

Sliwka, Dirk; Manthei, Kathrin

Conference Paper

Multitasking and the Benefits of Objective Performance Measurement - Evidence from a Field Experiment

Beiträge zur Jahrestagung des Vereins für Socialpolitik 2013: Wettbewerbspolitik und Regulierung in einer globalen Wirtschaftsordnung - Session: Labor Relations in Experiments, No. E05-V3

Provided in Cooperation with:

Verein für Socialpolitik / German Economic Association

Suggested Citation: Sliwka, Dirk; Manthei, Kathrin (2013) : Multitasking and the Benefits of Objective Performance Measurement - Evidence from a Field Experiment, Beiträge zur Jahrestagung des Vereins für Socialpolitik 2013: Wettbewerbspolitik und Regulierung in einer globalen Wirtschaftsordnung - Session: Labor Relations in Experiments, No. E05-V3, ZBW - Deutsche Zentralbibliothek für Wirtschaftswissenschaften, Leibniz-Informationszentrum Wirtschaft, Kiel und Hamburg

This Version is available at:

<https://hdl.handle.net/10419/79968>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Multitasking and the Benefits of Objective Performance Measurement - Evidence from a Field Experiment

Kathrin Manthei*

Dirk Sliwka†

University of Cologne

Preliminary version - Please do not circulate

February 22, 2013

Abstract

We examine the benefits of objective performance measurement in a field experiment conducted in a retail bank. At the outset objective performance measures of profits in each branch were only available on the branch level and managers allocated bonuses to their employees based on subjective assessments. In a subset of the branches, managers then obtained access to individual performance measures. We find a significant positive impact of objective performance measurement on effort and financial performance. This productivity increase is mainly driven by larger branches and higher sales for non-core products which is well in line with a formal economic model on the optimal allocation of monitoring efforts under subjective evaluations in multitask environments.

Key Words: Incentives, Subjective Performance Evaluation, Multitasking, Field Experiment, Bank

JEL Classification: M52, J33, D23

*University of Cologne, Albertus Magnus Platz, 50923 Köln, Germany, tel: +49 221 470-5887, fax: +49 221 470-5078, e-mail: manthei@wiso.uni-koeln.de.

†University of Cologne, Albertus Magnus Platz, 50923 Köln, Germany, tel: +49 221 470-5888, fax: +49 221 470-5078, e-mail: dirk.sliwka@uni-koeln.de.

1 Introduction

A key assumption in typical principal agent settings is that work incentives can be tied to objective and verifiable performance measures. In practice, however, individual employee performance is often assessed subjectively by a superior. Linking pay to an assessment that is at the discretion of a supervisor is often a necessity in jobs where key aspects of a job can only be measured poorly in objective terms (see Prendergast (1999)). However, as has been shown in many studies subjective evaluations are frequently biased (see e.g. Bretz et. al. 1992, Murphy and Cleveland (1995), Prendergast (1999)) and these distortions can reduce the effectiveness of the incentive schemes. Apart from the problem that supervisors may simply not be able to evaluate all aspects of performance time constraints might detain supervisors with a large span of control from monitoring each employee thoroughly. In that case performance assessments will be noisy and incentives can obviously not achieve their full power.

Hence, a straightforward conjecture is that the availability of objective performance measures should reduce these biases and, in turn, lead to higher incentives. However, due to multitasking incentives the use of objective performance measures may also lead to a reallocation of efforts across different tasks.

But although problems of subjective evaluations have been studied more and more by economists in recent years, (Prendergast and Topel (1993), Prendergast and Topel (1996), Gibbs et al. (2003)), the amount of empirical evidence on the incentive consequences is still rather limited (e.g. Engellandt and Riphahn (2011), Bol (2011), and Berger et al. (2013)). Moreover, there is no study we are aware of (i) in which the effects introducing objective measures of performance in an environment where prior to the intervention performance has been assessed subjectively is investigated (ii) this change is imposed exogenously for a randomly selected group allowing for a clean evaluation of causal effects and (iii) this is carried out in a field setting in a firm in an industrialized country.

The key goal of this paper is to study if and how the introduction of objective performance measures can affect the financial success of a company. To do so we make use of a randomized field experiment conducted in a retail bank in Germany. In a randomly selected subgroup of the bank's branches the bank established a set of objective performance measures for a 6-month test period. Prior to the intervention and the remaining branches only the overall branch performance was measured objectively while individual performance was assessed subjectively by the branch managers. This exogenous change allows us to obtain

clean evidence on the impact of objective performance evaluation on financial outcomes. We can track monthly profits for different product categories as well as actively recruited customer appointments as a more direct effort measure.

Our key finding is that the introduction of objective measurement indeed increased profits significantly. But we also find substantially heterogeneous treatment effects which are to some extent well organized by a formal model of subjective evaluations in a multitask environment. In the model a supervisor has to allocate attention on different products and employees and higher anticipated attention leads to higher employee efforts. For instance, we find that the increase in profits is mainly driven by larger branches and through higher sales of products that have a lower share in overall sales volume. With limited attention, larger branches should benefit more from objective performance measures as it is harder for branch managers to track the performance of the employees subjectively. Minor product categories should benefit more as supervisors naturally should focus more attention on a key product when they want to assess the overall profit contribution of each employee as accurately as possible given a limited “budget of attention”. But interestingly, we also find no evidence for overall benefits of objective performance measurement in the smallest branches. In these branches, the availability of objective performance measures for all products apparently led to a too strong shift of sales performance from a core product to more fringe products. In line with Holmström and Milgrom (1991) multitasking problems indeed arise when there is no division of labor – in the small branches all employees typically sell all different products and, hence, the interdependency between the incentives for different tasks.

