic motivation (Frey & Jegen, 2001). When the principal decides to implement an extrinsic incentive, the agent might consider the implementation a bad signal. The bad signal could mean that the principal does not trust the agent's intrinsic motivation and competence. This could mean that the implementation is bad news for the agent and can therefore lower his intrinsic motivation to do his work (Gneezy, Meier and Rey-Biel, 2011). As discussed already, multitasking is also a threat by implementing extrinsic incentives. If the incentives only capture a part of the whole objective, effort is shifted to the rewarded tasks. One last mechanism causing extrinsic motivation to have a negative effect is work under pressure. We also have to deal with this threat in our experiment and it is therefore useful to discuss this specifically, hence there is a lot of more to discuss, but that is beyond of the scope of this thesis. Extrinsic incentives can be set too high, and the agent can get stressed or lose focus or lose faith in reaching the target due to the high incentives. In an experiment in which subjects were split in two groups; with one group having double the payoff of the other group, the group with the lower payoff had a better performance, while both groups were performing the same tasks (Ariely, 2009). Similar results are found in a study with accountants, who performed better in the absence of monetary incentives and feedback (Ashton, 1990).

2.3 Feedback

Despite the small discussions above, which are lacking lots of other theories and mechanisms, we are now more able to define the two incentives that are implemented in the field experiment conducted to write this thesis. As mentioned before, one treatment will consist of giving more feedback to the workers, and the other treatment will consist of a bonus pay paid to the workers when they reach a certain target, to be discussed in the next subsection.

In the feedback treatment, workers will be informed before they start working on a project about the performance levels the organization expects the workers to work at. Both quantitative and qualitative measures are communicated, and the workers are informed that they will receive feedback during the project that consists of an individual performance measure. Workers know therefore that they will get confirmation about how they perform and are able to compare the individual performance with the goals set at the beginning of the project. More details about the feedback treatment can be found in the experimental design and the appendix.

Most literature on feedback consists of either feedback in the form of recognition or feedback on performance relative to other workers. This is one of the features of this thesis that stands out from other literature. Due to the performance goals communicated before the project, workers know they will get recognition if they perform well, and will therefore be incentivized to work harder. So the recognition is credible when it is given, and it can already have a positive effect before it is given! Furthermore, the feedback they get is relative to the organization's goals, instead of relative to workers, which takes away threats of bad competition among workers; e.g. causing sabotage etc.

Recognition having positive effects on the workers productivity is widely proven in economic literature. Research shows that what really motivates workers to perform better, even more than monetary incentives, is thoughtful and personal recognition that credibly shows appreciation when a task is performed well (Nelson, 2005). Empirical evidence of the positive effects of recognition is amongst others given by Stajkovic and Luthans (2003) and Gino (2010).

Other literature this thesis contributes to is literature about relative performance feedback. Barankay (2011) finds that workers who are given feedback regarding their performance relative to others, are on average less productive on the job compared to a control group, which did not receive feedback. This argues for the set up in our experiment to not use feedback relative to other workers, but feedback relative to the organization's goals. However other research proves the opposite, for example a paper showing that the revealing of relative rankings in performance increases productivity significant, also in the long term (Blanis I Vidal & Nossol, 2009). In another experiment people turn out to work harder when they find out that they might hear their relative ranking (Kuhnen & Tymula, 2012). This highlights the positive effects of announcing that feedback will be given. Also people who performed worse than they had expected increased effort, while people who performed better than they had expected decreased effort. The paper suggests that relative feedback helps create increasing effect in productivity mainly because they fight to be at the top of the rank. Despite the fact that the feedback given in our field experiment is not relative to other workers, the mechanisms that drive the workers actions can still be similar to what is described in the above literature.

Strongly related literature is a study that compares the motivational impact of goal setting and performance feedback. This study finds that feedback on four different measures on performance is superior to non-feedback, and that goal setting is superior to non-goal setting (Ivancevich & McMahon, 1982). In an experiment where electricity consumers set goals on electricity usage and got feedback on their goals, their performance increased compared to a control group. The positive result was due to the joint effect of goal setting and feedback (Becker, 1978). A paper about a field experiment that was conducted to measure gender differences has also large similarities with this thesis. The field experiment contained both a monetary incentive and feedback incentive. 128 stores of a large discount retail chain in the Netherlands took part in the field experiment. The stores were treated with either a control treatment, a tournament treatment or a tournament treatment with monetary rewards. In a tournament treatment, a store competed over a period of 6 weeks in a pool with 5 stores. They received weekly feedback on their relative performance compared to the other stores in their pool. The difference between the tournament treatments is that one treatment just gave the feedback and the other treatment gave the feedback with a monetary reward. Result turned out to be significant for both treatments, but there was no difference between the "feedback" treatment or the reward treatment (Delfgaauw et al, 2013). Different from this thesis is that feedback and incentives are given on personal level and not on store or team level. Also feedback and incentives that were given in this thesis field experiment are not relative to others.

We should take in mind that the feedback given in the experiment is a real measurement of the performance and relative to the organizations goals. It will only serve as recognition of good work if the workers actually do perform well. This can cause the threat of too much pressure for the worker if he does not perform according to the organizational objective as discussed already. This also makes the incentive to work harder not purely intrinsic. It is partly extrinsic in the sense that the worker feels that he has to perform on a certain level in order to work for the organization, which is extrinsic motivation as we have seen in the discussion on intrinsic and extrinsic motivation. The intrinsic motivation is caused by the fact that workers experience joy in reaching goals, and the recognition they will get for this.

It deals with the multitask problem because both quantitative and qualitive goals are set, and feedback is given on both performance measures.

2.4 Bonus Pay

Interestingly with this thesis is the comparison between a non-monetary incentive and a monetary incentive. In the bonus pay treatment, workers will also before the start of the project be informed about both the quantitative and qualitative objective of the firm. If the workers reach this target they will get a certain bonus above a fixed wage. For more details see the experimental design and the appendix.

A lot of literature of monetary incentives has already been discussed. Recall the research of Lazear (2000) and the discussion in subsection 2.1.

The Lazear paper is outstanding, but as mentioned before, the supply of empirical evidence acquired by real life field experiments on the effects of monetary incentives is rather scarce. Especially field experiments with large timeframes are not common. This thesis hopes to attribute to this supply, despite the small timeframe of its field experiment. However, there are still good examples of researches that saw the urgency to provide evidence on these important topics. The field experiment on gender differences, discussed in subsection 2.3, also found that on average, the tournaments increased percentage sales by about 5 percentage points, whilst there was no difference between the treatments with and without a monetary incentive (Delfgaauw et al, 2013). A field experiment taking place in a fruit picking context, tests the effect of introducing a tournament with monetary rewards for teams resulting in a 24% increase in productivity (Bandiera et al, 2013).

Additional literature to Lazear and this thesis, on the effects of monetary incentives on the personal level, instead of team incentives, is provided by Paarsch and Shearer. They first developed the optimal contract for a British Colombia tree planting firm following the principal agents theory, and estimated that profits would increase by 17 % if the firm would implement this contract based on piece rates (Paarsch & Shearer, 1999). In following research in the same context, they conclude that structural estimation accounted for the firm's optimal choice of a compensation system suggests that incentives caused a 22.6 percent increase in productivity. However, a part of the increase is due to a decrease in quality output, hence the multi task problem (Paarsch & Shearer, 2000). In a field experiment conducted again within the tree-planting firm, they find that the average productivity gain is 20 % when the worker

switched from fixed wage to a piece rate. Using econometrics estimation methods to control for the different planting conditions during the experiment, they suggest that the increase in productivity, controlled for experimental influences, would have been at least 21,7 % (Shearer, 2004). This thesis is different in that the same workers have not been observed in both control and treatment.

In addition, literature in psychology and behavioural economics says that paying workers for good work may undermine intrinsic motivation, empirical evidence of this mechanism is provided in a paper that performance pay has positive effects on productivity but can have negative psychological effects, depending on heterogeneity, suggesting that workers respond differently to incentives (Huffman & Bognanno, 2015). This is stressing the interesting feature of this paper to compare different motivation strategies in the same setting.

Threats with the bonus pay is that the target is set too high, and workers will choke under pressure, the feedback treatment deals with the same threat. Therefore, historical data have been used to estimate a reasonable target for the workers. The multitask problem is tackled by setting a qualitative constraint to the performance target.

Section 3

3.1 Experimental Context

The methodology that is used to examine the effects of the feedback incentive and the bonus pay incentive is by executing a field experiment at a private firm that provides stock-taking services. This Stock Take Company is specialized in counting stocks of large external retail firms. Last year, the Stock Take Company executed projects for at least 20 different retail firms located mainly in the Netherlands, but also in foreign countries as Belgium, Germany and even Sweden. The average external Retail firm has circa 100 stores. This varies between Retail Firms with 10 stores and large Retail Firms with 190 stores. On average 25 workers are needed to perform a stock take at one store of a retail store. However, some stores only need 5 workers to get a project done, but there are also stores that need almost 100 workers to get all the stock to counted. The average worker, works on average 3.7 hours on a project. The field experiment executed for this thesis has been performed at one of the large customers of the Stock Take Company.

Using a field experiment as the methodology, is believed to be a good way to measure the change in behaviour of the workers when they are confronted with incentives. This specific experiment will test whether giving more performance related feedback to workers during a project or implementing a bonus pay during a project, will indeed be good incentives to increase the productivity of the worker.

The field experiment can be described as a *natural field experiment* . One of the greatest features of a *natural field experiment* is that the subjects don't know that they are taking part in an experiment. This helps to offer unbiased insights in the reactions of the subjects to the incentives. It is therefore believed to be an ideal experiment, it is possible to observe the subjects in a controlled setting, but the subjects don't perceive the controls as being

unnatural (Harrison and List, 2004). To provide extra strength to the methodology of using a field experiment, a control group is used, which observes no unnatural changes.

3.1.1 Description of the Stock Take Company

A project is done when the Stock Take Company sends a team of workers to one store of a Retail Firm and counts all the articles in the store. All stores of a Retail Store are mostly counted within a short time frame. The core task is performed by workers who scan all the articles in the store into a database using a scanning device. Each product in a store needs to be scanned, including the products in the warehouse located at the same store. As a rule, a store is divided in zones, and the scanning process is done zone by zone. This helps to give a clear overview of which products are scanned and which are not. Scanning is done by scanning the barcode on the product with the scanning device. By doing this, every product is scanned into a large database which can be compared with the life stock in the store. Every scan made can perfectly be traced, which gives large opportunities for data analyses. It is therefore surprising to notice that performance related feedback is seldom given to workers.

One can imagine that there are a lot of products in a store, and that mistakes are easily made. A worker can for example count ten of the same products, than scan the barcode of the products, and fill in ten pieces into the scanning device, whilst there were actually eleven pieces. Instead of counting errors, the worker can count the products right, but one product might have a slightly different colour, and a count error is made again. To tackle these problems, the Stock Take Company also provides quality checks. Therefore, beside workers who are scanning and counting all the products in the store, there are also workers walking around with tablets. These workers can immediately see when zones with products are counted, and can then do a recount of the zone to double check. If an error is found, it can be corrected by the checker in the tablet. It is immediately visible in the database that a wrong scan was made, and how this scan was corrected. The errors made are therefore also easy to assign to the worker who made the scan error. However, it doesn't often happen that a worker immediately gets notice when an error has been made. Mostly positive feedback is given to the worker by the project leader if it is possible to prevent that the error is made again. Only when the worker makes the errors frequently, he will get notice of this from the project leader.

Depending on the agreements with the concerning external retail firm, quality checks are performed over a minimal percentage of all the products counted. The products to be checked are partly random checked and partly selected; zones with high value products are more likely to be checked, or zones where mistakes are relatively easily made are more likely to be checked. The quality check is performed by the Stock Take Company, but employees from the Retail Firm's store mostly help executing the quality checks. This is important in order to credibly execute the quality checks and make them reliable.

In addition, every project has a project leader, who is in charge of the scanners and checkers. It is the project leader's responsibility that all the workers do their work, and to communicate with the store's manager. There is also a preparation team, who before a project starts, splits up the whole store in zones, and there is an IT manager who sets up and manages the software. However, in this thesis the main focus is on the performance of the workers who perform the scanning activities.

