ETLA
**ELINKEINOELÄMÄN TUTKIMUSLAITOS**
THE RESEARCH INSTITUTE OF THE FINNISH ECONOMY
Lönnrotinkatu 4 B 00120 Helsinki Finland Tel. 358-9-609 900
Telefax 358-9-601 753 World Wide Web: http://www.etla.fi/

# Keskusteluaiheita – Discussion papers

No. 1216

Antti Kauhanen* – Sami Napari*

## PERFORMANCE MEASUREMENT AND INCENTIVE PLANS

* The Research Institute of the Finnish Economy, Lönnrotinkatu 4 B 00120 Helsinki, Finland. E-mail: antti.kauhanen@etla.fi, sami.napari@etla.fi. Funding by TT-säätiö is gratefully acknowledged. Comments by Petri Böckerman, Panu Kalmi and Terhi Maczulskij are appreciated.

ISSN 0781-6847 05.05.2010

**KAUHANEN,** Antti – **NAPARI**, Sami, **PERFORMANCE MEASUREMENT AND INCENTIVE PLANS.** Helsinki: ETLA, Elinkeinoelämän Tutkimuslaitos, The Research Institute of the Finnish Economy, 2010, 28 p. (Keskusteluaiheita, Discussion Papers, ISSN 0781-6847; No. 1216).

**ABSTRACT:** This paper explores performance measurement in incentive plans. Based on theory, we argue that differences in the nature of jobs between blue- and white-collar employees lead to differences in incentive systems. We find that performance measurement for white-collar workers is broader in terms of the performance measures, the organizational level of performance measurement and the time horizon. The intensity of incentives is also stronger for white-collar employees. All of these findings are consistent with theory.

**KAUHANEN,** Antti – **NAPARI**, Sami, **PERFORMANCE MEASUREMENT AND INCENTIVE PLANS.** Helsinki: ETLA, Elinkeinoelämän Tutkimuslaitos, The Research Institute of the Finnish Economy, 2010, 28 s. (Keskusteluaiheita, Discussion Papers, ISSN 0781-6847; No. 1216).

**TIIVISTELMÄ:** Artikkelissa tarkastellaan suorituksen arviointia tulospalkkausjärjestelmissä. Teoriaan nojautuen esitämme, että erot työnkuvassa toimihenkilöiden ja työntekijöiden välillä johtavat erilaisiin kannustinjärjestelmiin. Tulokset osoittavat, että toimihenkilöillä suoritusmittarit ovat laajempia, suoritusta mitataan useammalla organisaatiotasolla ja pidemmällä aikavälillä. Kannustimet ovat toimihenkilöillä myös voimakkaampia kuin työntekijöillä. Esitetyt tulokset ovat teorian mukaisia.

**Keywords:** incentive pay, performance measurement, risk versus distortion trade-off, agency theory.

**JEL codes:** J33, M52, M54

# *1. Introduction*

The success of incentive plans depends on the performance measures that are used. Performance measurement is often difficult because it is hard to reliably measure an employee's contribution to the objective of the firm. A good performance measure would capture the contribution of an agent to the objective of the firm and nothing else. In practice, such measures are hard to find. Typically, the broader measures that focus on many aspects of performance measures tend to reflect not only additional factors that the employee can control but also factors to which the employee cannot respond. Thus, the problem with broad performance measures is that they impose risk on the employee. Narrower measures are less risky for the employee, but they tend to miss some aspects of performance that the employee can control and may lead to distorted incentives.

A particular problem with narrow measures arises from employees having specific knowledge (in the sense of Jensen and Meckling 1992) because, in such cases, narrow measures do not provide incentives for the employee to utilize her information (Raith 2008). This trade-off between risk and distortion affects the performance measures used and the weights put to different measures in incentive plans (Baker 2002). In general, firms want to combine both broad and narrow measures to reach a better balance between risk and distortion (Baker 2002, Raith 2008). The trade-off between risk and distortion depends on the nature of the job, particularly on the relevance of the specific knowledge and breadth of tasks and decision rights.