Our study also contributes to the recent literature using field experiments within firms to evaluate the causal effects of incentive schemes (see Bandiera et al. (2011) for an overview). In a series of field experiments within a company producing soft fruits Bandiera et al. (2005), Bandiera et al. (2007), Bandiera et al. (2009), Bandiera et al. (2010) examine the interplay of work incentives and social relations. They find various forms of evidence that social ties in fact strongly interact with incentives in the workplace. Barankay (2012) studies the effect of rank feedback in a randomized field experiment among furniture salespeople who call sell products from different firms and, hence, are also subject to multitasking incentives. Interestingly, they find that abandoning rank feedback by one firm increases sales of this firm as rank feedback led to a shift of attention from products that generate negative rank feedbacks. Hossain and List (2012) conduct a field experiment at a Chinese high-tech manufacturing factory

studying the impact of different framings of incentives, observing positive effects on productivity for both losses and gains framings. In another study with a large European agricultural producer Englmaier et al. (2012) investigate how the salience of quantity incentives influences performance of harvesting teams. Salience is exogenously varied by giving workers in the treatment group a special briefing on the incentive scheme and visibly placing a note with this information at the workplace. They find that salience indeed has a significant and positive impact on quantity, however, quality is negatively affected.

The paper proceeds as follows. We first present the setting of the field experiment in section 2. An illustrative formal model of subjective evaluations is analyzed in section 3. Section 4 presents the results of the econometric analysis and section 5 concludes.

2 The Field Experiment

The field experiment was conducted by a retail bank with a large set of branches in Germany. We can use monthly data of the bank's branches (>250) of the year 2003 (Jan-Dec). Each branch consisted of a branch manager and a team of employees¹. The job of the branch employees was to serve clients by administering tasks at the counter and to sell the bank's products to customers. The assignment of shifts at the counter was subject to a rotation systems trying to share this responsibility among team members in a fair manner. Potential new customers were brought into the branches in several ways. First, there was a central marketing department initiating sales campaigns (e.g. direct mailings, a company website, promotion campaigns). In addition a central call center belonging to the bank organized sales appointments in the different branches. Furthermore the bank employees themselves were supposed to call current customers to make appointments in the branches. At the time of the experiment the bank sold products in the following key categories: loans, investment products, saving plans with building societies, and credit cards.² Given the strategy of the bank loans to private customers were the most important product category in the portfolio in terms of sales revenues and profits. Branch performance was assessed using a profit measure called "customer net revenue" (CNR in

¹In addition to the branch employees the bank contracts independent sales representatives (mobile sales force). Although they are associated to a specific branch they are self-employed and face therefore different compensation conditions.

²The bank also sold insurances. However, these are not part of this analysis as the available data on insurance sales is incomplete.

the following) basically tracking the profit contribution of each product in each branch. This key indicator was used throughout the bank for evaluating financial performance.

In both the treatment and control group employees could receive a bonus in addition to their fixed salaries which was based on quarterly financial targets for each branch measured in Euro CNR. When the respective target was met a bonus pool was paid out to the branch and it was the branch manager’s responsibility to allocate the bonus to the employees in the branch. Prior to the treatment intervention and in the control group, the branch managers did not have access to objective performance indicators on the sales of individual employees and, hence, the allocation of the bonus was based on their subjective assessment of the employees’ performance. During the intervention, branch managers in the the treatment group had access to the CNR measures for all individual employees in their branch. They could use this information to set individual sales targets and to allocate the bonus pool. Everything else remained unchanged. The bank picked 23 branches that were assigned to the treatment group³ such that they were representative in terms of size, performance, and geographical distribution.⁴ Branches in the treatment group were informed about the intervention within the two month prior to the intervention (i.e. in month 4). In June (i.e. the last month before the intervention) workshops were conducted with the branch managers in the treatment group informing them about the way in which the objective key figures were made available and could be handled.

Year 2003	month 1-6	month 7-12
Treatment (23 branches)	subjective assessment	objective assessment
Control (>250 branches)	subjective assessment	subjective assessment

Table 1: Structure of the experiment

Table 1 shows the structure of the experiment. From January to June 2003

³The reason to conduct the field experiment was a discussion with the firm’s works council about the way in which performance should be assessed. In Germany, firms need the consent of a works council (which is elected by the employees) when important HR policies are changed. Hence, firm and works councils agreed (wisely) to run a “pilot” experiment first to analyze effects of a change in the way performance is assessed.

⁴Due to bank’s wish to stay anonymous we are not able to report descriptive statistics. However, examining the means of our main variables in the pre-test phase with a standard t-test does not yield any significant differences for any of the variables of interest. The analysis includes the total CNR the CNR of each product category, the number of self-initiated appointments, the number of appointments initiated by the call center, or the number of full time equivalent employees.

purely subjective assessments were used in all branches. The intervention ran from July to December 2003 in 23 branches.

3 An Illustrative model

To fix ideas we analyze the following stylized model in which we highlight some aspects of the connection between team size, multitasking incentives, and the benefits of objective performance measurement. There is a group of n risk neutral agents $i = 1, \dots, n$ whose performance is to be evaluated. Each agent works on a set of $j = 1, \dots, J$ tasks and exerts an effort e_{ij} on each task with a cost function $c(e_{i1}, e_{i2}, \dots, e_{iJ}) = \sum_{j=1}^J c_j(e_{ij})$. For each task there is a performance outcome $\pi_{ij} = e_{ij} + a_{ij}$ where $a_{ij} \sim N(m_j, \sigma_a^2)$ are independently distributed error terms. The performance outcomes of all tasks of an agent i generate a profit for the firm which is equal to

$$\pi_i = \sum_{j=1}^J b_j \cdot \pi_{ij}$$

such that b_j describes the importance of task j for the firm. To simplify notation the tasks are ranked according to profitability such that $b_1 > b_2 > \dots > b_J$.

We compare two appraisal regimes, one in which objective performance measures are available and one in which performance is assessed subjectively. In both cases, the agents receive a wage which is linear in the agent's profit contribution. When objective performance measures are available the wage is $w_i = \alpha + \beta \cdot \pi_i$. Under subjective performance evaluation it is $w_i = \alpha + \beta \cdot \tilde{\pi}_i$ where $\tilde{\pi}_i$ is the supervisor's subjective assessments of the agent's profit contribution.