As mentioned, the amount of persons needed for a project to be performed vary between the Stock Take Company's customers. For some projects almost 100 workers, who mainly perform scanning activities, are needed to execute a stock take project. And mostly all projects for one customer are executed in a short timeframe, which explains a high demand for employees. For this reason, the Stock Take Company has a large employee pool where they get their workers from. When a Project is planned and needs to be done, the Stock Take Company will open up spots available in an online application, containing the date of the project and name of the project, and these spots will be filled up by workers out of the employee pool who can register to the project. The employees who work for the Stock Take Company are therefore mainly flex workers supplied by either the Stock Take Company's own Employment Agency or external Employment Agencies. There is a constant supply of new unexperienced workers, but most workers have executed projects before. It is important to understand that the workers don't have full time contracts at the Stock Take Company, in fact they are mainly flex workers. Most Project Leaders and IT workers are neither full time employees, but they tend to have more experience. The work that needs to be executed is low skilled work, and every regular person should be able to work for the Stock Take Company, this explains the age differences of the workers ranging from workers being16 years old to workers being 70 years old. The average age of a worker is 29 years. Salary is paid a fixed wage per hour, and is starting at circa 5/6 € and stops at circa 12€. The salary is increasing with age until the worker is an adult. It is important to understand that the workers only are workers working for the Stock Take Company on a flexible base. The workers among others are students, people with a full time job who perform stock take projects as a side activity, or retired people who still want to work. There is no information on motives why workers want to work for the Stock Take Company, but it is clear that these workers are not naturally low skilled workers.

In helping to understand the context see Zielhuis(2017), where similar data from the same company are used for a small research note on the effect of the level of monitoring on the quality performance, and how this interacts with gender. The research was performed in the timeframe from august 2016 to February 2017 , and consisted of 256 projects at 12 different customers/retail firms.

3.2 Experimental design

The field experiment is performed from 9 June 2017 till 13 June 2017, and in this period 82 projects were executed. All projects were stores of one of the Stock Take Company's large customers. This has the advantage that every project is comparable with other projects. The stores are located in the Netherlands and situated all over the whole country.

The projects are divided in three groups. The first group is the control group, the second group is the "feedback" treatment, and the third group is the "bonus pay" treatment. At first the Retail Firm which has been used for the experiment will be shorty described, then the treatments and timing of the experiment is described.

3.2.1 Description of the Retail Firm used for the experiment.

The Retail Firm at which the field experiment has been performed is one of the customers of the Stock Take Company. It has 83 stores located in the Netherlands and 5 large external warehouses, also located in the Netherlands. For the experiment we only use data from the stores, since warehouses differ too much from stores. The working protocol at warehouses is different, and the counting of warehouses is done different as well, this would bias the experiment. However, the store's own warehouses, located at the same location as the store itself are included in the analyses.

The Retail Firm is specialized in selling travelling articles. It is good to know which kind of products are in the store, since counting and scanning these products is the main task of the scanners, and we are interested in their performance.

The stores of the Retail Firm are relative small compared to other customers, and a project can be executed by a team of circa 10 workers. 6 or 7 of these workers are workers who perform the scanning activities. The average shift is similar to the earlier discussed average shift, and lasts for circa 3.7 hours. Mostly after circa 2 hours of work the workers have a break.

The Stock Take Company has designed performance measures in order to communicate to the Retail Firm, but also for planning purposes. Mainly to calculate how many workers are needed to perform a project. This can be done because the Stock Take Company gets information from the Retail Firm on the expected amount of products that are stored in the store when the project is executed, and the Stock Take Company is then able to determine how much workers are needed to count all the products.

The so called calculation norm is therefore the first one of this performance measures, and this is the amount of products an average worker is expected to count per scan hour. A scan hour starts when the worker logs in on his scanning device with his username and log inn code. The calculation norm is set on 650 products per hour per worker. Which is a purely quantitative performance measure.

The second performance measure designed for this specific customer is a qualitative performance measure. This measure is designed in order to guarantee the quality of the stock count to the customer, and is part of the agreements between the two parties. It is called the error rate, which is the percentage of all checked scans that are corrected. It can be calculated over a whole project, but also on personal level. It is obtained by dividing the amount of corrections that are made with the amount of scans that are checked and multiplying by 100. The maximum error rate is set on 0,5 percent. Meaning that if a worker makes 1000 scans, and if 200 of his scans are checked, he is only allowed to make one mistake in all his scans checked.

These performance measurements are used to design the treatments. Hence, the performance measures are only used in the design of the treatments, but are not specifically designed for the treatments. They were only used by the Stock Take Company for planning and contracting purposes, but never used to communicate to the workers.

Furthermore, for this specific Retail Firm, the protocol is that every product needs to be scanned. This means that is not allowed to scan more than one product per scan, even when

there are similar products. If for example in a zone, 10 similar products are stored, then every barcode of each unique product needs to be scanned. It is not allowed to scan one barcode, and fill in 10 pieces on the scanning device. Therefore the average amount of scans per hour and the average amount of products scanned per hour should be equal for a worker.

3.2.2 Feedback and Bonus Pay implementation

For the design of the treatments, the two performance measures are used. In contrast to the control group, where the workers did typically not learn about the performance measures, the workers in the feedback and bonus pay treatments did. For the feedback treatment, the performance measures were communicated right before the project started, and they were told that those were goals set by the firm, and that the workers were expected to obtain those goals. In addition, these workers were also told that they would get personal feedback after circa one hour of work on how the performed in relation with these goals. To recall the exact measures that were communicated as being the targets:

Products Counted per Hour: 650 (calculation norm)

Maximal Error Rate: 0,5 %

The communication with the workers was done by the project leaders of the projects. There was one manager who coordinated all projects, and he sent an email with extra work instructions to the project managers just before the project started. In addition was agreed that an extra text message was send to the project leaders in order to confirm the receiving of the email. This was done on short notice, trying to let other projects unaffected by the treatment of one project by keeping the implementation of the treatments as secret as possible. In the extra instructions (see appendix) was exactly told what the project leaders were expected to do, and accurate steps were given to make them feel comfortable in executing the instructions.

First, the project leaders gathered all the scanners, and communicated the two targets to them. The scanners were told that the Stock Take Company expected these targets from their workers for the specific project. Then the project leader told the workers that they would get feedback on their own performance regarding to the targets after about one hour. Specific instructions in how to get personal performance measures was given accurately described in the extra instructions. The last communication was done in the break, whereby average performance measurements were given to the whole team.

After the project, every project leader needed to fill out an evaluation, and for the ‘‘feedback treatment, some extra questions were added. This information is used to see if projects were executed in the right way, and if the project leader succeeded in implementing the feedback treatment.

The ‘‘Bonus Pay’’ treatment is actually almost similar to the ‘‘Feedback’’ treatment. Again the information communication is done by the project leaders of the project, and the project

leaders are informed by the project coordinator with an email on short notice. The email contained accurate instructions on how to communicate to the workers (see appendix..). Again the workers were informed about the Stock Take Company's targets, and that the workers were expected to reach these targets. But now, the workers were also told that they would get a bonus if they actually succeeded in reaching some new designed targets, also with help of historical data.

Recall the wage per hour which is discussed , was ranging from 5€ till 12€, and the average workhours per shift were 3.7 hours, this provides some information to compare the height of the bonus pays relative to the salary from one shift.

Bonus A = €6,- on top of the workers normal pay out

Calculation norm + 50 = Products Counted per Hour = 650 + 50 = 700

Maximal Error Rate = 0,5%

Bonus B = €8,50 on top of the workers normal pay out

Calulation norm + 150 = Products Counted per Hour = 650 + 150 = 800

Maximal Error Rate = 0,5%

The constraint of the Maximal Error Rate is added to make sure that the extra effort that workers might put in reaching the bonus targets is not reducing the effort put in quality.

The targets for the bonus pay are relatively high, considering the distribution of the Products Per Hour for the similar project performed last year. This partly is a decision of the Stock Take Company to make the Bonus Pay profitable, because it costs money when rewards are given to scanners who perform under the calculation norm, and partly for the reason that scanners are not incentivized to target their performance and then stop putting in extra effort. In the same line of reasoning the second bonus is added ,to keep incentivizing top workers. It is not a perfect design yet, but it is supposed to be a step in the good direction.

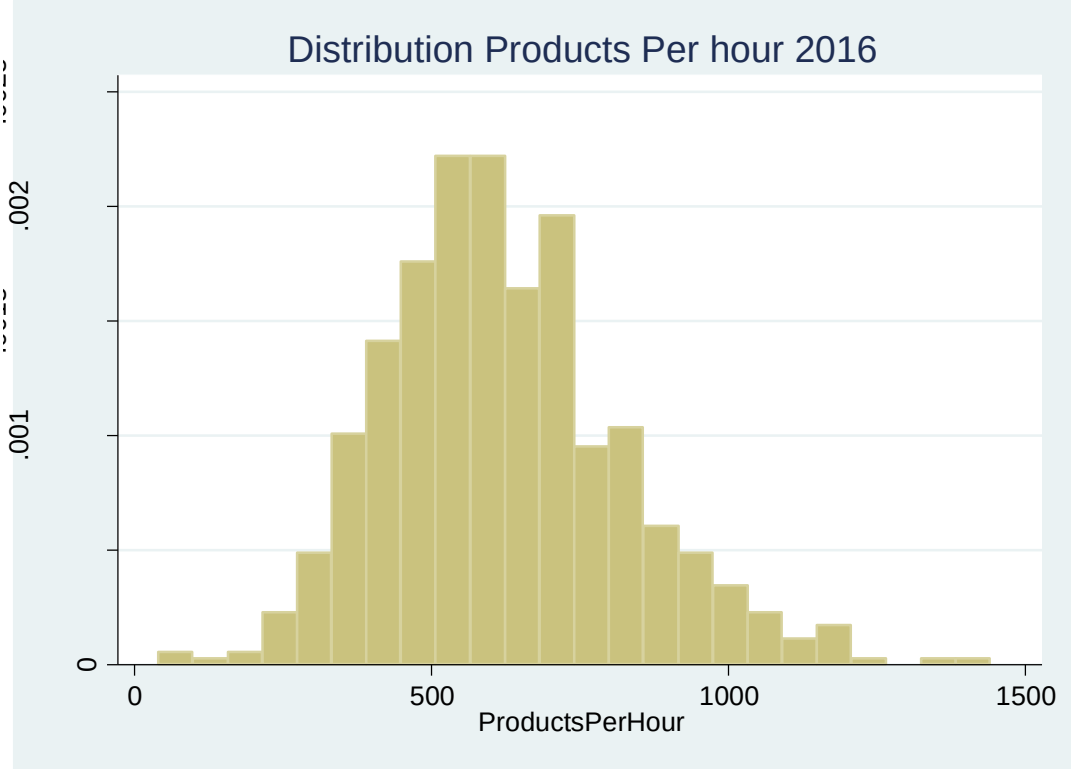

FIGURE 1

To be sure, products per hour is the average amount of products that is scanned per hour by one scanner on one project. In the 2016 project, 40% of the observations would have performed according to at least the calculation norm, 31% of all the observations would have got at least bonus A, and 18% of the observations would have got bonus B. Predicting that the bonus pay will incentivize the workers, it should be a reachable but ambitious bonus target at least for some workers, if it is assumed that the projects in 2016 are comparable to the projects in the field experiment.

Table 1 provides a summarizing scheme of the differences of all treatments, in order to get a clear overview of what is going on.

TABLE 1

Treatment	**Control**	**Feedback**	**Bonus Pay**
Before Project	none	Calculation Norm is communicated Error Rate is communicated Workers are informed that performance related feedback will be given during the project, and before the break	Calculation Norm is Communicatd Error Rate is communicated Workers are informed about the bonus pay and its constraints
Between start project and break	None	Personal performance is communicated, to each worker personal.	None (workers might ask for performance update, no further information on this)
During break	Percentage of all products Scanned is communicated	Average Error Rate is communicated Average Products per Hour is communicated Percentage of all products Scanned is communicated	Average Error Rate is communicated Average Products per Hour is communicated Percentage of all products Scanned is communicated

3.2.3 Planning and Timing

As mentioned before the project took place from 9 June 2016 till 13 June 2016, yet on Friday 16 June there was one additional project which also took place, this one is dropped from the data together with the 5 external warehouses. Because there were workers and project leaders who performed more than one project, it was chosen to distribute the treatments to the projects in a logic way in order to keep as much observations as possible and to let other projects unaffected by the treatment of one project. This was done instead of doing a randomization. Therefore the first and official planning was designed in the way that neither the project leader or the workers in a project faced another sequence than Control → Feedback → Bonus Pay when they performed more than one project. It was therefore not possible to be at first in the Bonus Pay Treatment and in a later project in the control group.