Despite the large theoretical literature on performance measurement in incentive plans, empirical research on the topic is scarce. It has been argued in the recent literature that our current knowledge of the features of incentive plans is scant, almost non-existent (e.g. Ittner and Larcker 2002 pp. S59, Gibbs et al. 2009 pp. 238). For example, Ittner and Larcker (2002) argue that the prior studies have typically considered aggregate performance measure classifications such as

financial vs. non-financial classifications instead of specific measures such as profitability and productivity, among others, which may lead to incorrect inferences concerning the choice of performance measures in incentive plans. Moreover, most studies have concentrated on incentive contracts for top management, while the contracts for other employees have received little attention (Ittner and Larcker 2002 is an exception). However, it is likely that incentive plans are different for executives and other employees, as the performance measures that provide the optimal trade-off between risk and distortion vary substantially depending on the nature of the job (see, for example, Baker 2002). Therefore, incentive plans for different employee groups may use very different measures and combinations of measures.

We utilize a unique survey covering a large part of the Finnish manufacturing and service sector to study incentive plans. We make two contributions to the literature. First, we provide a descriptive account of the key features of incentive systems. We describe what kind of performance measures are used, how these measures are combined, the organizational level of performance measurement, how often the rewards are paid and their share of regular earnings. Second, we explore the implications of the recent theoretical work on performance measurement. Based on prior research (Hopp et al. 2009), we argue that blue- and white-collar employees differ in terms of the relevance of specific knowledge and the breadth of tasks and decision rights, leading to differences in incentive systems. We develop a set of hypotheses based on recent theory and test them with these data. We find that performance measurement for white-collar workers is broader in terms of the performance measures used, the organizational level of performance measurement and the time horizon in comparison to blue-collar workers. The intensity of incentives is also stronger for white-collar employees. All of these findings are consistent with theory.

The rest of the paper is organized as follows. The next section introduces the relevant theory and develops hypotheses. This is followed by a description of the pertinent Finnish labor market institutions. In the main body of the paper, we describe the survey used and present our findings. The final section concludes.

## *2. Theory*

By building on the theoretical literature on properties of performance measures, this section discusses what kind of performance measures should be used in incentive systems and why these measures are likely to differ between employee groups. The review of the theoretical literature is then summarized to four hypotheses to be tested in the empirical part of the paper.

### Properties of performance measures

A central property of a performance measure is its risk profile. Performance measures often combine two kinds of risk: 1) uncontrollable risk and 2) controllable risk.[i] Uncontrollable risk is pure noise; it reflects all environmental factors affecting the performance measure that the employee is unable to control. Controllable risk refers to environmental uncertainty affecting the performance measure that the employee can control or react to, such as, for example, knowledge about the correct action to take given the state of the world. The key feature of controllable risk or specific knowledge is that it is costly to communicate to others in the organization (Jensen and Meckling 1992). Therefore, the employee herself has to act based on this knowledge, or this piece of knowledge will not used in decision making at all, due to costly communication.

These two types of risk generally have the opposite implications for incentive plans. Performance measures that have a lot of uncontrollable risk are costly to use, as a risk-averse agent has to be compensated for carrying the risk (Holmstrom 1979). Therefore, a higher level of uncontrollable

risk in the performance measure implies a lower level of optimal incentives. However, given that the employee has knowledge that is valuable to the firm, output-based incentives can be used to motivate the employee to utilize this knowledge (Raith 2008). Therefore, optimal incentives are stronger when controllable risk is important.

Another important property of performance measures is distortion. Distortion refers to situations where an agent's actions have different impacts on the performance measure and the goal of the organization (Holmstrom and Milgrom 1991, Baker 1992, Feltham and Xie 1994, Baker 2002). Holmström and Milgrom (1991) focus on situations where agent has multiple tasks that can be measured with varying accuracy. They show that firms avoid strong incentives to discourage sub-optimal effort allocation between tasks. Baker (1992) shows that if the principal cannot pay for what she really cares about, incentives are reduced to avoid rewarding the wrong behavior. Feltham & Xie (1994) show that if the agent carries out more tasks than can be measured, the optimal effort levels will be inefficient.

An ideal performance measure would impose little risk on the employee (who are typically risk averse) and would precisely measure the employees' contribution to the objective of the organization. In this situation, the employee could be given strong incentives based on this measure, which would lead to high effort in tasks that are in the interests of the organization with little risk for the employee. However, it is usually the case that the undistorted measures contain a lot of uncontrollable risk and vice versa, which lead to a trade-off between risk and distortion (see Baker 2002 for examples and discussion). Usually, broader measures that focus on many aspects of performance are less distorted but are more risky for the employee. In effect, broader measures tend to include both more factors that employee can control and more factors that she cannot control. For example, in publicly traded firms, the market value of the firm would be an undistorted measure, as it captures the impact of possible actions of the employee but imposes considerable

risk on the employee, as many factors outside the employee's control affect the market value as well. Narrower measures are less risky, but they may also be more distorted, as they tend to miss some aspects of performance. For example, using divisional profit as a performance measure partly removes uncontrollable risk but at the same time might reduce employees' incentives to support and help other divisions, resulting in negative effects on the value of the firm.