First, suppose that there are no objective performance measures. Then a supervisor S has to evaluate the performance subjectively. The supervisor monitors the agents by collecting signals on their performance π_{ij} for the different tasks. The quality of each signal depends on the time the supervisor spends on monitoring each agent and task. Let t_{ij} be the time spend on the performance of agent i for task j . The supervisor has an overall time budget T that she can allocate on the different tasks and agents such that

$$\sum_{i=1}^n \sum_{j=1}^J t_{ij} = T.$$

By spending time on monitoring a task, the supervisor collects more and more precise information on the true performance outcome π_{ij} for this task. In each unit of time τ the supervisor observes a signal $\eta_{ij\tau} = \pi_{ij} + \varepsilon_{ij\tau}$ where the $\varepsilon_{ij\tau}$ are iid and $\varepsilon_{ij\tau} \sim N(0, \sigma_\varepsilon^2)$. Hence, when investing time t_{ij} the supervisor has observed a vector $\eta_{ij} \in \mathbb{R}^{t_{ij}}$. Note that the mean of the observed signals $s_{ij} = \frac{1}{t_{ij}} \sum_{\tau=1}^{t_{ij}} \eta_{ij\tau}$ is a sufficient statistic for π_{ij} and – given equilibrium efforts e_{ij}^* – this “observed performance” s_{ij} is normally distributed with mean $m_j + e_{ij}^*$ and variance

$$V \left[\frac{1}{t_{ij}} \sum_{\tau=1}^{t_{ij}} \eta_{ij\tau} \right] = \sigma_a^2 + \frac{1}{t_{ij}} \cdot \sigma_\varepsilon^2.$$

For ease of notation we treat the t_{ij} as continuous variables.⁵ The supervisor’s task is to assess the agent’s profit contribution. Following an approach used, for instance, by Prendergast and Topel (1996) or Prendergast (2002) we assume that the supervisor has a preference to report the agent’s profit contribution accurately and her expected utility is⁶

$$-E \left[\left(\tilde{\pi}_i - \pi_i \right)^2 \middle| s_{i1}, s_{i2}, \dots, s_{iJ} \right].$$

Given the normality assumptions, ex-post the supervisor thus optimally reports her own conditional expectation about π_i given the observed signals.

From an ex-ante perspective the supervisor’s decision problem under subjective performance evaluation is how to allocate the time budget T on the different tasks and agents in order to obtain the best estimate of the agents’ profit contribution and minimize expected posterior deviations between reported and actual profit contributions. The agents, in turn, will anticipate this allocation of the monitoring intensity and choose their effort levels in order to maximize their expected payoffs.

We start by analyzing the straightforward benchmark case in which objective performance measures are available. Each agent maximizes her expected utility

$$E \left[\alpha + \beta \cdot \pi_i - \sum_{j=1}^J c_j (e_{ij}) \right]$$

⁵ As the supervisor’s objective function will be strictly concave the optimal discrete choice must be one of the nearest neighbours in the discrete grid.

⁶ If the supervisor is completely selfish, a different interpretation is that the principal can verify the report with a certain probability and then imposes a fine $(r_t - y_t)^2$.

thus choosing an effort level of

$$e_{ij} = c_j'^{-1}(\beta b_j). \quad (1)$$

We now compare these benchmark efforts to those chosen under subjective evaluation where the information structure is less precise and endogenously determined.

Note that the supervisor's decision problem when choosing the optimal report is

$$\min_{\tilde{\pi}_i} E \left[\left(\tilde{\pi}_i - \sum_{j=1}^J b_j \cdot \pi_{ij} \right)^2 \middle| s_{i1}, s_{i2}, \dots, s_{iJ} \right]$$

This boils down to computing the least squares estimator of π_i based on the signals s_i which – as π_i and the signals s_i are normally distributed is identical to reporting the conditional expectations on the agents' profit contributions

$$\begin{aligned} E \left[\sum_{j=1}^J b_j \cdot \pi_{ij} \middle| s_{i1}, s_{i2}, \dots, s_{iJ} \right] &= \sum_{j=1}^J b_j \cdot E[\pi_{ij} | s_{ij}] \\ &= \sum_{j=1}^J b_j \cdot \frac{\sigma_\varepsilon^2 (m_{ij} + e_{ij}^*) + t_{ij} \sigma_a^2 s_{ij}}{t_{ij} \sigma_a^2 + \sigma_\varepsilon^2} \end{aligned}$$

where the latter follows from applying a standard result on the conditional expectation of normally distributed random variables (see, for instance, DeGroot (1970), pp. 169; details given in the Appendix).

We now move to the ex-ante perspective and determine the supervisor's optimal allocation of attention on the tasks and agents. Her ex-ante expected disutility of misreporting is equal to

$$\sum_{i=1}^n E \left[\left(\sum_{j=1}^J b_j \cdot \left(\frac{\sigma_\varepsilon^2 (m_{ij} + e_{ij}^*) + \sigma_a^2 t_{ij} \left(\pi_{ij} + \frac{1}{t_{ij}} \sum_{\tau=1}^{t_i} \varepsilon_{ij\tau} \right)}{\sigma_\varepsilon^2 + t_{ij} \sigma_a^2} - \pi_{ij} \right) \right)^2 \right]$$

which (after some rearrangement – see the Appendix) simplifies to

$$\sum_{i=1}^n \sum_{j=1}^J b_j^2 \frac{\sigma_\varepsilon^2 \sigma_a^2}{\sigma_\varepsilon^2 + t_{ij} \sigma_a^2}.$$

Hence, the supervisor allocates her time budget in order to minimize the ex-

pected standard errors in his posterior estimate of the agents' profit contributions. Note that for each task j of each agent i this is a decreasing and convex function of t_{ij} . Hence, there are decreasing marginal returns to allocated attention to a task – the more time a manager has spent on collecting information about the performance in a task, the less informative are additional signals. The manager thus optimally allocates attention by balancing the marginal returns under the constraint that $\sum_{i=1}^n \sum_{j=1}^J t_{ij} = T$.