At the start of all the projects, all the control groups mainly took place, during the start and in the middle of all projects, the "Feedback" Treatments took place and at the end of all the projects mainly "Bonus Pay" Treatments took place. For a detailed scheme of the planning see (appendix.) . This planning strategy therefore made possible to use an subject who was in the control group and in the bonus pay group, it was still possible to use an subject who was in the feedback treatment and the bonus pay treatment.

Arguing that an observation can be used that has already been in the control group is obvious. However, arguing that someone who was in the feedback treatment still could be

used in the bonus pay treatment is rather odd. But when looking at the two treatments, it is actually clear, that the bonus pay treatment is almost equal to the feedback treatment, and the only thing added is a bonus pay. This makes reasonable to state that someone who was in the feedback treatment and afterwards in the bonus pay treatment is not biased by the feedback treatment, because he at least gets the same incentives in the bonus pay treatment as in the feedback treatment. In the bonus pay there is also the goal setting before the project and performance related feedback is live available. Only now there is the extra monetary incentive. However there are not much observations taking part in a feedback treatment and a bonus pay treatment. Because a worker can learn about his own performance in the first project due to the feedback, it needs to be said that in arguing that a worker can be used in both feedback treatment and bonus pay treatment , long term effects of the treatments, for example learning effects, are assumed to be zero.

Because all the information and implementation went via the project leaders, the planning was also designed in way that a project leader didn't have to execute more than two different groups. This was thus because the communication of the treatment information to the project leaders was done at rather short notice what caused the risk that project leaders might get confused by all the extra instructions. Although those were very brief and clear.

One might argue that choosing to distribute the treatments to projects in the most logic way is harming the strength of the field experiment in the way that there is no randomization and there is no possibility to control for day fixed effects. This is partly true but there was a trade-off between more observations or a better randomization. To defend this strategy, there is no reason to believe that the assignments of projects to days is not random, following information from the Stock Take Company, in which they stated that projects were planned in the most practical way and that the potential difficultness of a project was not taken into account. If this is the case, the assignment of treatments is still random.

In making sure that there are no day specific effects that could bias the results and to strengthen the argumentation that the assignment was random, the historical data from the 2016 project is used to give this strategy power. In 2016 the same projects were performed from 4 november til 8 november, these data are used to check whether it matters on which day a project is done, and wether we can detect a trend. Looking at the graph below,the bar represents the average products per hour counted on that specific date, and the lines represent the top and bottom of a 5 % confidence interval.

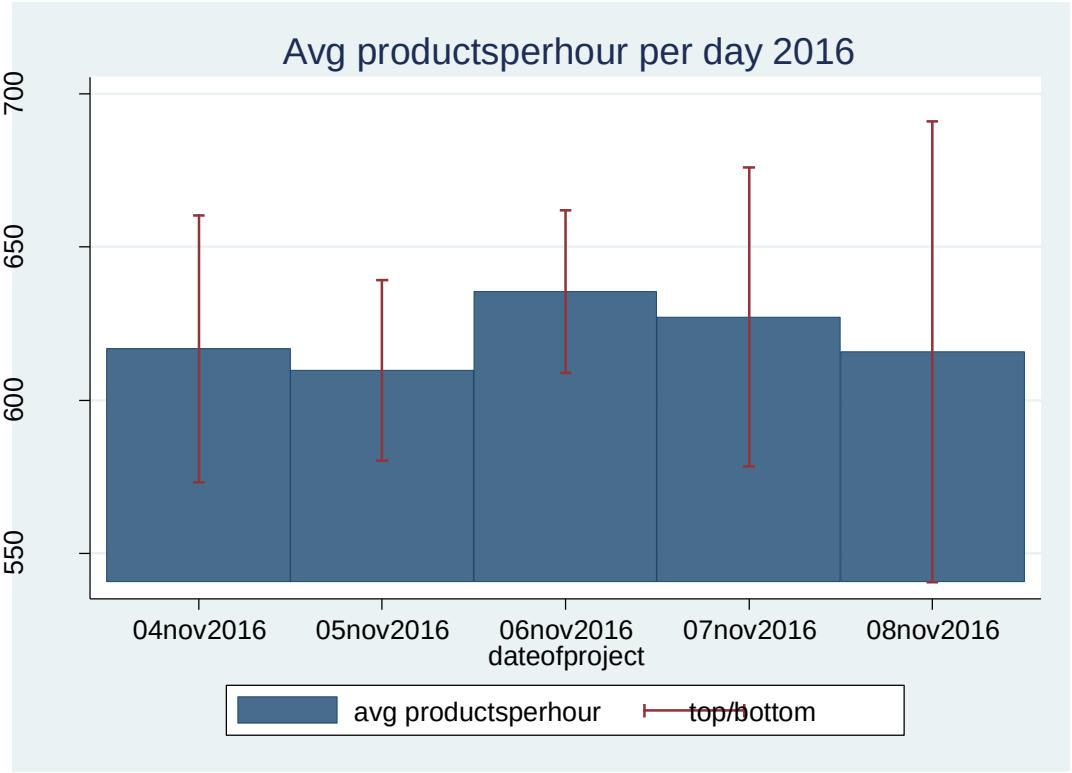

FIGURE 2

We can see that there are no large differences between the days, and in addition a simple OLS regression is run with the same data. Products Per Hour is used as the dependent variable and the regression contained worker fixed effects and project fixed effects plus dummies for every specific date. None of the coefficients for the date specific dummies had significant values. Based on these small checks with the historical data, there is no reason to believe that the way of randomization used in the field experiment in any way biases the results.

Section 4

4.1 Data description

The main data are gathered directly from the field experiment, as mentioned earlier, the workers use scanners to count all the products in the store. Every scan made is going directly into a large database. This causes that the Stock Take Company owns a large data base, with really detailed data. The database can exactly tell at which time a scan was made, by who, which product was scanned, what the product costs, how much products were counted in the scan, whether the scan was checked, whether the scan was corrected by the checker, in which way the scan was corrected etc. This is data on scan level, the data set provides also data on worker level; e.g. hours worked, hours scanned, number of products scanned, number of mistakes made and on project level; e.g. number of workers on the project,

number of hours spent working on the project, how many products were counted on the project etc.

For the empirical analysis, mainly data on worker level is used. This is done because we are interested in the treatment effects on worker level and since all the projects are performed at one customer, the stores are similar stores, selling similar products and we believe there is not much variance between the stores. Despite this, still two variables are created that might be important controls for the project level. In addition we have information from an evaluation form which is filled in by the project leader after a project. This is used as a control for major unlucky events that might have happened during a project which will bias the results of the experiment but which are not caused by implementing the experiment. Examples of these events are bad internet/not working software, bad preparation by the preparation team etc.

The data on person level is merged with a data set containing demographic information as age, gender, employment agency and tenure. The most important variables that are used will be described. Hence we are only interested in workers who scan and count all the products, so we only use data from workers who have at least scanned at a project, during two hours.

TotalHours is the amount of hours a worker has worked on a project, this is the amount of hours the worker gets payed for.

ScanHours is the amount of hours a worker has actually used scanning. This starts when a worker logs in on his scanner and starts using his scanner.

CheckHours is the amount of hours a worker has performed checking activities. This starts when a worker logs in on his tablet and starts checking.

LossHours is the amount of *TotalHours – ScanHours – CheckHours,* this is the time the worker is actually not performing work.

ProductsPerHour is the average amount of products counted per *ScanHour* by the worker, per project.

ScansPerHour is the average amount of scans per *ScanHour* by the worker, per project.

Errorrate is the percentage of corrections made per scan that is checked. These corrections are controlled for mistakes that are not the worker's fault (e.g. wrong bar code at product), so the *Errorrate* is a reliable variable to measure the workers quality performance.

Age is the workers age at the time of the project.

Gender is a dummy variable that takes on the value 1 if the worker is a male and 0 if the worker is a female.

Employment Agency is a dummy that takes on the value 1 if the worker is provided by the Stock Take Company's own Employment Agency and 0 if the worker is provided by an External Employment Agency.

Tenure is the amount of projects that the worker already has performed at the beginning of a project.

FeedbackTreatment is a dummy variable that takes on value 1 when the worker works in a project that received a Feedback Treatment and 0 otherwise.

BonusPayTreatment is a dummy variable that takes on the value 1 when the worker works in a project that received a Bonus Pay Treatment and 0 otherwise.

The two control variables that are created on the project level are *PercentageWarehouse* and *PercentageChecked*. These are respectively the percentage of products in the store that are stored in the store's own warehouse (which is at the same location), and the percentage of scans that are checked. These are added because scanning products in the warehouse is harder than scanning products in the store, and when some stores have a respectively larger warehouse than others, this might bias the results. Furthermore, the percentage of scans checked is a measurement of the percentage of monitoring, which might influence the quality or quantity performance of the workers. However in the research note performed by (Zielhuis, 2017) on these effects, there is no significant evidence for this mechanism to happen.

The Stock Take Company and the Retail Firm have agreed upon a minimum percentage of all scans that should be checked, but this percentage still varies between projects. Mostly because some project leaders performed more quality checks than others, and some employees from the Retail Stores work harder than others. This variable is therefore determined with the data after the project is done. It is clear that the percentage warehouse is given already before a project starts. In Table 2 the summary statistics per treatment group on project level are described.

TABLE 2

Summary Statistics on project level

	Bonuspay	Control	Feedback	Total
Totalscans	12906.8	12690.0	12503.8	12685.3
	(1277.6)	(1604.0)	(1730.8)	(1555.7)
Totalscans checked	4155.3	4666.0	4672.6	4518.9
	(1402.4)	(2365.8)	(1808.4)	(1910.6)
Percentage checked	**32.3**	**37.7**	**38.0**	**36.2**
	(11.2)	**(20.4)**	**(14.8)**	**(16.11)**
Products in store	9535.3	9881.4	9463.1	9627.1
	(815.8)	(1945.6)	(1641.8)	(1565.5)
Products in warehouse	4017.3	3545.9	3853.6	3796.4
	(908.3)	(1671.8)	(1128.7)	(1288.2)
Percentage warehouse	**29.5**	**26.3**	**29.0**	**28.2**
	(4.93)	**(11.1)**	**(7.03)**	**(8.24)**
Total workers on project	9.833	10.75	9.833	10.15
	(1.239)	(1.555)	(1.177)	(1.389)
Total scanners on project	6.167	6.714	6.267	6.390
	(0.565)	(1.863)	(1.413)	(1.421)
Hours Scanned	2.997	3.209	2.922	3.043
	(0.760)	(0.832)	(0.695)	(0.772)
N	24	28	30	82

Standard deviations in parentheses

Percentage Checked and Percentage Warehouse are in bold, because they can be used as Control variables. Hence *PercentageChecked* is calculated by dividing the Total scans Checked by the Total scans and multiplying with 100 and *PercentageWarehouse* by dividing the Products in Warehouse by the total amount of Products in Warehouse and the Products in store and multiplying with 100.

Further, we see that an average project has circa 10 workers, and that on average 6.4 workers perform scanning activities for at least 2 hours. Total scan hours lays around 3 hours per

project. For every variable a one-way ANOVA was conducted to determine if the value of the variable was different for projects within different treatments. A significant difference was found for Total workers on Project at the 5% significance level, however this variable is of no importance for this research and is only provided to get a grasp of how a project looks. For none of the other variables a statistically significant difference was found and therefore it is not allowed to state that the variables differ in values between the groups. For the output of every one way ANOVA test see appendix.

As mentioned, we cleaned the data by dropping all the observations who have scanned less than two hours on a project, this is done because we are only interested in the performance of workers who mainly execute scanning activities and the field experiment is aimed at them. The average of Total working hours on a project was 3,7 hours per project. Also all observations from project leaders are dropped, which is a logic choice. The summary statistics of the cleaned data are described in table 3.