The trade-off between distortion and risk as well as the importance of specific knowledge depend on job characteristics. Consider a job that is broad in terms of tasks and decision rights. Narrow performance measures lead to distortions, as they may miss some tasks or do not take into account all possible consequences of the employees' decisions. Broader measures possibly fare better in terms of the trade-off between risk and distortion, as they lead to lower distortions while not increasing risk too much, due to the employee's broad decision rights. However, in a narrow job, narrow performance measures may not lead to distorted incentives, as they may capture all aspects of performance that the employee can control. Using broader measures would tend to increase risk but not diminish distortions. This match of job design and performance measurement is discussed in greater detail in the next section, where we explicitly explore the differences in incentive systems between white-collar and blue-collar employees.

So far, the discussion has dealt with single performance measures. However, incentive plans usually use multiple measures. The informativeness principle states that the incentive plan should include any performance measure that is marginally informative of the employees actions if it can be costlessly incorporated to the contract (Holmstrom 1979). The trade-off between risk and distortion provides another reason for including multiple measures. Gibbs et al. (2009) argue that multiple performance measures are used to mitigate weaknesses in each other. The combination of performance measures should thus produce a better result in terms of risk and distortion than any of the measures used alone. Baker (2002) shows that even if the firm could reward for an

undistorted measure, it may also want to use a distorted measure if it helps in reducing uncontrollable risk. Raith (2008) comes to a similar conclusion but with different reasoning. In his model, to encourage the employee to respond to controllable risk, she is rewarded for her output but is partly shielded from uncontrollable risk by also being rewarded for input-based measures.[ii] An employee having higher levels of specific knowledge and productivity should be compensated more by output and less by input, according to Raith (2008). Hwang et al. (2009) extend Raith's (2008) model to include personal tasks and group tasks, showing that the specific knowledge and the value of knowledge sharing (which here means the pooling of specific information held by multiple agents) lead to greater emphasis on output measures. Essentially, this result follows from the value of using specific information in both individual and groups tasks (for more detailed discussion, see Hwang et al. 2009 pp. 1152-1153 ).

One common way of improving the trade-off between distortion and risk is using subjectivity in performance measurement (Gibbs et al. 2004). Subjective evaluation can decrease distortion if it allows taking into account dimensions of an employee's job that the quantitative measures cannot. In this way, subjective evaluation is helpful in jobs with multiple tasks or where the employer cannot measure what she really cares about.

## Incentive systems for blue- and white-collar employees

As discussed above, the choice of performance measures depends on the trade-off between distortion and risk and the importance of controllable risk (specific knowledge) and knowledge sharing. The trade-off between distortion and risk in turn depends on the characteristics of the job, especially on the breadth of tasks and decision rights. Next, we will argue that blue-collar and white-collar jobs differ notably in terms of the importance of controllable risk (or specific knowledge) and knowledge sharing and in the breadth of tasks and decision rights.

Hopp et al. (2009) develop a framework to analyze the main differences between blue-collar and white-collar tasks. They argue that the key differences arise from the presence or absence of an intellectual or knowledge component in a task and whether the task is mainly routine or involves creativity. White-collar tasks are knowledge intensive and involve creativity in the sense that the employee has to find novel and appropriate solutions to problems.[iii] Moreover, they emphasize the importance of knowledge sharing for white-collar workers. For blue-collar tasks, such a knowledge-based component is usually missing, and the tasks involve the application of known methods to familiar problems. Therefore, blue-collar work is more standardized and routine, which leaves less room for specific knowledge. This fact does not mean that blue-collar workers cannot possess specific knowledge; rather, specific knowledge plays a larger role for white-collar workers. Fama (1991) also points out that the relationship between inputs such as hourly effort and skill can differ between workers and tasks in white-collar jobs, whereas in blue-collar jobs, the principal has a better idea of how inputs map to outputs. Again, if the principal has less information on the relationship between inputs and outputs than the employees, then the employees have specific knowledge.