First note that the marginal returns to attention must be equal for all tasks that receive positive attention in equilibrium as the supervisor could otherwise shift attention from a task with lower marginal returns to another task with higher returns and reduce the expected measurement error. Moreover, the marginal returns to attention are identical across all agents for a specific j , hence, $t_{ij} = t_j$ for all agents i . The marginal returns to attention for a task j are thus equal to

$$\frac{b_j^2 \sigma_a^4 \sigma_\varepsilon^2}{(\sigma_\varepsilon^2 + t_j \sigma_a^2)^2}$$

and for any tasks j and j' that receive positive attention their marginal returns must be identical which leads to the condition that

$$\frac{b_j}{b_{j'}} = \frac{\sigma_\varepsilon^2 + t_j \sigma_a^2}{\sigma_\varepsilon^2 + t_{j'} \sigma_a^2}. \quad (2)$$

Using this expression we can characterize the optimal allocation of time on the different tasks and obtain the following result.

Proposition 1 *The supervisor allocates his attention on a subset of the $\bar{J}$ most productive tasks, i.e. $t_j > 0$ for $j \leq \bar{J}$. The degree of attention spent on a task j for each agent i is equal to*

$$t_j = \begin{cases} \frac{b_j}{\sum_{j' \leq \bar{J}} b_{j'}} \left(\frac{T}{n} + \bar{J} \cdot \frac{\sigma_\varepsilon^2}{\sigma_a^2} \right) - \frac{\sigma_\varepsilon^2}{\sigma_a^2} & \text{if } j \leq \bar{J} \\ 0 & \text{if } j > \bar{J} \end{cases}. \quad (3)$$

The least productive task which is still monitored $\bar{J}$ is the smallest j for which

$$\frac{\sigma_\varepsilon^2}{\frac{T}{n} \sigma_a^2 + j \cdot \sigma_\varepsilon^2} \sum_{j'=1}^j b_{j'} > b_{j+1}. \quad (4)$$

Hence, the supervisor monitors the most important tasks. The time spend on monitoring a task is a function of the “relative productivity share” of this task

$b_j / \sum_{j' \leq \bar{J}} b'_{j'}$. But the least productive tasks may not receive any attention.⁷ To understand this result recall that the supervisor wants to assess the overall profit contribution as accurately as possible. As more productive tasks contribute more to overall profits the supervisor will invest the most attention in these tasks. If productivity differences are sufficiently large the marginal gains from the last unit of monitoring a productive task may well exceed the marginal gains from starting to monitor a less productive task.

This result has a number of implications regarding the potential benefits of objective performance measurement when we investigate the agents' effort reaction. Given the supervisor's evaluation strategy each agent maximizes

$$\beta \cdot E \left[\sum_{j=1}^J b_j \frac{\sigma_\varepsilon^2 (m_{ij} + e_{ij}^*) + t_{ij} \sigma_a^2 s_{ij}}{t_{ij} \sigma_a^2 + \sigma_\varepsilon^2} \right] - \sum_{j=1}^{\bar{J}} c(e_{ij}).$$

with first order conditions

$$\beta b_j \frac{t_j \sigma_a^2}{t_j \sigma_a^2 + \sigma_\varepsilon^2} = c'(e_{ij}).$$

By inserting the optimal monitoring choices, rearranging terms and comparing optimal efforts under objective measurement (1) we obtain the following result:

Proposition 2 *Under subjective performance evaluation the efforts exerted on each monitored task are characterized by*

$$e_{ij} = c'^{-1} \left(\beta b_j \left(1 - \frac{\sigma_\varepsilon^2}{\frac{b_j}{\sum_{j' \leq \bar{J}} b'_{j'}} \left(\frac{T}{n} \sigma_a^2 + \bar{J} \cdot \sigma_\varepsilon^2 \right)} \right) \right) \text{ for } \forall j \leq \bar{J}.$$

Efforts for all tasks are strictly lower than under objective performance evaluation and are decreasing in team size n . The loss in marginal incentives is the larger the lower the relative profitability $b_j / \sum_{j' \leq \bar{J}} b'_{j'}$ of a task.

First of all, note that subjective performance evaluations lead to less differentiated assessments as compared to objective measurement. The reason is that the higher the uncertainty about the true performance the closer is the optimal estimate of performance to the prior expectations: If a supervisor knows that her

⁷Note that all tasks will receive attention when T is sufficiently large.

assessment is noisy she will rationally attribute a deviation from prior expectations to a larger extent to errors of perception⁸. But this rating compression leads to lower powered incentives as the marginal returns to effort are reduced.

Moreover, the size of this distortion is affected by team size. A supervisor who has to monitor a larger number of agents can spend less time on each agent and thus has less precise information on individual performance. In turn, reported ratings will be less differentiated and thus incentives will be lower-powered.

Furthermore, the distortion will depend upon the importance of the tasks. Ratings will be the more accurate the more important the task.⁹ Efforts are therefore biased towards the most important tasks. Agents do not only work less for the less profitable tasks because of their lower profitability (this is also the case when performance is measured objectively). But they also work less on these tasks because the less profitable tasks are monitored less intensively by supervisors with limited capacities for attention.

4 Results

4.1 Customer Appointments

We now analyze the effects of the treatment intervention on the different available outcome variables. We start with the most direct available key figure for employee effort – the number of self-initiated sales appointments. In each month it is tracked in a data base how many appointments were arranged by the employees in a branch by actively calling up a customer and inviting him to the branch). Calling customers is the most direct way in which an employee can try to raise his financial performance.