TABLE 3

Summary Statistics worker demographics

	Bonuspay	control	feedback	Total
age	30.72	29.02	29.52	29.67
	(14.07)	(13.69)	(12.64)	(13.43)
tenure	23.58	20.03	20.35	21.14
	(21.57)	(21.83)	(20.29)	(21.23)
gender	0.454	0.453	0.537	0.483
	(0.500)	(0.499)	(0.500)	(0.500)
Own Emp. Agency	0.677	0.682	0.695	0.685
	(0.469)	(0.467)	(0.462)	(0.465)
N 132	172	173	477	

Standard deviations in parentheses

The average age of all workers was circa 30 years, and we see that there is not much difference between the groups. An average worker has already performed 21 projects in the past, right before he started working on the project. We see that for the bonus pay treatment, the tenure is slightly higher, and for the control and feedback a bit lower. Also small a difference is observable for gender, where the feedback group exists for 54 percent out of men, and the bonus pay and control circa for 45 percent. Own employment agency is the percentage for how much percent of the workers are employed by the Stock Take Company's own employment agency, and this around 69 percent on an average, and is not varying so much between the treatment groups. The randomization of the projects is not giving a perfectly balanced sample, suggesting that we should control for these demographics. Still testing if the means within the groups are equal by using a oneway ANOVA test again, doesn't provide enough evidence that the variables tenure and age differ significant between the treatment groups. Since gender and own employment agency are dummy variables, a

Pearson chi square test is run and still there is not enough evidence to state that the variables differ between the treatment groups. This suggests that the allocation of the projects to the treatments is successful in terms of balance, however one should still be careful in stating that the variables are perfectly equal among the treatment groups. See appendix for stata output.

Since most observations are unique and in such a short time frame it is straightforward to mention that the data are cross sectional. This is also because the planning of the treatment assignment made it possible to use an observation again when it already performed a project.

4.2 Empirical Strategy

In estimating the effects of the treatments, ordinary least squares is used. There will be different models that are used to measure the effects, starting off with a simple OLS estimation with a dummy for the "Feedback" treatment and a dummy for the "Bonus Pay" treatment as the explanatory variables. For the dependent variable, three variables are used: *ProductsPerHour, LossHours, Errorrate. ProductsPerHour* is a logical way to measure the effects of both treatments on quantity and the *Errorrate* is a straightforward measurement to estimate the effects of the treatments on quality. *LossHours* is used as a dependent variable to measure whether the effect of the treatment on the worker is actually due to putting in more effort or whether the worker starts using his scanner smarter, by logging out on his scanner when he is not performing scanning work. This will decrease his Scan Hours, but increase his *LossHours*. Since the averages are based on the *ScanHours* the worker might be incentivized to game the system.

Let γ_{ip} be the dependent variable for worker i in project p. What means that γ_{ip} takes on the value of the *ProductsPerHour, LossHours* or *Errorrate* for worker *i* on project *p*. Furthermore, let F_{ip} be a dummy variable, which is equal to one, if person *i* in project *p* was assigned to the Feedback Treatment rather than to the control group or Bonus Pay Treatment. Similary let B_{ip} be a dummy variable, which is equal to one, if person *i* in project *p* was assigned to the Bonus Pay Treatment rather than the control group or Feedback Treatment.

To assess the average effects of the Feedback Treatment and Bonus Pay Treatment we estimate the following:

$$\gamma_{ip} = \alpha + \beta_1 F_{ip} + \beta_2 B_{ip} + \varepsilon_{ip} \qquad (1)$$

Where α is a constant , β_1 is the average effect of the Feedback Treatment relative to the control group and B_2 is the average effect of the Bonus Pay Treatment relative to the control group. The standard error is given by ε_{ip} . Since treatmens are assigned per project at a store is it not possible to include store fixed effects because this would cause multicollinearity, the same line of reasoning can be used to choose not to include day fixed effects, since in the randomization process there is selected on days. In order to control for this, we cluster the standard errors at the project level, which will also help controlling for heteroscedasticity across projects. Adding person specific effects cannot be done either, because this would take a lot of the effects of the treatments, since most observations are unique. In the

robustness check the same estimation will be performed, but with standard errors cluster at the individual level, to account for workers who scan multiple projects.

In addition to the single regression controls are added on worker level and on project level. The final regression, after adding all controls will than look like equitation 2. Including a squared term for age is not implemented, since a scatter plot implied a linear relationship between age and the different dependent variables.

$$\gamma_{ip} = \alpha + \beta_1 F_{ip} + \beta_2 B_{ip} + \beta_3 Age_i + \beta_4 Tenure_i + \beta_5 Gender_i + \beta_6 Ownemp.Agency_i + \beta_7 Percentagechecked_p + \beta_8 Percentagewarehouse_p + \varepsilon_{ip} \quad (2)$$

Again, we are interested in β_1 and β_2, since this coeffients display the effects of the treatments on the dependent variable. Further, $\beta_3 - \beta_6$ are the coefficients for the worker controls, respectively age, tenure, gender and whether or not the worker was employed by the Stock Take Company's own employment agency. Last β_7 and β_8 are the coefficients for the project controls; percentage of all scans that have been checked, and the percentage of all products which are stored in the warehouse.

The literature has widely discussed the use of incentives, how they work, how they can motivate workers or how they might demotivate workers. Even the whole Economics of Organizations literature is replete with research and theories concerning the use of incentives. All this is meant to solve the principal agent theory. However other research has come up with another potential solution. This is selecting the right "type" of worker for the job, and create a perfect worker-job match. This will cause the worker to perform better, because he feels more tied to his job. Akerlof et al. (2005) even goes further, and suggests that workers who feel more identified with their job or employer, and therefore have the right identity, have such a high intrinsic motivation that the principal agent problem is solved. In this thesis, there is no talk of a worker-job match, however, there is reason to believe that there is a clear worker-company match when workers from the Stock Take Companies own employment agency and workers from the external employment agency are compared.

Supposing, that if workers employed from the own employment agency are assumed to feel more identified with the Stock Take Company it is interesting to see if these groups react different to the Feedback or Bonuspay treatment, since following Akerlof et al, they should already be extra intrinsic motivated. The following estimation strategy will be used to test how the workers employed by the own agency company compared to how the workers employed at external agency company react to the treatments. The final regression will look like equitation 3, starting off without controls and by adding the controls step by step.

$$\gamma_{ip} = \alpha + \boldsymbol{\lambda_1 F_{ip}} * \boldsymbol{Ownemp.Agency_i} + \boldsymbol{\lambda_2 B_{ip}} * \boldsymbol{Ownemp.Agency_i} + \beta_1 F_{ip} + \beta_2 B_{ip} + \beta_3 Age_i + \beta_4 Tenure_i + \beta_5 Gender_i + \beta_6 Ownemp.Agency_i + \beta_7 Percentagechecked_p + \beta_8 Percentagewarehouse_p + \varepsilon_{ip} \quad (3)$$

The coefficients of interest are now given by λ_1 and λ_2. Which are the average effects of the treatments for a worker who is employed by the own employment agency compared to a worker who is employed by an external employment agency.

Section 5

5.1 Results

The sample averages per treatment group are displayed in figure 3. First looking at the means of the products per hour, we see that workers in the control group counted on average circa 590 products per hour. However the workers in both the feedback treatment and bonuspay treatment counted on average circa 640 products per hour. The raw figure implies that the workers who got either a feedback treatment or a bonus pay treatment counted on average 50 products more per hour, compared to the control group. This is an increase of 8.5 percent.

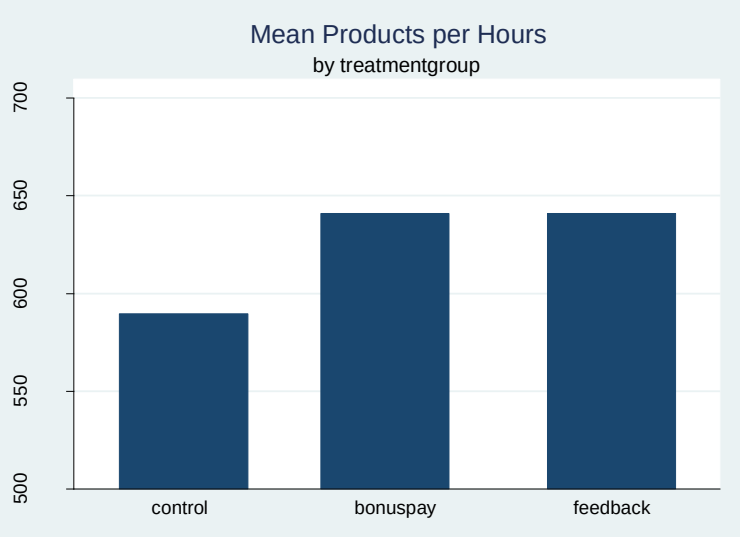

FIGURE 3

Before discussing the empirical results, it is interesting to see whether the difference in the means might be due to multitasking or gaming. As said, the multitask problem is tried to be counteracted with a quality constraint in both treatments, so we should expect no differences in quality performance between the groups. However, there might be some form of gaming when workers used their scanners smarter, by logging out when they were not performing scanning activities, since the targets were based on products per scan hour. This would increase the *Losshours* as discussed before, so it would be interesting to see if the difference

in productivity is due to a kind of gaming. In figure 4 and 5, respectively the means of the *errorrate* and of the *losshours* are displayed.

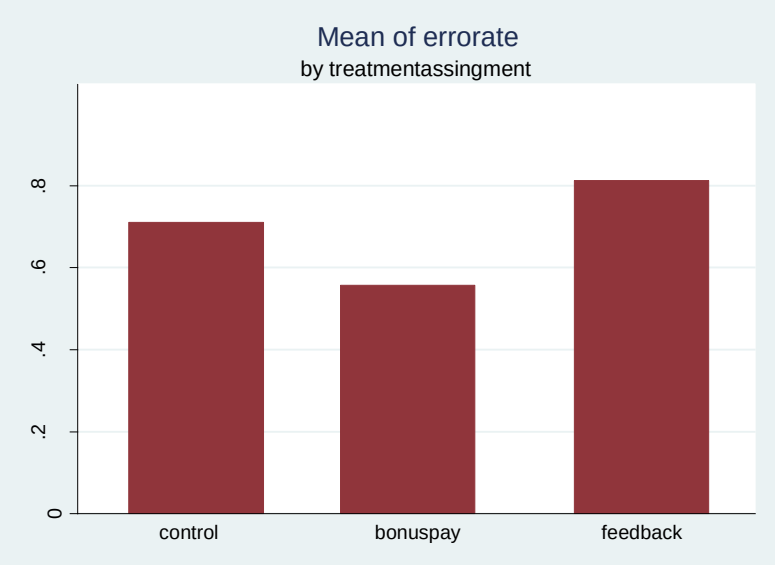

FIGURE 4

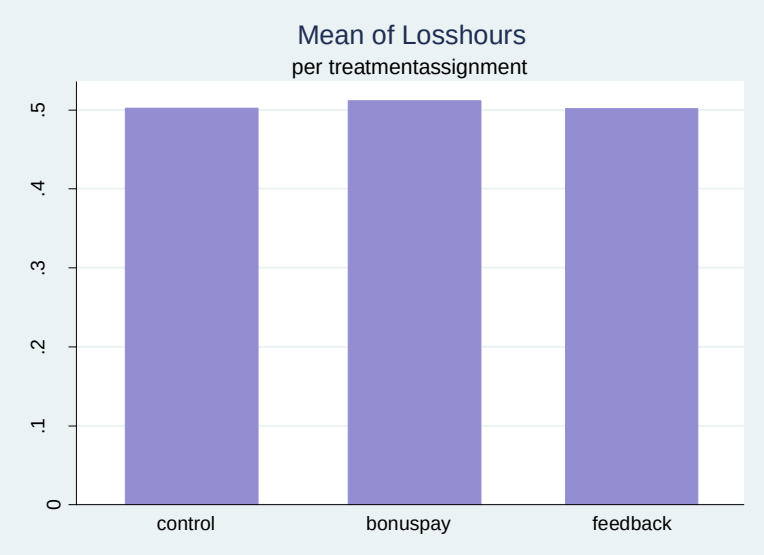

FIGURE 5

At first sight, there is no reason to believe that the positive effects of the treatments are due to gaming, looking at figure 5. However there are minor differences in the error rate averages, if these are significant is questionable, and will be discussed in the empirical analyses.