Because white-collar work is intellectually more complex and procedures are not easily pre-specified, white-collar employees are typically given more discretion in their jobs (Hopp et al. 2009), for example, with respect to the selection and scheduling of tasks as well as output quality.

Blue- and white-collar jobs also differ in terms of what can be observed or measured. Fama (1991) argues that in blue-collar jobs, either their hourly output can be monitored or their skill and effort can be monitored and the principal knows how these inputs translate to outputs. For the white-collar employees, the monitoring of hourly output or skill and effort is not as easy, as the outputs are typically intangible and can only be measured over longer horizons (Fama 1991, Davenport and Prusak 2002).

To summarize, controllable risk (specific knowledge) and knowledge sharing are more important for white-collar employees. Additionally, the jobs of white-collar employees are likely to be broader in terms of tasks and decision rights. Next, we discuss how these differences in the nature of jobs lead to differences in performance measurement in incentive plans.

## Hypotheses

According to Raith's (2008) model, more complex jobs, in terms of specific knowledge and difficulties faced in specifying the optimal allocation of effort between tasks, should place more emphasis on broader performance measures (e.g., output-based instead of input-based incentives) and stronger incentives. Hwang et al. (2009) add to this model a group task and show that when knowledge sharing is more important, broader measures should be used. Moreover, because the white-collar jobs are usually broader in terms of tasks and decision rights, performance measurement systems that include only narrow performance measures would lead to larger distortions for white-collar workers than for blue-collar employees (See Lazear and Gibbs 2009 pp. 245). This breadth of tasks also implies that white-collar workers should be evaluated on multiple measures to ensure that all aspects of the job are captured by the performance measures. In general, narrow and broad measures should be combined either to balance the trade-off between distortion and risk (Baker 2002) or to balance the trade-off between risk and incentives to utilize specific knowledge (Raith 2008). However, using multiple measures in complex tasks is not without problems. Quantitative measures are unlikely to capture all of the relevant dimensions of white-collar jobs, and finding optimal weights for multiple measures is likely to be difficult (Holmstrom and Milgrom 1991). Therefore, subjective evaluation may be warranted (Gibbs et al. 2004). We summarize this discussion in the following hypothesis:

*Hypothesis 1: Because white-collar employees have broader jobs in terms of tasks and decision rights and because specific knowledge and knowledge sharing play a larger role for them, their performance measurement should be broader. More specifically, their performance measurement should i) emphasize outputs more than inputs, ii) not be based solely on inputs, iii) be based on multiple measures, i.e., combinations of broad and narrow measures and iv) possibly include subjective evaluation.*

Because white-collar employees typically perform multiple tasks and because they have broader decision rights, white-collar employees are often able to affect outcomes at several organizational levels. Therefore, performance measures that fail to account for this fact and focus on evaluating performance only at a narrow level (say, in one particular department) are likely to lead to increased distortion. For blue-collar workers, using performance measures covering only one or few organizational levels will not necessarily lead to large distortions, given the narrower nature of their jobs and limited discretion, which leads us to the following hypothesis:

*Hypothesis 2: The performance of white-collar employees should be measured at a larger number of organizational levels than the performance of blue-collar employees.*

In white-collar jobs, outputs typically contain knowledge components. Because of this intangible nature of outputs, the performance of white-collar employees can often been measured only long after the task has been completed (Hopp et al. 2009). In blue-collar work, because tasks are more routine and less knowledge based and also because outputs are often physical, it is easier to measure hourly output or skill and effort over short periods of time. In fact, Fama (1991) argues that this is one of the reasons why blue-collar workers are paid by the hour, whereas salary payoffs are more common for white-collar employees. In terms of incentive plans, this fact suggests that blue collar workers can be rewarded more often than white-collar employees. Thus:

*Hypothesis 3: The incentive payments of white-collar employees should be paid less frequently than the payments of blue-collar employees.*

The intensity of incentives is primarily determined by the value of employee effort and the properties of available performance measures. As discussed, for example, in Rosen (1982) and Lazear and Gibbs (2009, pp. 272), the value of employee effort tends to be higher at higher levels in the hierarchy. Given that the intensity of incentives increases with the value of effort, this finding would suggest stronger incentives for white-collar employees than for blue-collar employees. Also according to Raith (2008), greater importance placed on specific knowledge, leads to the need for stronger (output-based) incentives, which also implies stronger incentives for white-collar employees. We thus conclude the following:

*Hypothesis 4: White-collar employees should have stronger incentives, i.e., the proportion of earnings that is tied to employee performance should be higher for white-collar employees than for blue-collar workers.*

A common theme in all of these hypotheses is that the performance measures for white-collar employees should be broader in terms of measures used, the organizational level of performance measurement and the frequency of measurement.