⁸This result may be viewed as an economic rationale for a result commonly discussed in by personnel psychologists that there is rating compression in subjective assessments (sometimes called the “centrality bias”). See for instance Murphy and Cleveland (1995), or for a discussion from an economics perspective Prendergast and Topel (1993), Prendergast and Topel (1996), Gibbs et al. (2003) or Kampkötter and Sliwka (2011).

⁹This may be interpreted as an economic rationale for some aspects of the so-called “Halo” effect in subjective performance evaluations according to which in subjective assessments the most important characteristics that are to be assessed “outshine” other less important characteristics.

	<i>Appointments</i>	
	(1)	(2)
Treatment (month 7-12)	15.07* (7.958)	24.80*** (8.931)
Information (month 5-6)		29.19*** (7.659)
Constant	258.6*** (2.935)	258.6*** (2.925)
R-squared	0.323	0.326
Branch fixed effects and month dummies included,		
Robust standard errors clustered on branch level, *** p<0.01, ** p<0.05, * p<0.1		

Table 2: The impact of objective measurement on self-initiated appointments

We can now estimate the causal effect of the intervention by estimating fixed effects models with the treatment intervention as the key independent variable estimating robust standard errors clustered on branch level. There are two points in time after which the treatment group is differently affected by the treatment as compared to the control group. Before month 5 employees in the treatment group were informed that objective performance measures were used starting with month 7. Hence, it is conceivable that already the announcement of individual measurement already affects performance. As customer appointments typically predate actual product sales employees may have an incentive to increase the number of appointments at the earlier date. We estimating the effect of the intervention itself, i.e. the dummy “treatment” takes value 1 in the branches in the treatment group in the months 7-12 in specification (1) and we include a dummy for the month 5 and 6 in which employees in the treatment group were already informed about the new system but it is not yet in place in column (2).

Considering the estimates in column (1) the number of self-initiated sales appointments increase by about 6% relative to month 1 to 6. But as column (2) reveals, the treatment effect already starts with month 5. Already the announcement of the treatment increased the number of monthly appointments by about 11%. When the treatment is in place the effect continues with an increase of roughly 10% relative to the month prior to the announcement. As can be seen from Figure 1 which shows the development of the number of customer

appointments over time, the branches in the treatment group on average made a lower number of appointments prior to the intervention. From month 1 to 4 time trends run nearly exactly in parallel in both treatment and control. But already in month 5 the gap becomes substantially narrower. Hence, indeed the announcement of the treatment affects the incentives to invest in customer contacts that may later on increase financial performance.

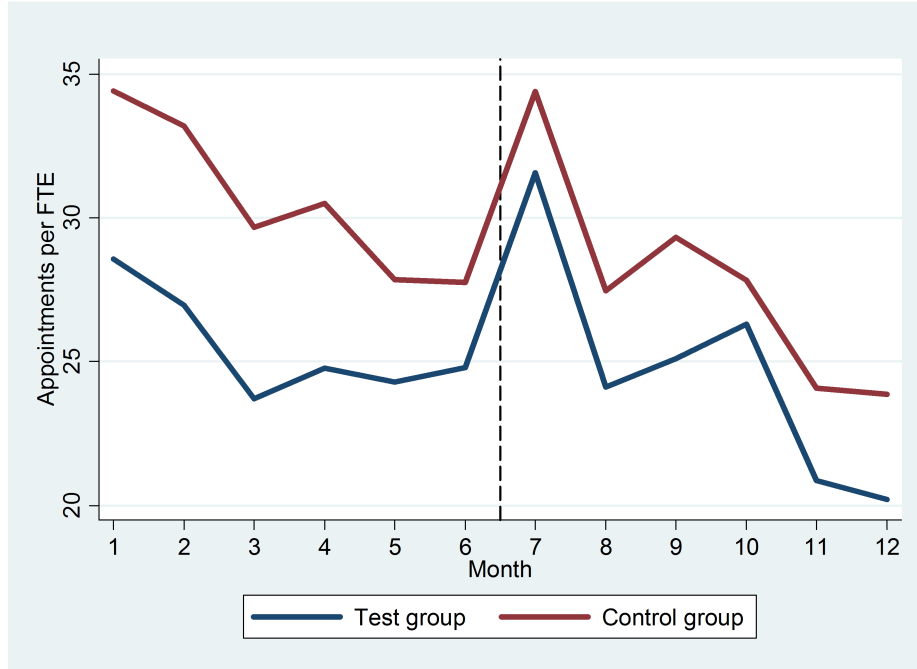


Figure 1: Appointments per FTE over time

4.2 Financial Performance

But how does the treatment affect financial performance? To study this we analyze the profit measure CNR separately by single product categories and aggregated over all categories. Table 3 below shows fixed effects regressions with the profits of the different product categories in columns (1) to (4) and overall profits in column (5). The table reveals a very interesting pattern. As can be seen from column (1) the treatment had no effect on the bank's key product consumer loans. But it has substantial effects for the other products that were of

minor importance before the intervention. For instance, profits from investment products increased by roughly 20%, those from savings accounts with building societies by even 50% and of credit cards (albeit weakly significantly) by 10%. As loans were still the predominant product after the intervention the effect of the treatment on overall profits is at roughly 2%.

	(1)	(2)	(3)	(4)	(5)
	<i>log CNR</i>	<i>log CNR</i>	<i>log CNR</i>	<i>log CNR</i>	<i>log CNR</i>
	<i>loans</i>	<i>investment</i>	<i>savings</i>	<i>credit cards</i>	<i>total</i>
Treatment	0.00652 (0.00906)	0.182*** (0.0419)	0.416** (0.177)	0.0961* (0.0552)	0.0201** (0.00906)
R-squared	0.308	0.382	0.086	0.381	0.410

Branch fixed-effects, month dummies and call center initiated appointments included
Robust standard errors clustered on branches, *** p<0.01, ** p<0.05, * p<0.1.