Table 4 provides all the regressions that are run with *Productsperhour* as a dependent variable, once again, this is the amount of products that a worker counted on a project, divided by the amount of Scan Hours he made at the same project. Model 1, is the estimation as defined in equitation (1) in the description of the empirical strategy. Interpreting the results of model 1,

we see that a worker who was in the feedback treatment, counted on average 51 products more per hour compared to a worker who was in the control group. Exactly the same fact is true for a worker who was in the bonus pay treatment. Despite the loss of degrees of freedom by using robust standard errors, the coefficient for the bonus pay treatment is still significant at a 10 percent level , and the coefficient for the feedback treatment even at the 5 percent significance level, which makes it legit to interpret them. The significance stays the same in model 2 where the demographics of the workers are added. The bonus pay coefficient becomes even significant at the 5 percent level. Which also holds good for model 3 where all the control variables are added to the regressions and the estimation is as defined by equitation (2). The more we control, the more a gap exists between feedback and bonus pay treatments, suggesting that the samples weren't perfectly balanced, which already was noticed at the data description sector. Age turns out to have a negative significant effect on the amount of average products per hour. Someone who is 20 years old counted on average circa 17 products more per hour compared to someone who is 30 years old. As expected tenure has a positive effect, where someone who has performed 10 projects compared to someone who is performing his first project counts on average 9 products more per hour. Percentage checked is also significant, but not of economically importance. Further, the control variable ownemployees turns out to have a large positive significant effect, and will be discussed later in the thesis.

TABLE 4

Dep Variable: Products per Hour	(1)	(2)	(3)
VARIABLES	model 1	model 2	model 3
feedback	51.29**	48.21**	46.95**
	(21.59)	(21.11)	(21.47)
bonuspay	51.45*	54.49**	58.88**
	(26.18)	(24.31)	(24.30)
age		-1.675***	-1.735***
		(0.580)	(0.582)
gender		-13.92	-11.43
		(16.31)	(16.43)
tenure		0.917*	0.914*
		(0.466)	(0.473)
ownemployees		64.22***	77.66***
		(21.71)	(25.19)
percentagechecked			1.055**
			(0.489)
percentagewarehouse			-0.116
			(1.225)
Constant	589.5***	582.0***	538.7***
	(17.61)	(24.74)	(45.37)
Observations	477	464	464
R-squared	0.019	0.074	0.081

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Interpreting the dependent variable model 3, implies that a worker who got a bonus pay if he succeeded in reaching the Stock Take Company's targets, scanned on average 58 products per hour more on a project compared to the control group, which means that this worker works on average at least 10,8 percent harder than a worker in the control group. The result for the

feedback treatment becomes a bit smaller, but there is still a positive effect; A worker who was told the firm's expectations and that he would get feedback on this, counted on average 47 products more per hour, meaning that he worked almost 9,1 percent harder than the average control group worker, everything else kept equal.

TABLE 5

Dep Variable: Error rate VARIABLES	(1) model 1	(2) model 2	(3) model 3
feedback	0.102	0.0354	0.0369
	(0.168)	(0.151)	(0.150)
bonuspay	-0.153	-0.159	-0.157
	(0.138)	(0.128)	(0.122)
age		0.000485	0.000439
		(0.00288)	(0.00298)
gender		0.188*	0.190*
		(0.100)	(0.110)
tenure		-0.000856	-0.000883
		(0.00373)	(0.00381)
ownemployees		-0.469***	-0.463***
		(0.143)	(0.145)
percentagechecked			0.000165
			(0.00599)
percentagewarehouse			-0.000821
			(0.00668)
Constant	0.710***	0.948***	0.961***
	(0.111)	(0.166)	(0.297)
Observations	474	461	461
R-squared	0.007	0.053	0.053

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

As discussed with the graphs earlier in this section, we want to know if these positive effects are actually caused by the incentives, in the sense that workers put in more effort in their tasks and not due to shifting effort from quality work or gaming the system. To check for this all the same regressions are run again but now with the *errorrate* and the *losshours* as dependent variables. Results are displayed in table 5 and table 6.

In both tables, we see no significant coefficient for the effects of the treatments on the error rate or the amount of loss hours. This would suggest that the positive effects on the productivity is not affecting the quality or the difference between the hours worked and the hours registered. Where table 6 is not providing any significant information, however, it interesting to see the signs of the coefficients in table 5, with the error rate as dependent variable. The bonus pay treatment coefficient turns out to have a negative sign, suggesting that workers in the bonus pay treatment, besides putting in more effort, also take more care of quality. This implies that

the quality constraint on the bonus pay targets succeeded its function by concurring the multi task problem. We shouldn't interpret this result as causal but it still is an interesting reaction of the workers.

Looking at the effects of the control variables, there are no important significant effects produced by table 6. In table 5, with the error rate as dependent variable, a significant coefficient for gender is shown, meaning that a male has an error rate of 0.19 percent points higher, if all other factors are kept equal. Implying that gender doesn't matter for quantity but does matter for quality.

Again the dummy variable for ownemployees has a striking significant coefficient. Someone employed for the Stock Take Company's own employment agency has an error rate of circa 0.46 percent points lower compared to someone of the external agency employments. To set this difference in perspective, recall the error rate target of a 0,5 percent! Considering the quantity performance in table 4; a worker employed at the own employment agency counts on average 78 products per hour more compared to someone who is employed at an external employment agency. This is an increase of 14.5 percent! These huge performance differences do give support to Akerlof et al. (2005), in saying that having workers with the right identity is a huge step in solving the principal agency theory. One should still be careful with concluding though, because there is no detailed information available on the skills of the workers.

These effects cannot be explained by experience since there are controls for tenure but it can indeed be that workers from the own employment agency are better trained or more skilled, however, the case might also be that these workers feel more identified with the firm, due to a more personal relation, and therefore have a higher intrinsic motivation. Although all this is speculating, it would be interesting to see if these workers react different to the feedback and or bonus pay treatment. Table 7 provides the results with interaction terms for the own employee and both treatments as described in equitation (3) in the empirical strategy. We only use products per hour as dependent variable, after learning the insignificance of the other two dependent variables.

TABLE 6

Dep Variable: Losshours VARIABLES	(1) model 1	(2) model 2	(3) model 3
feedback	-0.000603	0.000977	0.00796
	(0.0492)	(0.0513)	(0.0496)
bonuspay	0.00976	0.00568	0.0287
	(0.0522)	(0.0543)	(0.0531)
age		0.000960	0.000605
		(0.00103)	(0.000975)
gender		-0.0186	-0.000223
		(0.0308)	(0.0296)
tenure		0.00165*	0.00150
		(0.000952)	(0.000937)
ownemployees		-0.0539	0.00121
		(0.0506)	(0.0506)
percentagechecked			0.00258*
			(0.00138)
percentagewarehouse			-0.00514
			(0.00310)
Constant	0.502***	0.487***	0.501***
	(0.0355)	(0.0570)	(0.108)
Observations	477	464	464
R-squared	0.000	0.015	0.041

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

TABEL 7

Dep Variable: Products per Hour VARIABLES	(1) model 1	(2) model 2	(3) model 3
ownemployeeXfeedback	20.37	25.57	16.04
	(41.60)	(41.82)	(44.09)
ownemployeeXbonuspay	52.87	59.65	46.67
	(53.58)	(53.84)	(53.50)
feedback	33.20	30.83	35.75
	(29.67)	(30.46)	(33.03)
bonuspay	19.09	14.08	26.20
	(43.30)	(43.83)	(43.39)
age		-1.721***	-1.752***
		(0.581)	(0.585)
gender		-14.40	-12.81
		(16.40)	(16.20)
tenure		0.928*	0.934*
		(0.467)	(0.473)
ownemployee	48.68	38.42	56.23
	(29.42)	(28.26)	(35.00)
percentagechecked			0.942*
			(0.502)
percentagewarehouse			0.133
			(1.183)
Constant	556.1***	600.9***	551.3***
	(19.55)	(24.08)	(46.45)
Observations	465	464	464
R-squared	0.055	0.078	0.083

Robust standard errors in parentheses

The results of the interaction terms with own employees and the treatments appear to have insignificant coefficients. However, it is still possible to look at the signs of the coefficients and again these are surprising. It turns out that for both treatments the coefficient with the interaction term is positive. This means that workers of the Stock Take Company's own employment agency react stronger to the incentives compared to workers of external employment agencies. One should be careful in stating this because of the insignificant coefficient, but the insignificance is also partly due to the clustered standard errors on the project level.

It is hard to explain why workers who were assumed to be already highly intrinsic motivated, do also react stronger to incentives in trying to give extra motivation. It is perhaps because

these workers have so much connection with the Stock Take Company, so when the Stock Take Company tries to improve the productivity of the workers they are more willing to collaborate.

Shifting back to the main results, a small cost benefit analyses is performed. The coefficient of table 4, model 3 are used as the increase in productivity because of the incentive implementations. Since has been shown that error rate and loss hours don't have a significant impact, it is reasonable to believe that this increase is because the workers put in more effort. The increase of the productivity is used to calculate the benefits as the amount of total hours saved. For the base rate , the average hours worked per project is taken, which is 3.7 hours . Be aware, the base rate is not the same as the control group, the control group is not used because the amount of products in a store or warehouse do differ between treatment assignments. The average hours worked per project for the control group is 3.8 hours, so the results would not differ so much anyway . The only monetary costs that are made, are the bonus pay outs. A total of 8 workers obtained bonus A, and 10 workers obtained bonus B. Recall bonus A was 6€ and bonus B 8,50€, which makes the total costs sum op to 133€. Table 8 shows the calculation of the benefits.

TABEL 8

Treatment	**Baserate**	**Bonus Pay**	**Feedback**
productivity increase	0%	10%	9%
Hours for same output	3,7	3,34	3,39
benefit in hours	0	0,36	0,31
observations		132	172
Total benefit in hours		47,52	53,32
cost per hour		18,50 €	18,50 €
Total benefit in €		879,12 €	986,42 €

One should take into account that the costs per hours that are saved are inclusive taxes. And that the costs of the bonus pay are without taxes. There is no information how much taxes exactly are. Further, it is assumed that project leaders don't get a bonus, even if they obtained the bonus.

Section 6

6.1 Robustness checks

To give strength to the results discussed, and to make sure that all important issues that might biased results have beendiscussed, a couple of robustness checks are performed. These checks are described in this section.

By agreement with the Stock Take Company, the project coordinator would execute the whole field experiment. There was no further control in this, also because the field experiment was executed as a "pilot" by the Stock Take Company, and this would not be credible if a student would help in executing the "pilot". Still, the agreements were clear,

and the project coordinator would send to all the project leaders the information with the detailed instructions to implement the projects. In addition, as mentioned earlier, an extra text message would be send to the project leaders to get confirmation on receiving the email. Because of varied reasons , the Stock Take Company did not fully succeed in distributing all necessary communication to the project leaders or in compelling all project leaders to execute the extra instructions. It turned out that the strategy of sending the information at short notice turned out not perfectly to succeed.

To control for this, the same regressions are run as in table 4, with products per hour as dependent variable. Instead of using all projects, only the projects that succeeded in executing the treatments were included. This information was gathered by reading the external evaluations that needed to be filled in by the project leaders after the project, and from intern information received from the Stock Take Company. The external evaluations were also checked on unlucky events that might happened during projects , but no significant events happened that might have harmed the projects. In total 9 bonus pay treatments were dropped and 6 feedback treatments needed to be dropped. Results of these estimations are displayed in table 9. Only one project of the control group was dropped, because this project was delayed, however this project had been dropped already from the original data.

TABEL 9

Dep Variable: Products per Hour	(1)	(2)	(3)
VARIABLES	model 1	model 2	model 3
Feedback	46.07*	40.76*	40.64*
	(23.17)	(21.98)	(22.51)
Bonuspay	71.63**	72.11**	74.67***
	(29.61)	(27.76)	(27.43)
Age		-1.733***	-1.757***
		(0.597)	(0.589)
Gender		-7.454	-5.458
		(17.79)	(17.75)
Tenure		1.281**	1.252**
		(0.507)	(0.516)
Ownemployee		52.82**	63.67**
		(20.91)	(25.56)
percentagechecked			0.607
			(0.519)
percentagewarehouse			-0.344
			(1.274)
Constant	589.5***	581.3***	561.7***
	(17.64)	(25.80)	(49.06)
Observations	395	387	387
R-squared	0.027	0.088	0.091

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The coefficient for the feedback treatment decreases, which is not as expected. It stays however positively and significantly at the 10 percent level. Instead of an increase of 9.1 percent as the original test, the effect is now 7.3 percent compared to the control group and everything else kept equal. Why this decreased is hard to explain. It might be that one or two projects were dropped, and that by chance these projects contained workers who were skilled above average. The change is however not large, and the positive effect remains significant. The coefficients for the bonus pay treatment do however increase , after the projects were dropped who did not get a bonus pay although they should have. A worker who was in the bonus pay treatment counted on average 75 products more compared to someone who was in the control group. Implying an increase of 13.2 percent in productivity compared to the control group, everything else kept equal.