The above discussion focuses on comparing white-collar and blue-collar employees. However, in the empirical analysis that follows, we further separate the white-collar workers into distinct groups: clerical and upper white-collar workers. Because upper white-collar employees work at higher levels in the organizational hierarchy than clerical employees, the differences between upper white-collar employees and blue-collar workers in terms of controllable risk and the breadth of tasks and decision rights are likely to be even larger than the corresponding differences between clerical and blue-collar workers. It thus follows that any support to our

stated hypotheses should also be stronger for the upper white-collar employees than for the clerical workers, which is why we think it is useful to analyze these employee groups separately.

## *3. Institutional setting*

In this section, we describe the relevant features of Finnish labor markets. One central aspect is the definition of the blue-collar employee and the white-collar employee. As mentioned above, we divide employees into three groups. The first one, blue-collar *workers*, consists of hourly waged non-managerial, non-supervisory employees who are mostly low-skilled, that is, without post-secondary education. The second group, *clerical employees*, consists of salaried lower white-collar employees. The third group, *upper white-collar employees* includes supervisors, managers and other upper white-collar employees. These definitions are often used in the context of the Finnish labor market and collective bargaining.

In Finland, both employees and employers are widely organized, and collective bargaining coverage is around 95% (Asplund 2007). Collective bargaining takes place at the sectoral level, and guidelines for these negotiations usually come from negotiations between the central organizations of employee unions and employer federations. Recently, there has been a movement toward local bargaining, but the range of issues that can be bargained locally has been limited, and the sectoral agreements set the minimum standards in any case. The collective agreements are typically generally binding meaning that the agreements are extended to cover non-signatory parties as well.

The base pay mechanisms differ between blue-collar and white-collar employees. The hourly wages of the blue-collar employees are fixed hourly wages, piece-rate or reward-rate, or some combination of these.[iv] The minimum hourly wage rates for given job-complexity levels are

defined in the collective agreement, and these agreements also provide guidelines for setting the piece-rates or stipulate the rates. The white-collar employees receive a monthly salary. For clerical employees, minimum wages for different jobs are stipulated in the collective agreements, while the wage setting of upper white-collar employees is based on bargaining at the personal level. These differences in base pay methods can be understood in the light of Fama's (1991) arguments presented earlier.

In the Finnish labor market, performance-related pay (PRP) is typically defined as a wage supplement that augments base pay mechanisms and is based on meeting financial or operative targets. In practice, this definition covers all usual performance-related pay components except for those stipulated in the collective agreements for blue-collar employees, e.g., piece-rates.

Performance-related pay has become more widespread since the early 1990s, while in the 1980s, it was virtually non-existent apart from piece- and reward-rates in blue-collar manufacturing jobs. There is little reliable data from the early 1990s, but in 1998, 39% of blue-collar workers in Finnish manufacturing were in PRP schemes, while the share had risen to 66% in 2004. For the white-collar employees, the respective figures were 57% and 85%. (EK 2006)

## *4. Data*

The survey used here was carried out in 2008 and covers all members of the Confederation of Finnish Industries (EK), which is the central organization of employer associations. There are over 16,000 member firms, most of which are small and medium-sized enterprises. The member firms represent over 70% of the Finnish GDP and over 90% of exports. The sample consisted of 3,180 firms, and 1,738 of these responded, giving a response rate of 55%, which is quite good for

this kind of survey[1]. The sampled firms represent around 720,000 employees, while the responding firms employ roughly 530,000 persons. Therefore the respondents represent 73% of employees in the sample. The survey was sent to 25% of randomly chosen firms with 10-99 employees and to all firms employing over 100 persons. More details of the sample and respondents are given in the Appendix.

The survey deals mainly with i) the prevalence and characteristics of PRP systems ii) the prevalence of various base-pay systems and iii) views on the wage increases stipulated by the collective agreements. It also includes some background variables at the firm level. Each question concerning the PRP plans asks separate answers for blue-collars, clerical employees and white-collars. The top management, e.g., CEOs and equivalents, were not included in the survey. In this paper, we focus on the PRP systems.