Table 3: The impact of objective measurement on profits

One interpretation in the light of the model laid out in section 3 is that supervisors kept track of the core product even under subjective evaluation. But as the non-core products had only a weak share in the overall profitability these products were not the key focus of their attention. Objective measurement now provided supervisors precise information about these minor products at no costs. Hence, employees now had an incentive to exert substantially more effort on these product categories.

As we have seen in the above, employees already increased their efforts directly after the announcement of the treatment. A key question is now whether financial performance already increases at that point. But when we include again a dummy for the months after the announcement but before the treatment (see Table A1 in the Appendix) we find no significant effect of the announcement for any of the product categories. The coefficients of the treatment remain virtually unchanged (but the effect on credit cards while increasing somewhat in size is no longer significant). Hence, employees start to prepare for the new system by immediately increasing appointments after the announcement, but product sales go up only after they are measured precisely.

4.3 The Impact of Branch Size

To study further heterogeneous treatment effects we look more closely at the role of branch sizes. For this purpose we interact the treatment variable with the branch size. Table 4 reports the respective regression results. Model (1) just adds an interaction term with the branch size (centered at the mean) to the base line regression. Column (2) instead includes dummies for the 20% smallest (≤ 6 fte) and for the 20% largest (> 10.1 fte) branches. Column (3) analogously considers the 10% smallest and largest branches.¹⁰

	<i>log CNR total</i>		
	(1)	(2)	(3)
Treatment	0.0192** (0.00885)	0.0223* (0.0118)	0.0212** (0.00979)
Treatment x branch size (centered)	0.00466* (0.00269)		
Treatment x 20percentile		-0.0328** (0.0145)	
Treatment x 80percentile		0.0284* (0.0147)	
Treatment x 10percentile			-0.0379** (0.0146)
Treatment x 90percentile			0.0436*** (0.0149)
R-squared	0.411	0.412	0.412
Branch fixed-effects, month dummies and call center initiated appointments included			
Robust standard errors clustered on branches, *** p<0.01, ** p<0.05, * p<0.1.			

Table 4: The role of branch sizes

As the results in column (1) show, the treatment effect depends on the size of the branch. The profit increase from introducing objective performance measures is substantially higher in large branches. An interpretation in the light of the model presented in the above is that the larger the branch, the harder it is for a supervisor to keep track of the performance of the employees. Hence,

¹⁰There is some slight variation over time in the number of fte. For calculating the dummy variables for the 20% smallest and 20% largest branches we use the threshold values from the last month before the announcement of the intervention.

larger branches benefit substantially more from the use of objective performance measurement. Indeed, in the largest 10% of the branches overall profits increase by 6.5%. However, in the smallest branches there seems to be no or even a negative effect.

But a further interesting pattern emerges when we run separate regression for the smallest branches and the rest as reported in Table 5: Even though profits from investment products increased by more than 30% in these branches, here the intervention had a significant negative effect on the sale of loans.

	(1)	(2)	(3)	(4)	(5)
	<i>log CNR</i>	<i>log CNR</i>	<i>log CNR</i>	<i>log CNR</i>	<i>log CNR</i>
	<i>loans</i>	<i>investment</i>	<i>savings</i>	<i>credit cards</i>	<i>total</i>
<i>Small branches (≤ 6 full time employees)</i>					
Treatment	-0.0275**	0.279***	0.0188	0.0302	-0.00511
	(0.0111)	(0.0603)	(0.270)	(0.109)	(0.0130)
<i>Larger branches (> 6 full time employees)</i>					
Treatment	0.0159	0.152***	0.517**	0.112*	0.0262***
	(0.00976)	(0.0498)	(0.206)	(0.0627)	(0.00988)
R-squared	0.174	0.351	0.090	0.309	0.251

Branch fixed-effects, month dummies and call center initiated appointments included

Robust standard errors clustered on branches, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 5: Treatment effects in small and larger branches

Hence, the intervention here caused a shift from the major product loans to the former fringe investment products. This effect is absent in the larger branches.

One difference between large and small branches in the bank is the division of labor. In smaller branches each employee typically sells all products. As already shown by Holmström and Milgrom (1991) multitasking problems – i.e. externalities of an increase in incentives for one task that lead to a reduction in the incentives for other tasks – can be avoided when there is a division of labor. If one agent is responsible for multiple tasks, higher incentives for one task may reduce efforts in another task as agents shift their efforts. If, however, different agents are responsible for the different tasks, this effect does not occur

as a change in the incentive structure for one agent does not affect the behavior of other agents.

5 Conclusion

We studied data from a field experiment in a bank to investigate the benefits of objective performance measurement. As the intervention was exogenously administered we can give clean evidence on the causal effects of introducing objective performance measures in an environment where performance priorly was only assessed subjectively. The introduction of objective measures indeed increased not only employee efforts as measured by the initiated customer appointments but also increased financial success significantly. The effects are driven by higher powered incentives for products that previously have not been in the main focus of supervisors. Moreover, the effects are substantially larger in larger branches where it is harder for supervisors to keep track of the performance of individual employees without objective key figures.

Furthermore the study yields insights on multitasking incentives. In smaller branches where employees are typically responsible for all products and there is no division of labor the use of objective performance measures shifted sales efforts away from the key product to other products which were not the key focus before the intervention. This shifting effect is absent in larger branches – and this adds to the revealed pattern that objective performance measures are more important when team sizes are large but may yield no benefits in smaller teams.