That the original results for the feedback treatment tend to be overestimated, is also found in regressing the same estimations again, with the original sample, but with scans per hour as dependent variable instead of products per hour. Officially, following the protocol that is communicated by the Stock Take Company, it was not allowed to scan more than one product per scan, and therefore the amount of scans per hour and the amount of products per hour should be equal. The results of the regressions with scans per hour do however differ somehow from the results of the regressions with products per hour. The coefficient for the bonus pay coefficient increased a bit, and the coefficient of the feedback treatment decreased a bit. The relative change in the reactions are not as strong as in the previous robustness checks, and still both coefficients are significant. The table is reported in the appendix.

As promised in the empirical strategy section, estimations (1) and (2) are performed again but now with the standard errors clustered at the individual level instead of project level, in order to control for workers who worked on multiple projects. This produces almost similar results as the original results, except for the significance of the coefficients. The significance increased as a logical reaction to the increase in the degrees of freedom. Suggesting that results are not biased by workers who work on multiple projects.

To strengthen the strategy used to assign the treatments to projects, several tests are performed with day fixed effects for 9, 10 and 11 June. On these days there was enough variation between treatments, and it was therefore possible to check if there were day specific effects which are not dealt with in the field experiment. It is straightforward to explain that 12 and 13 June couldn't be included in these tests, since on those day only bonus pay treatments took place. With products per hour as dependent variable, a simple regression was done with 9 and 10 June as explanatory variables and 11 June as the base date. In addition to this regression, also individual fixed effects and project fixed effects were added. It is clear that adding the individual and project fixed effects causes biases, since projects and workers were treated differently, but this is only done to test whether there exist significant day specific effects. As a last test, the day fixed effects are added to the original estimations discussed in the empirical strategy. Not once did the tests produce significant coefficients, which provides a good argument to state that the randomization of treatments to projects is successful.

The cost-benefit analyses used the variable Total Hours as a measure to calculate the benefits of the experiment. This was done by calculating how much costs were saved by not paying

out the hours that were saved. Because the average total hours were used as the base rate for the calculations, it is interesting to see if the real amount of hours that were saved do confirm the cost benefit calculations. Table 10 provides the regressions result from a regression with total hours as dependent variable and the treatments as explanatory variables. The absolute values of amount of hours saved are much lower than the cost-benefit analyses suggested. The coefficient for the bonus pay treatment tells that on average 0.176 hours are saved and the coefficient for the feedback treatment tells that 0.172 hours are saved, everything else kept equal. The cost benefit analyses predicted that 0.36 hours would be saved for the bonus pay treatment and 0.31 hours for the feedback treatment. However, estimating the relative effects gives almost equal results as the cost-benefit analyses. Following table 10, the relative increase in productivity that cause the decrease in total hours is for both treatments almost 10 percent. Which is comparable to the increases in the productivity used in the cost-benefit analyses, which were 10 percent for the bonus pay treatment and 9 percent for the feedback treatment. The differences in absolute values, by comparing the results from the regressions, and the calculations of the cost-benefit analyses is mainly due to the fact that both treatment groups had on average more products per worker per project to be counted. *Prodstoreperperson* is the amount of products per person for that specific project stored in the store and *prodwareperperson* is the amount of products per person stored in the store's own warehouse. These are included in the regressions as control variables.

TABEL 10

Dep Variable: Total Hours VARIABLES	(1) model 1
bonuspay	-0.176**
	(0.0760)
feedback	-0.172**
	(0.0708)
prodstoreperperson	0.00107***
	(0.000103)
prodwareperperson	0.000746***
	(0.000163)
Constant	1.973***
	(0.183)
Observations	573
R-squared	0.177

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The latest test being performed is inspired by Bandiera et al. (2007). Bandiera finds that, when a relative performance pay is implemented, and workers work with relatively more friends, the workers' productivity decreased. There is no detailed information on which workers are friends with each other in this thesis setting, but the interesting split between the employment agencies is still available. If it is assumed that workers of the own employment agency are relatively close with each other compared to workers of the external employment agencies, it is possible to test if the mechanism found by Bandiera, also exists in the current context. Therefore the variable *owncol* is created, which is the percentage of all workers who are employed at the Stock Take Company's own employment agency. All the observations of

the external employment agencies are dropped from the data after the creation of the variable *owncol* because there are multiple external employment agencies and the variable *owncol* would have a reversed effect on them. This doesn't matter since the only interest is in testing the mechanism. Table 11 shows the results of the regressions that are run, and consists interaction terms with *owncol* and the treatments.

TABLE 11

Dep Variable: Products Per Hour	(1)	(2)
VARIABLES	model 1	model 2
feedback	298.5***	292.3***
	(106.6)	(100.6)
bonuspay	457.8***	435.8***
	(111.7)	(102.7)
owncolXfeedback	-234.3**	-236.1**
	(115.1)	(108.5)
owncolXbonuspay	-382.3***	-363.9***
	(120.4)	(115.9)
age		-1.511**
		(0.634)
gender		-20.30
		(20.74)
tenure		1.261***
		(0.417)
percentagechecked		1.082
		(0.900)
percentagewarehouse		0.0832
		(1.602)
owncol	295.4***	313.3***
	(92.06)	(87.62)
Constant	319.7***	287.3***
	(86.78)	(102.1)
Observations	408	408
R-squared	0.076	0.118

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The table is hard to interpret, but at least it is shown that the coefficients of the interaction terms have significant negative values. Suggesting that there is some similarity with the findings by Bandiera. Because most project leaders are employed at the Stock Take Companies own employment agency, a second regression is run without the project leaders. This didn't matter for the sign and significance of the coefficients, which did not change.

Section 7

7.1 discussion

This thesis studied how communicating performance related targets, before workers start on their job, and giving them feedback regarding these targets or giving them a monetary reward if they succeed in obtaining these targets, help to increase productivity. By using a field experiment at a large Firm specialized in Stock Take services the two different incentives were tested.

The Feedback treatment could be described as a partly intrinsic and partly extrinsic way of motivating workers, and the Bonus Pay treatment could be described as a mainly extrinsic way to motivate workers. Both incentives turned out to have a positive effect on the productivity of the workers, which is in line with the discussed literature. Offering a bonus pay to the workers, caused workers to increase their productivity with 10.8 percent, and additional tests proved that this result was underestimated, which means that the monetary reward increased the workers' productivity with at least 10.8 percent. Also the workers in the feedback treatment increased their productivity, but not as much as with the bonus pay treatment. The official result was an increase of 9.1 percent compared to the control group, however some robustness checks suggested that this was a bit overestimated. Following the results of the robustness checks, the expected effect of the feedback treatment is on average between 7.3 and 9.1 percent.

The positive effects of the given incentives is almost only caused because workers put in more effort as a reaction to obtain the targets. It is shown that these effects are not because workers shift effort from quality to quantity performance, or because workers game the system by optimizing their use of Scan Hours. Workers in the bonus pay treatment did even increase their quality performance, but not with significant values.

The workers reacted to the incentives as expected, following the literature. But the result that workers who got a monetary reward increased their productivity relatively more than workers who only got feedback is not fully in line with the literature on intrinsic and extrinsic motivation. E.g. Delfgaauw et al. (2013) found that there was no difference in the size of the effect between workers in a feedback treatment or workers in a monetary reward treatment compared to a control group. Also other literature suggests that intrinsic motivation has equal or sometimes a better effect than extrinsic motivations. This can be explained by the nature of the work the workers performed in the Stock Taking context . The work description is clear, productivity is easy to measure, and any moderate person is able to perform the work without much schooling. It is therefore harder to gain intrinsic value out of the job. Recall Delfgaauw et al (2013) , where subjects were salespersons. It is clear that selling gives a higher satisfaction when a worker succeeds, and that intrinsic motivation in this setting is relatively more important. It might be that because the nature of the stock taking work is relative easy compared to other work activities, and that therefore workers react relative stronger to monetary incentives.

A last important finding is the finding that workers who worked for the Stock Take Company's own employment agency had a much higher productivity compared to workers of an external employment agencies. In addition, the reaction to both treatments was also stronger for the Stock Take Company's own employment agency employees compared to the external employment agency employees, this result however was not significant. These results

suggest that the worker-company relationship is an important relationship, and that this relationship should not be undervalued.

By using a natural field experiment as the methodology to test the incentives and in combination with the existent of a control group, it is reasonable to conclude that the results are a trustworthy reflection of the change in the workers' behaviour. But still, one should be careful in interpreting all results as causal. First they do hold good for this specific experiment at this specific firm, but one should take the external validity into account when making conclusions. Second, the timeframe of the field experiment was too narrow to into account the long term effects of the given incentives. Third, also due to the narrow time frame of the experiment, it was not possible to observe subjects enough times to get a trustworthy measure of individual effects. Fourth, in the cost-benefit analyses, the only costs accounted for are monetary costs, but one should also take other costs in account, for example the training of project leaders or overhead costs.

7.2 Further Research

The obvious suggestion for further research is implementing the experiment over a wider time frame, this will increase the robustness of the results and give a better understanding of the mechanisms that take place. In addition it will solve the problem that there is no control for personal capabilities.

The detailed data that the Stock Take Company owns gives also broad possibilities for more detailed data research, but also for creating dynamic performance targets, to increase the fairness of these targets. Since information is available about every worker, every scan, every zone, every product, there are unlimited ways to improve the accuracy and the fairness of the performance measures. This can for example be done by adding multipliers which display the difficulty of a zone, and base performance measures on those outcomes, instead of just taking averages of the whole store. In line with this, some project leaders wrote suggestions in the external evaluation, to split the products in the store and the products in the store's warehouse in different categories and take this into account when measuring the performance, which is a first step in creating accurate performance measures.

A last suggestion is based on the finding regarding the huge difference in productivity by workers of the Stock Take Company's own employment agency and the external employment agencies. Assuming that the workers of the own employment agency have a better relationship with the Stock Take Company, it suggests that a good worker-company match can be a good way in preventing the principal agency problems. If it is possible to measure the real identity of workers with help of surveys for example, then can be determined how great the effects of the right identity and the perfect worker-company or worker-manager match can be. Matching workers with the right identity to the right company could possibly have much greater impacts than the use of incentives.

Section 8

8.1 Appendix: Literature

Akerlof, G. A., & Kranton, R. E. (2005). Identity and the Economics of Organizations. The Journal of Economic Perspectives, 19(1), 9-32.

Albanese, R., & Van Fleet, D. D. (1985). Rational behavior in groups: The free-riding tendency. *Academy of Management review*, *10*(2), 244-255.

Ariely, D., Gneezy, U., Loewenstein, G., & Mazar, N. (2009). Large stakes and big mistakes. *The Review of Economic Studies*, *76*(2), 451-469.

Arrow, K. J. (1968). The economics of moral hazard: further comment. *The American Economic Review*, *58*(3), 537-539.

Ashton, R. H. (1990). Pressure and performance in accounting decision settings: Paradoxical effects of incentives, feedback, and justification. *Journal of Accounting Research*, 148-180.

Baker, G. (1992). Incentive Contracts and Performance Measurement. *Journal of Political Economy, 100*(3), 598-614.

Barankay, I. (2011). Rankings and social tournaments: Evidence from a crowd-sourcing experiment. *Wharton School of Business, University of Pennsylvania Working Paper*.

Becker, L. J. (1978). Joint effect of feedback and goal setting on performance: A field study of residential energy conservation. *Journal of applied psychology*, *63*(4), 428.

Berlin, J. M. (2014). common incentives for teamwork - the unspoken contracts significance . Team Performance Management , 65-80

Blanes i Vidal, J., & Nossol, M. (2011). Tournaments without prizes: Evidence from personnel records. *Management science*, *57*(10), 1721-1736.

Bradler, C., Dur, R., Neckermann, S., & Non, A. (2016). Employee recognition and performance: A field experiment. *Management Science*, *62*(11), 3085-3099.

Brickley, J. A., & Zimmerman, J. L. (2001). Changing incentives in a multitask environment: evidence from a top-tier business school. *Journal of Corporate finance*, *7*(4), 367-396.

Deci, E. L., & Ryan, R. M. (1985). The general causality orientations scale: Self-determination in personality. *Journal of research in personality*, *19*(2), 109-134.

Delfgaauw, J., Dur, R., Sol, J., & Verbeke, W. (2013). Tournament incentives in the field: Gender differences in the workplace. *Journal of Labor Economics*, *31*(2), 305-326.

DeVaro, J. (2006). Strategic promotion tournaments and worker performance. *Strategic Management Journal*, *27*(8), 721-740.