Taking the theoretical predictions to the data involves three challenges. First, much of the theory focuses on relative weights put on different performance measures. In our data, we do not observe how strongly different measures are rewarded; we only observe whether a measure is used as part of the incentive system. Therefore, we have to assume that, following e.g., Hwang et al. (2009), the same considerations that would lead to more weight put on a performance measure will also lead to a greater likelihood that a measure will be used.

Second, much of the theoretical discussion focuses on controllable risk or specific knowledge, which we cannot directly measure in our data. However, as the previous discussion shows, on average,

---

[1] For example, in the European Company Survey, which considers inter alia payment systems, carried out by the European Foundation for the Improvement of Living and Working Conditions had a response rate of 42 % in its management interview module (Eurofound 2010). The survey used by Ittner and Larcker (2002) had a response rate of roughly 34%.

white-collar employees are more likely to have specific knowledge than blue-collar employees. This does not mean that blue-collar employees would never have specific knowledge but rather that white-collar employees on average are more likely to possess it. Similar notions apply to knowledge sharing. Hwang et al. (2009) use the association between investments in new production technologies as a proxy for specific knowledge. This proxy is justified on the grounds that such investments are linked to employee skills and capabilities. We feel that our proxy is better grounded in theory.

Third, classifying performance measures as distorted/undistorted or outputs/inputs or broad /narrow in an absolute sense is not easy. However, the task is easier in a relative sense. In this paper, we take profits to be a relatively undistorted measure. Because we concentrate on incentive plans that are not based on stock prices or related measures (the reason for this approach is explained below), profitability is arguably the broadest measure among performance measures, excluding stock-based pay. We will thus refer to profits as a broad (undistorted/ output) measure and all other measures as narrow (distorted/ input).[v]

All of the analyses are carried out using the number of employees and the inverse of the sampling probability as weights. Therefore, the figures show, for example, not how many firms use PRP but rather how many employees are in PRP plans.

Table 1 shows summary statistics of the prevalence of the PRP and stock-based compensation systems among the EK member firms. From panel A, we see that the PRP systems are generally fairly commonly used, although clerical and upper white-collar employees are clearly more often participating in the PRP systems than blue-collar workers. When it comes to stock-based compensation, it is mostly upper white-collar employees who are participating in these plans; even among them, stock-based compensation is quite rare. In this respect, our focus on the PRP systems is well justified. Panel B shows the shares of persons participating in a compensation

system given that a system is used in the personnel group. Interestingly, if the PRP system is used, it seems to be somewhat more selective higher in the organization. This pattern also holds for stock-based compensation, although one should be careful in drawing too strong conclusions in this respect because of the small number of observations.

**Table 1 Summary statistics of the prevalence of Performance-Related Pay**

**Table 1**

*Panel A. Prevalence of Compensation Plans*

| **Performance-related pay** | Obs | Mean |
|---|---|---|
| Workers | 1730 | 0.44 |
| Clerical | 1730 | 0.69 |
| Upper white-collar | 1730 | 0.75 |
| **Stock based compensation, options, etc.** | | |
| Workers | 1729 | 0.01 |
| Clerical | 1729 | 0.05 |
| Upper white-collar | 1729 | 0.26 |

*Panel B. Share of Persons Participating if a Plan is Used in the Personnel Group*

| **Performance-related pay** | Obs | Mean |
|---|---|---|
| Workers | 473 | 0.90 |
| Clerical | 768 | 0.85 |
| Upper white-collar | 837 | 0.84 |
| **Stock based compensation, options, etc.** | Obs | Mean |
| Workers | 14 | 0.58 |
| Clerical | 37 | 0.22 |
| Upper white-collar | 180 | 0.20 |

Notes. Firms included in the analysis are those that have reported i) the number of employees in each employee group, and ii) have reported the number of employees participating in different compensation plans or iii) reported that a compensation plan in question is not used for any personnel group. The figures are weighted by employment and sampling probability.

## *5. Results*

We now turn to the empirical examination of the hypotheses developed in Section 2. As discussed above, by comparing the incentive systems between blue-collar and white-collar employees, we aim to provide new information on the implications of the recent theoretical work on performance measurement. The analysis that follows also adds to our current knowledge of the main features of the incentive plans used in practice.