6 Appendix

Conditional Expectation of profit contribution:

$$\begin{aligned}
\tilde{\pi}_{ij} &= E[\pi_{ij}|s_{ij}] = E\left[\pi_{ij} + \frac{1}{t_{ij}} \sum_{\tau=1}^{t_i} \varepsilon_{ij\tau}\right] \\
&= (m_{ij} + e_{ij}^*) + \frac{Cov\left[\pi_{ij}, \pi_{ij} + \frac{1}{t_{ij}} \sum_{\tau=1}^{t_i} \varepsilon_{ij\tau}\right]}{V\left[\pi_{ij} + \frac{1}{t_{ij}} \sum_{\tau=1}^{t_i} \varepsilon_{ij\tau}\right]} (s_{ij} - (m_{ij} + e_{ij}^*)) \\
&= \frac{\sigma_\varepsilon^2 (m_{ij} + e_{ij}^*) + t_{ij} \sigma_a^2 s_{ij}}{t_{ij} \sigma_a^2 + \sigma_\varepsilon^2}
\end{aligned}$$

Ex-ante expected disutility of misreporting:

$$\sum_{i=1}^n E \left[\left(\sum_{j=1}^J b_j \cdot \left(\frac{\sigma_\varepsilon^2 (m_{ij} + e_{ij}^*) + \sigma_a^2 t_{ij} \left(\pi_{ij} + \frac{1}{t_{ij}} \sum_{\tau=1}^{t_i} \varepsilon_{ij\tau} \right)}{\sigma_\varepsilon^2 + t_{ij} \sigma_a^2} - \pi_{ij} \right) \right)^2 \right]$$

Using $E[X^2] = V[X] + (E[X])^2$ this is equivalent to

$$\begin{aligned} \sum_{i=1}^n & \left[V \left[\sum_{j=1}^J b_j \cdot \left(\frac{\sigma_\varepsilon^2 (m_{ij} + e_{ij}^*) + \sigma_a^2 t_{ij} \left(\pi_{ij} + \frac{1}{t_{ij}} \sum_{\tau=1}^{t_i} \varepsilon_{ij\tau} \right)}{\sigma_\varepsilon^2 + t_{ij} \sigma_a^2} - \pi_{ij} \right) \right] + \right. \\ & \left. + E \left[\sum_{j=1}^J b_j \cdot \left(\frac{\sigma_\varepsilon^2 (m_{ij} + e_{ij}^*) + \sigma_a^2 t_{ij} \left(\pi_{ij} + \frac{1}{t_{ij}} \sum_{\tau=1}^{t_i} \varepsilon_{ij\tau} \right)}{\sigma_\varepsilon^2 + t_{ij} \sigma_a^2} - \pi_{ij} \right) \right]^2 \right]. \end{aligned}$$

But the expected value of the squared deviations is equal to zero as

$$\begin{aligned} & E \left[\sum_{j=1}^J b_j \cdot \left(\frac{\sigma_\varepsilon^2 (m_{ij} + e_{ij}^*) + \sigma_a^2 t_{ij} \left(\pi_{ij} + \frac{1}{t_{ij}} \sum_{\tau=1}^{t_i} \varepsilon_{ij\tau} \right)}{\sigma_\varepsilon^2 + t_{ij} \sigma_a^2} - \pi_{ij} \right) \right] \\ &= \sum_{j=1}^J b_j \cdot \left(\frac{\sigma_\varepsilon^2 + \sigma_a^2 t_{ij}}{\sigma_\varepsilon^2 + t_{ij} \sigma_a^2} (m_{ij} + e_{ij}^*) - (m_{ij} + e_{ij}^*) \right) \\ &= 0 \end{aligned}$$

such that the disutility of misreporting is equal to :

$$\begin{aligned} & \sum_{i=1}^n V \left[\sum_{j=1}^J b_j \cdot \left(\frac{\sigma_\varepsilon^2 (m_{ij} + e_{ij}^*) + \sigma_a^2 t_{ij} \left(\pi_{ij} + \frac{1}{t_{ij}} \sum_{\tau=1}^{t_i} \varepsilon_{ij\tau} \right)}{\sigma_\varepsilon^2 + t_{ij} \sigma_a^2} - \pi_{ij} \right) \right] \\ &= \sum_{i=1}^n \sum_{j=1}^J b_j^2 V \left[\left(\frac{\sigma_a^2 t_{ij} - \sigma_\varepsilon^2 - t_{ij} \sigma_a^2}{\sigma_\varepsilon^2 + t_{ij} \sigma_a^2} \right) \pi_{ij} + \frac{\sigma_a^2}{\sigma_\varepsilon^2 + t_{ij} \sigma_a^2} \left(\sum_{\tau=1}^{t_i} \varepsilon_{ij\tau} \right) \right] \\ &= \sum_{i=1}^n \sum_{j=1}^J b_j^2 \left(\frac{\sigma_\varepsilon^4 \sigma_a^2}{(\sigma_\varepsilon^2 + t_{ij} \sigma_a^2)^2} + \frac{t_{ij} \sigma_\varepsilon^2 \sigma_a^4}{(\sigma_\varepsilon^2 + t_{ij} \sigma_a^2)^2} \right) \\ &= \sum_{i=1}^n \sum_{j=1}^J b_j^2 \frac{\sigma_\varepsilon^2 \sigma_a^2}{\sigma_\varepsilon^2 + t_{ij} \sigma_a^2} \end{aligned}$$

Proof of Proposition 1:

First, note that (??) is equivalent to

$$t_{j'} = \frac{b_{j'}}{b_j} \left(\frac{\sigma_\varepsilon^2}{\sigma_a^2} + t_j \right) - \frac{\sigma_\varepsilon^2}{\sigma_a^2}.$$

Summing up this expression across the first $\bar{J}$ tasks we obtain that the total time spend must be equal to

$$\begin{aligned} T &= n \sum_{j' \leq \bar{J}} \left(\frac{b_{j'}}{b_j} \left(\frac{\sigma_\varepsilon^2}{\sigma_a^2} + t_j \right) - \frac{\sigma_\varepsilon^2}{\sigma_a^2} \right) \\ &\Leftrightarrow \left(\frac{\sigma_\varepsilon^2}{\sigma_a^2} + t_j \right) \frac{\sum_{j' \leq \bar{J}} b_{j'}}{b_j} - \bar{J} \cdot \frac{\sigma_\varepsilon^2}{\sigma_a^2} = \frac{T}{n}. \end{aligned}$$