Donato, K., et al. "Personality Traits and Performance Contracts: Evidence From A Field Experiment Among Maternity Care Providers In India." American Economic Review (2017).

Eisenhardt, K. M. (1989). Agency theory: An assessment and review. *Academy of management review*, *14*(1), 57-74.

Fehr, E., & Schmidt, K. M. (2004). Fairness and Incentives in a Multi-task Principal–Agent Model. *The Scandinavian Journal of Economics*, *106*(3), 453-474.

Frey, B. S., & Jegen, R. (2001). Motivation crowding theory. *Journal of economic surveys*, *15*(5), 589-611.

Gibbons, R. (2005). Four formal (izable) theories of the firm?. *Journal of Economic Behavior & Organization*, *58*(2), 200-245.

Grant, A. M., & Gino, F. (2010). A little thanks goes a long way: Explaining why gratitude expressions motivate prosocial behavior. *Journal of personality and social psychology*, *98*(6), 946.

Gneezy, U., Meier, S., & Rey-Biel, P. (2011). When and why incentives (don't) work to modify behavior. *The Journal of Economic Perspectives*, *25*(4), 191-209.

Hall, M. Lazear, 2000, Performance Pay at Safelite Auto Glass. *Harvard Business School, Harvard Business School Case*, *800291*.

Hamilton, B. H., Nickerson, J. A., & Owan, H. (2003). Team incentives and worker heterogeneity: An empirical analysis of the impact of teams on productivity and participation. *Journal of political Economy*, *111*(3), 465-497.

Harrison, G. W., & List, J. A. (2004). Field experiments. *Journal of Economic literature*, *42*(4), 1009-1055.

Holmstrom, B., & Milgrom, P. (1991). Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design. *Journal of Law, Economics, & Organization*, *7*, 24-52.

Huffman, D., & Bognanno, M. (2015). Performance Pay and Workers' Non-Monetary Motivations: Evidence from a Natural Field Experiment.

Ivancevich, J. M., & McMahon, J. T. (1982). The effects of goal setting, external feedback, and self-generated feedback on outcome variables: A field experiment. *Academy of Management Journal*, *25*(2), 359-372.

Jensen, M. C., & Meckling, W. H. (1976). Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of financial economics*, *3*(4), 305-360.

Laffont, J. J., & Martimort, D. (2009). *The theory of incentives: the principal-agent model*. Princeton university press.

Lazear, E. P., & Gibbs, M. (2010). Personnel Economics in Practice . United States of America : wiley .

Marschak, J. (1955). Elements for a theory of teams. *Management science*, *1*(2), 127-137.

Milgrom, P., & Roberts, J. (1988). An Economic Approach to Influence Activities in Organizations.*American Journal of Sociology, 94*, S154-S179

Nelson, B. (2005). 1001 Ways to Reward your Employees. Vol. 2.

Oosterbeek, H., Sloof, R., & Sonnemans, J. H. (2006). Rent-seeking versus Productive Activities in a Multi-task Experiment.

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.

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*, 643-667.

Prendergast, C. (1993). The Role of Promotion in Inducing Specific Human Capital Acquisiton . The Quarterly Journal of Economics , 523-543.

Ross, S. A. (1973). The economic theory of agency: The principal's problem. *The American Economic Review*, *63*(2), 134-139.

Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary educational psychology*, *25*(1), 54-67.

Shearer, B. (2004). Piece rates, fixed wages and incentives: Evidence from a field experiment. *The Review of Economic Studies*, *71*(2), 513-534.

Stalnaker, R. (2006). Saying and meaning, cheap talk and credibility. In *Game theory and pragmatics* (pp. 83-100). Palgrave Macmillan UK.

Stajkovic, A. D., & Luthans, F*. (2003). Behavioral management and task performance in organizations: conceptual background, meta-analysis, and test of alternative models. Personnel Psychology, 56(1), 155-194.*

Thomas, K. W. (2009) The four intrinsic rewards that drive employee engagement. *Ivey business journal*

Zielhuis, G. E. (2017) Performance of workers and gender differences in stock taking environment

8.2 Appendix: Additional Tables

8.2.1 One way ANOVA project level

```
. oneway totalscans treatmentassignment, tab
```

treatmentassignment	Summary of AantalScansTotaal		
	Mean	Std. Dev.	Freq.
bonuspay	12906.792	1277.5817	24
control	12690.036	1603.9658	28
feedback	12503.767	1730.7549	30
Total	12685.329	1555.7164	82

Analysis of Variance

Source	SS	df	MS	F	Prob > F
Between groups	2166663.82	2	1083331.91	0.44	0.6447
Within groups	193873870	79	2454099.62		
Total	196040534	81	2420253.51		

```
Bartlett's test for equal variances:  chi2(2) =   2.2853  Prob>chi2 = 0.319
```

```
. oneway totalscanschecked treatmentassignment, tab
```

treatmentassignment	Summary of AantalScansGeconroleerd		
	Mean	Std. Dev.	Freq.
bonuspay	4155.3333	1402.4045	24
control	4665.9643	2365.7814	28
feedback	4672.6	1808.4301	30
Total	4518.939	1910.5964	82

Analysis of Variance

Source	SS	df	MS	F	Prob > F
Between groups	4486629.2	2	2243314.6	0.61	0.5466
Within groups	291194037	79	3686000.47		
Total	295680667	81	3650378.6		

```
Bartlett's test for equal variances:  chi2(2) =   6.5283  Prob>chi2 = 0.038
```

```
. oneway percentagechecked treatmentassignment, tab

treatmentas |     Summary of percentagechecked
   signment |        Mean   Std. Dev.       Freq.
------------+------------------------------------
   bonuspay |   32.301267     11.2171          24
    control |   37.716142   20.425385          28
   feedback |   38.011907   14.800147          30
------------+------------------------------------
      Total |   36.239507   16.116111          82

                          Analysis of Variance
    Source              SS         df      MS            F     Prob > F
------------------------------------------------------------------------
Between groups      527.528365       2   263.764182      1.02     0.3667
 Within groups      20510.5241      79   259.626887
------------------------------------------------------------------------
    Total           21038.0524      81   259.729042

Bartlett's test for equal variances:  chi2(2) =   8.6359  Prob>chi2 = 0.013

. oneway productsstore treatmentassignment, tab

treatmentas |   Summary of AantalProductenWINKEL
   signment |        Mean   Std. Dev.       Freq.
------------+------------------------------------
   bonuspay |   9535.3333   815.83282          24
    control |   9881.4286   1945.5983          28
   feedback |   9463.0667    1641.831          30
------------+------------------------------------
      Total |   9627.0732   1565.4686          82

                          Analysis of Variance
    Source              SS         df      MS            F     Prob > F
------------------------------------------------------------------------
Between groups      2820439.5       2   1410219.75      0.57     0.5682
 Within groups      195685602      79   2477032.94
------------------------------------------------------------------------
    Total           198506042      81   2450691.87

Bartlett's test for equal variances:  chi2(2) =  16.0494  Prob>chi2 = 0.000

. oneway productswarehouse treatmentassignment, tab

treatmentas | Summary of AantalProductenMAGAZIJN
   signment |        Mean   Std. Dev.       Freq.
------------+------------------------------------
   bonuspay |   4017.2917   908.29872          24
    control |   3545.8929   1671.8114          28
   feedback |   3853.5667   1128.7063          30
------------+------------------------------------
      Total |   3796.4268   1288.1724          82

                          Analysis of Variance
    Source              SS         df      MS            F     Prob > F
------------------------------------------------------------------------
Between groups      3026183.06       2   1513091.53      0.91     0.4068
 Within groups      131384249      79   1663091.76
------------------------------------------------------------------------
    Total           134410432      81   1659388.05

Bartlett's test for equal variances:  chi2(2) =   9.6718  Prob>chi2 = 0.008
```

```
. oneway totalworkersonproject treatmentassignment, tab
```

treatmentassignment	Summary of totalworkersonproject Mean	Std. Dev.	Freq.
bonuspay	9.8333333	1.2394482	24
control	10.75	1.5545632	28
feedback	9.8333333	1.1768846	30
Total	10.146341	1.3888618	82

Analysis of Variance

Source	SS	df	MS	F	Prob > F
Between groups	15.4939024	2	7.74695122	4.35	0.0162
Within groups	140.75	79	1.78164557		
Total	156.243902	81	1.92893707		

```
Bartlett's test for equal variances:  chi2(2) =   2.4382  Prob>chi2 = 0.295

. oneway totalscannersonproject treatmentassignment, tab
```

treatmentassignment	Summary of totalscannersonproject Mean	Std. Dev.	Freq.
bonuspay	6.1666667	.5646597	24
control	6.7142857	1.863035	28
feedback	6.2666667	1.4125871	30
Total	6.3902439	1.4207987	82

Analysis of Variance

Source	SS	df	MS	F	Prob > F
Between groups	4.59790941	2	2.2989547	1.14	0.3241
Within groups	158.914286	79	2.01157324		
Total	163.512195	81	2.01866908		

```
Bartlett's test for equal variances:  chi2(2) =  27.4038  Prob>chi2 = 0.000
```

8.2.2 one way ANOVA worker demographics

```
. oneway age treatmentassignment, tab

treatmentas |           Summary of age
   signment |        Mean   Std. Dev.       Freq.
------------+------------------------------------
   bonuspay |   30.724161   14.069056         130
    control |   29.020816   13.690662         170
   feedback |   29.516335    12.64375         164
------------+------------------------------------
      Total |   29.673187    13.42834         464

                       Analysis of Variance
    Source              SS         df      MS             F     Prob > F
------------------------------------------------------------------------
Between groups      219.975985       2   109.987992      0.61     0.5444
 Within groups      83268.3305     461   180.625446
------------------------------------------------------------------------
    Total           83488.3065     463   180.320316

Bartlett's test for equal variances:  chi2(2) =   1.8360  Prob>chi2 = 0.399

. oneway tenure treatmentassignment, tab

treatmentas |          Summary of tenure
   signment |        Mean   Std. Dev.       Freq.
------------+------------------------------------
   bonuspay |   23.592308   21.517866         130
    control |   20.029412   21.775462         170
   feedback |   20.175758   20.340381         165
------------+------------------------------------
      Total |   21.077419   21.216211         465

                       Analysis of Variance
    Source              SS         df      MS             F     Prob > F
------------------------------------------------------------------------
Between groups      1143.06462       2   571.532312      1.27     0.2815
 Within groups      207716.148     462   449.602053
------------------------------------------------------------------------
    Total           208859.213     464   450.127614

Bartlett's test for equal variances:  chi2(2) =   0.8488  Prob>chi2 = 0.654
```

8.2.3 Chi^2 test Worker Demographics

```
. tab treatmentassignment gender, chi2
```

treatmenta ssignment	gender 0	1	Total
bonuspay	71	59	130
control	93	77	170
feedback	77	88	165
Total	241	224	465

Pearson chi2(2) = 2.7291 Pr = 0.256

```
. tab treatmentassignment ownemployee, chi2
```

treatmenta ssignment	ownemployees 0	1	Total
bonuspay	42	88	130
control	54	116	170
feedback	51	114	165
Total	147	318	465

Pearson chi2(2) = 0.0686 Pr = 0.966

8.2.4 Scans per Hour as Dep Variable.

Dep Variable: Scans Per Hour	(1)	(2)	(3)
VARIABLES	model 1	model 2	model
feedback	42.18*	38.60*	37.68*
	(21.34)	(20.51)	(20.82)
bonuspay	54.36**	59.03**	62.87***
	(26.02)	(23.28)	(23.53)
age		-1.500***	-1.553***
		(0.562)	(0.564)
gender		-16.11	-13.87
		(15.56)	(15.87)
tenure		0.394	0.390
		(0.479)	(0.480)
ownemployees		78.58***	90.12***
		(21.51)	(24.81)
percentagechecked			0.884*
			(0.498)
percentagewarehouse			-0.162
			(1.019)
Constant	560.1***	549.0***	514.4***
	(17.64)	(23.55)	(41.04)
Observations	477	464	464
R-squared	0.018	0.075	0.080

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

8.3: Appendix: experimental design

8.3.1: Extra instruction project leaders bonus pay treatment

Werkinstructie projectleider
Voorraadinventarisatie 2017
CUSTOMER A – Company A

Aanvullende werkinstructie bonusbetaling

Company A voert tijdens het CUSTOMER A project een pilot uit met een bonusbetaling. De telmedewerkers zijn hier niet vooraf van ingelicht. Jij bent de eerste die het aan hen verteld. Deze extra werkinstructie dient als handleiding voor het informeren van de extra bonusbetaling aan de telmedewerkers.

De bonusbetaling is onderdeel van een pilot van Projectservice. Deze pilot onderzoekt mogelijkheden om de arbeidsproductiviteit de telmedewerkers te verhogen.

Bij aanvang van het project

- Verzamel alle telmedewerkers, en zorg dat je de aandacht hebt.
- Leg vervolgens uit wat van hen verwacht wordt, en wanneer ze aanspraak maken op de bonus:

Calculatienorm:

"Company A houdt er rekening mee dat jullie een bepaald aantal artikelen per uur scannen. Het is een maatstaf waarvan Company A verwacht dat de gemiddelde telmedewerker dit haalt; de calculatienorm. Deze norm is specifiek afgesteld voor dit huidige project en is gesteld op 650 artikelen per scan uur.

Foutenpercentage:

"Company A heeft ook afspraken gemaakt met de CUSTOMER A wat betreft de kwaliteit. Naast dat hier afspraken over hebben gemaakt met CUSTOMER A , is het voor Company A zelf ook een
prioriteit dat de kwaliteit van de tellingen uitstekend is. Hierom is er een maximaalfoutenpercentage gesteld van 0,5 %.

Bonus

- Informeer de telmedewerkers dat Company A voor dit specifieke project een bonus uitbetaald als de volgende normen worden gehaald.
  - o Bonus A bij een scansnelheid hoger dan 700 scans per uur.
    - Bonus van € 6,- o Bonus B bij een scansnelheid hoger dan 800 scans per uur.
    - Bonus van € 8,50
- Belangrijk: Naast de scansnelheid moet het foutenpercentage onder de 0,5% zijn.
- De telmedewerker moet minimaal 1 uur hebben gescand.
- De bonus wordt samen met het normale salaris uitgekeerd.

TIJDENS HET PROJECT

- Tijdens het project is het de taak van de projectleider dat de telmedewerkers random worden ingedeeld in de winkel. Voor het behalen van de bonus is het vanzelfsprekend dat telmedewerkers aantrekkelijke zones willen tellen. Pas er daarom op dat dit zo min mogelijk gebeurd.

8.3.2 Extra instructions project leader feedback treatment

Werkinstructie projectleider
Voorraadinventarisatie 2017
CUSTOMER A – Company A

Aanvullende werkinstructie projectleider

Deze extra werkinstructie heeft het doel om tellers meer te betrekken bij de werkzaamheden en de tellers meer verantwoording te geven voor hun prestaties. Verder is het de bedoeling de telmedewerkers extra te motiveren. De instructies zijn onderdeel van een pilot die Company A uitvoert en gelden alleen voor dit specifieke project. Het is niet de bedoeling dit op andere projecten toe te passen.

Telmedewerkers hebben meerdere malen aangegeven er baat bij te hebben als er meer feedback wordt gegeven op de eigen prestaties. Deze instructie geeft een handleiding hoe de telmedewerkers benaderd dienen te worden, en hoe de feedback gegeven dient te worden.

Bij aanvang van het project

- Verzamel alle telmedewerkers, en zorg dat je de aandacht hebt.
- Vervolgens deel je de telmedewerkers mee wat de maatstaven zijn waarmee de klant en Company A rekening houdt:

Scansnelheid:

> "Company A houdt er rekening mee dat jullie een bepaald aantal artikelen per uur scannen. Het is een maatstaf waarvan Company A verwacht dat de gemiddelde telmedewerker dit haalt; de calculatienorm. Deze norm is specifiek afgesteld voor dit huidige project en is gesteld op 650 artikelen per scan uur.

Foutenpercentage:

> "Company Aheeft ook afspraken gemaakt met de CUSTOMER A wat betreft de kwaliteit. Naast dat hier afspraken over hebben gemaakt met CUSTOMER A , is het voor Company ook een
> prioriteit dat de kwaliteit van de tellingen uitstekend is. Hierom is er een maximaalfoutenpercentage gesteld van 0,5 %.

Feedback geven

Als laatst wijs je alle telmedewerkers er op dat jezelf als projectleider, de telmedewerkers persoonlijk nog een keer benaderd tijdens het project. Hierin vertel je de telmedewerker hoe het er voor staat met zijn persoonlijke telsnelheid en zijn persoonlijke kwaliteit.
Het is belangrijk dat voor dat het werk begint de telmedewerkers duidelijk wordt gemaakt dat deze feedback gegeven

TIJDENS HET PROJECT

- Naast de reguliere taken als projectleider, verwachten we dat je bij dit project een extra taak uitvoert. Deze taak is het geven van persoonlijke feedback aan elke telmedewerker.
- De persoonlijke feedback bestaat uit een terugkoppeling op de telsnelheid en een terugkoppeling op de kwaliteit.

- Instructie voor het vinden van de persoonlijke telsnelheid van een telmedewerker op dat moment.
    - Je gaat naar het kopje “scans” in het menu.
    - Je filtert dan op naam, dit doe je door in de cel onder “gescand door” de naam betreffende naam in te vullen.
    - Onderaan de kolom “#” zie je het totaal aantal scans wat er is gemaakt tot dan toe.
    - Na een uur moet dit aantal dus ongeveer 650 zijn.
- Instructie voor het vinden van de persoonlijke kwaliteit van een telmedewerker op dat moment
    - Je blijft in hetzelfde overzicht als hierboven staat. Maar vult dan ook nog in de kolom “Revision” 2 in.
    - Hierdoor zie alle wijzigingen die zijn gemaakt.
    - In de kolom “#” zie je vervolgens hoeveel correcties er zijn gemaakt.
    - Na een uur moet dit niet meer zijn dan 3 of 4.
- De richtlijn tijden voor het geven van de feedback zijn vanaf een uur naar aanvang van scannen tot de eerste pauze. Dit zorgt ervoor dat je ook niet hoeft te berekenen hoeveel per uur er is gescand. Want je weet dat na een uur het aantal scans ongeveer 650 moet zijn.
- Denk er tevens aan dat het moment van feedback goed uitkomt voor de telmedewerker, stoor deze bijvoorbeeld niet wanneer hij/zij grote aantallen aan het tellen is of ergen extra gefocust op is.
- Let op, geef feedback van dichtbij, en complimenten van veraf.
- Denk eraan dat het belangrijk is dat de telmedewerkers gemotiveerd worden, denk er over na hoe je de feedback gaat geven.

Tijdens de pauze

- Verzamel alle telmedewerkers
- Herhaal nogmaals de maatstaven die voor het project genoemd zijn.
- Vertel hoe het ervoor staat met het totale aantal artikelen geteld.
    - Dit doe je op dezelfde manier als voor een persoon, alleen vul je in het overzicht “scans” geen naam in. Hierdoor krijg je de totalen voor de hele telling.
    - Als je na 2 uur pauze houd, kan je het aantal scans dus door 2 delen om zo het aantal scans per uur te weten.
- Vertel hoe het ervoor staat met de gehele kwaliteit van het project.
    - Dit overzicht krijg je op dezelfde manier als hiervoor genoemd. Alleen dan bij de kolom “Revision” 2 in te vullen.
    - Door het aantal correcties te delen door het totaal aan scans weet je het fouten percentage.
- Noem geen persoonlijke prestaties

8.3.3 Official randomization of treatment groups

Naam PL	Afdeling	vrij 9-6	zat 10-6	zon1 11-6	zon2 11-6	ma 12-6	di 13-6
[redacted]	[redacted]	Alkmaar	Den Helder				
				A'dam-Osdorpplein	Haarlem		
		A'dam Noord					
			Leidschendam				
						Hoofddorp	
		Beverwijk	Amstelveen			Purmerend	
							Zaandam
				Hhwaard	Hoorn		
			A'dam Zuid-Oost				
		Arnhem-K	Eindhoven-W	Eindhoven-C		Almelo	
		Deventer					
			Arnhem-C	Doetinchem			
			Helmond				
				Amersfoort	Veenendaal		
				Assen			
				Oosterhout			
		Gorinchem	Ede				
		Tiel	Tilburg			Apeldoorn	Waalwijk
		Den Bosch					
		Enschede		Oss	Uden	Hengelo	
			Nijmegen				
			Groningen				
			Heerenveen				
			Sneek				Roden
		Emmen	Roermond	Weert	Venlo		
			Meppel				
				Leeuwarden	Drachten		
						Hoogeveen	
		Zwolle					Harderwijk
				Hilversum	Almere		
					Lelystad		
		Middelburg		Goes	Terneuzen		
			Roosendaal	Etten-Leur	Bergen op Zoom		
			Heerlen	Maastricht	Sittard		
		Spijkenisse	Hellevoetsluis		Naaldwijk	Dordrecht	
		Rijswijk					
			R'dam-Alex	Gouda	Ridderkerk	R'dam-C	Delft
			R'dam-Zuidplein			Zoetermeer	
				Alphen adR	Leiden		
							Vlaardingen
			Breda	Hoogvliet			
			Den Haag Centrum GV2 (SLplein)	Den Haag GV1 (Wweg)			
		Utrecht K-eiland	Utrecht Centrum HC	Utrecht-Overvecht	Zeist	Nieuwegein	
Extern Magazijn = Wordt niet gebruikt maar is dus gelijke aan control							
Controle groep	29						
Feedback Groep	30						
Bonus betaling	24						
Routes hebben telkens de zelfde treatment (geen 1 route heeft bonus treatment)							
Projectleiders hebben maximaal 2 verschillende treatments							#NAAM?

8.3.4. Actual randomization of treatment groups

Naam PL	Afdeling	vrij 9-6	zat 10-6	zon1 11-6	zon2 11-6	ma 12-6	di 13-6
[redacted]	Alkmaar	Alkmaar	Den Helder				
	Alkmaar			A'dam-Osdorpplein	Haarlem		
	Alkmaar	A'dam Noord					
	Alkmaar		Leidschendam				
	Alkmaar					Hoofddorp	
	Alkmaar	Beverwijk	Amstelveen			Purmerend	
	Alkmaar						Zaandam
	Alkmaar			Hhwaard	Hoorn		
	Alkmaar		A'dam Zuid-Oost				
	Almelo	Arnhem-K	Eindhoven-W	Eindhoven-C		Almelo	
	Almelo	Deventer					
	Almelo		Arnhem-C	Doetinchem			
	Almelo		Helmond				
	Almelo			Amersfoort	Veenendaal		
	Almelo			Assen			
	Almelo			Oosterhout			
	Apeldoorn	Gorinchem	Ede				
	Apeldoorn	Tiel	Tilburg			Apeldoorn	Waalwijk
	Apeldoorn	Den Bosch					
	Enschede	Enschede		Oss	Uden	Hengelo	
	Enschede		Nijmegen				
	Groningen		Groningen				
	Groningen		Heerenveen				
	Groningen		Sneek				Roden
	Groningen	Emmen	Roermond	Weert	Venlo		
	Hoogeveen		Meppel				
	Hoogeveen			Leeuwarden	Drachten		
	Hoogeveen					Hoogeveen	
	Kampen	Zwolle					Harderwijk
	Kampen			Hilversum	Almere		
	Kampen				Lelystad		
	Middelburg	Middelburg		Goes	Terneuzen		
	Middelburg		Roosendaal	Etten-Leur	Bergen op Zoom		
	RAS Eindhoven		Heerlen	Maastricht	Sittard		
	RAS Rotterdam	Spijkenisse	Hellevoetsluis		Naaldwijk	Dordrecht	
	Rotterdam	Rijswijk					
	Rotterdam		R'dam-Alex	Gouda	Ridderkerk	R'dam-C	Delft
	Rotterdam		R'dam-Zuidplein			Zoetermeer	
	Rotterdam			Alphen adR	Leiden		
	Rotterdam						Vlaardingen
	Rotterdam		Breda	Hoogvliet			
	Rotterdam		Den Haag Centrum GV2 (SLplein)	Den Haag GV1 (Wweg)			
	Utrecht	Utrecht K-eiland	Utrecht Centrum HC	Utrecht-Overvecht	Zeist	Nieuwegein	
Voorzien van PL							
Extern Magazijn = Wordt niet gebruikt maar is dus gelijke aan control							
Controle groep	28						
Feedback Groep	24						
Bonus betaling	15						
project mislukt	16						
Routes hebben telkens de zelfde treatment (geen 1 route heeft bonus treatment)							
Projectleiders hebben maximaal 2 verschillende treatments							