Solving for t_j yields that for each task j for which $t_j > 0$ we must have

$$t_j = \frac{b_j}{\sum_{j' \leq \bar{J}} b_{j'}} \left(\frac{T}{n} + \bar{J} \cdot \frac{\sigma_\varepsilon^2}{\sigma_a^2} \right) - \frac{\sigma_\varepsilon^2}{\sigma_a^2}. \quad (5)$$

Now, note that the marginal returns to attention for a task j at $t_j = 0$ is equal to $b_j^2 \frac{\sigma_a^4}{\sigma_\varepsilon^2}$. As the objective function is strictly convex in each t_j and because $b_{j+1} < b_j$ it must be the case that if $t_j = 0$ then $t_{j+1} = 0$. Suppose that $\bar{J}$ is the last task that is actively monitored (i.e. $t_j > 0$ but $t_{j+1} = 0$). The marginal return the last unit of monitoring task j is then

$$\frac{b_j^2 \sigma_a^4 \sigma_\varepsilon^2}{\left(\sigma_\varepsilon^2 + \left(\frac{b_j}{\sum_{j'=1}^j b_{j'}} \left(\frac{T}{n} + j \cdot \frac{\sigma_\varepsilon^2}{\sigma_a^2} \right) - \frac{\sigma_\varepsilon^2}{\sigma_a^2} \right) \sigma_a^2 \right)^2} = \frac{\left(\sum_{j'=1}^j b_{j'} \right)^2 \sigma_a^4 \sigma_\varepsilon^2}{\left(\frac{T}{n} \sigma_a^2 + j \cdot \sigma_\varepsilon^2 \right)^2}$$

Hence, it is not worthwhile spend time on monitoring task $j+1$ if

$$\frac{\left(\sum_{j'=1}^j b_{j'} \right)^2 \sigma_a^4 \sigma_\varepsilon^2}{\left(\frac{T}{n} \sigma_a^2 + j \cdot \sigma_\varepsilon^2 \right)} > b_{j+1}^2 \frac{\sigma_a^4}{\sigma_\varepsilon^2}$$

which is equivalent to condition (4). ■

	(1)	(2)	(3)	(4)	(5)
	<i>log CNR</i>	<i>log CNR</i>	<i>log CNR</i>	<i>log CNR</i>	<i>log CNR</i>
	<i>loans</i>	<i>investment</i>	<i>savings</i>	<i>credit cards</i>	<i>total</i>
Information	0.00850 (0.00717)	0.0374 (0.0571)	-0.00725 (0.193)	0.0210 (0.0622)	0.0124 (0.00779)
Treatment	0.00936 (0.0109)	0.194*** (0.0482)	0.414** (0.178)	0.103 (0.0660)	0.0242** (0.0109)
R-squared	0.309	0.382	0.086	0.381	0.410

Branch fixed-effects, month dummies and call center initiated appointments included

Robust standard errors clustered on branches, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A1: The impact of treatment information on profits

References

- Bandiera, O., I. Barankay, and I. Rasul (2005). Social preferences and the response to incentives: Evidence from personnel data. *The Quarterly Journal of Economics* 120(3), 917–962.
- Bandiera, O., I. Barankay, and I. Rasul (2007). Incentives for managers and inequality among workers: Evidence from a firm-level experiment. *The Quarterly Journal of Economics* 122, 729–773.
- Bandiera, O., I. Barankay, and I. Rasul (2009). Social connections and incentives in the workplace: Evidence from personnel data. *Econometrica* 77(4), 1047–1094.
- Bandiera, O., I. Barankay, and I. Rasul (2010). Social incentives in the workplace. *The Review of Economic Studies* 77(2), 417–458.
- Bandiera, O., I. Barankay, and I. Rasul (2011). Field experiments with firms. *CEPR Discussion Paper No. DP8412*.
- Barankay, I. (2012). Rank incentives. evidence from a randomized workplace experiment. *mimeo Wharton School of Business, University of Pennsylvania*.
- Berger, J., C. Harbring, and D. Sliwka (2013). Performance appraisals and the impact of forced distribution: An experimental investigation. *Management Science*.

- Bol, J. C. (2011). The determinants and performance effects of managers' performance evaluation biases. *Accounting Review* Vol. 86, 1549–1575.
- DeGroot, M. (1970). *Optimal Statistical Decisions*. New York: McGraw-Hill.
- Engellandt, A. and R. T. Riphahn (2011). Evidence on incentive effects of subjective performance evaluations. *Industrial and Labor Relations Review* 64, 241–257.
- Englmaier, F., A. Roider, and U. Sunde (2012). The role of salience in performance schemes: Evidence from a field experiment. *IZA Discussion Paper*.
- Gibbs, M., K. A. Merchant, W. A. van der Stede, and M. E. Vargus (2003). Determinants and effects of subjectivity in incentives. *The Accounting Review* 79, 409–436.
- Holmström, B. and P. Milgrom (1991). Multitask principal-agent analyses: Incentive contracts, asset ownership and job design. *Journal of Law, Economics and Organization* 7, 24–52.
- Hossain, T. and J. List (2012). The behavioralist visits the factory: Increasing productivity using simple framing manipulations. *Management Science*, forthcoming.
- Kampkötter, P. and D. Sliwka (2011). Differentiation and performance - an empirical investigation on the incentive effects of bonus plans. *mimeo*.
- Murphy, K. R. and J. N. Cleveland (1995). *Understanding Performance Appraisal*. Thousand Oaks: Sage.
- Prendergast, C. and R. Topel (1996). Favoritism in organizations. *Journal of Political Economy* 104, 958–978.
- Prendergast, C. J. (1999). The provision of incentives in firms. *Journal of Economic Literature* 37, 7–63.
- Prendergast, C. J. (2002). Uncertainty and incentives. *Journal of Labor Economics* 20, 115–37.
- Prendergast, C. J. and R. H. Topel (1993). Discretion and bias in performance evaluation. *European Economic Review* 37, 355–65.