Table 2 shows data on different performance measures used. At least four observations can be made. First, from the last row, we see that the average number of measures increases with organizational level. For blue-collar workers, there are on average four measures in place, whereas for clerical and upper white-collar employees, there are 4.6 and 5 measures, respectively. Second, the prevalence of the profitability measure, which can be considered a relatively broad measure that captures many aspects of performance measures and seen as an "output", is clearly highest among upper white-collar employees followed by clerical employees. Third, the quality of service or products, which is an operational metric measuring an input of an employee (see Raith 2008 pp. 1063), is a more important measure for blue-collar workers than for upper white-collar employees. However, it should be noted that the importance of this measure does not steadily decrease with organizational level, as the quality of service or products is as important of a measure for clerical employees as it is for blue-collar workers. Finally, development goals are much more important for clerical and upper white-collar employees than for blue-collar workers. These goals may refer to goals agreed upon in development talks but may also refer to improvement in quantitative measures. Therefore, interpreting this result as evidence of subjective evaluation may not be warranted. However, it does include a subjective component, so we take this result as offering some support for the claim that subjective evaluation is more common for white-collar workers.

Overall, Table 2 supports hypothesis 1. Performance measurement in white-collar incentive plans is broader in the sense that profitability is more commonly used compared to blue-collar workers; additionally, the number of measures used is higher for white-collar employees. Furthermore, the use of development goals as a performance measure was found to be more common for white-collar workers. If this finding is interpreted as evidence of subjective evaluation, the result would support hypothesis 1.

**Table 2 Performance Measures**

| | Workers | | Clerical | | Upper white-collar | |
|---|---|---|---|---|---|---|
| | Obs | Mean | Obs | Mean | Obs | Mean |
| Cost savings | 494 | **0.41** | 724 | **0.45** | 744 | **0.45** |
| Quality of service or products | 494 | **0.61** | 724 | **0.62** | 744 | **0.55** |
| Productivity | 494 | **0.54** | 724 | **0.54** | 744 | **0.55** |
| Lead time | 494 | 0.29 | 724 | 0.26 | 744 | 0.23 |
| Inventory quantity or value | 494 | 0.16 | 724 | 0.17 | 744 | 0.19 |
| Development goal | 494 | 0.33 | 724 | **0.53** | 744 | **0.58** |
| Sales target | 494 | 0.15 | 724 | 0.34 | 744 | 0.39 |
| Profitability | 494 | **0.75** | 724 | **0.81** | 744 | **0.92** |
| Enhanced utilization of capital | 494 | 0.08 | 724 | 0.12 | 744 | 0.19 |
| Turnover | 494 | 0.19 | 724 | 0.27 | 744 | 0.36 |
| Market share | 494 | 0.08 | 724 | 0.13 | 744 | 0.19 |
| Other | 494 | 0.37 | 724 | 0.38 | 744 | 0.36 |
| Number of measures on average | | 4.0 | | 4.6 | | 5.0 |

Note: The figures are weighted by employment and sampling probability.

Table 3 also examines hypotheses 1 and 2 but focuses on combinations of performance measures. As argued above, theory shows that generally narrow and broad measures should be combined either to balance the trade-off between distortion and risk (Baker 2002), or to balance the trade-off between risk and incentives to utilize specific knowledge (Raith 2008). Following this logic, Table 3 shows information on the measures used in cases where i) only one measure is applied, ii) multiple narrow measures are used and iii) narrow measures are combined with a broad measure. It is apparent from the table that by far the most common combination is multiple measures that also include a broad measure. However, the prevalence of multiple narrow measures combined with a broad measure is clearly more pronounced for clerical and upper white-collar employees. Second, if only one measure is applied, for white-collar employees, the profitability measure tends to be used, whereas for blue-collar workers, measuring the quality of service or products is also quite common. However, if multiple narrow measures without a broad measure are used, quality of service or products and productivity are common for all three employee groups. For white-collar employees, development goals are also important in this case. From the third column, we see that when multiple narrow measures are combined with a broad measure, then productivity, quality of service or products and cost savings tend to go along with profitability. Additionally, here development goals play an important role for clerical and upper white-collar employees.

**Table 3 Performance measures by incentive scope**

| | *Panel A. Workers* | | | | | |
|---|---|---|---|---|---|---|
| | Only one measure | | Multiple measures, all narrow | | Multiple measures, including a broad measure | |
| | Obs | Mean | Obs | Mean | Obs | Mean |
| Cost savings | 114 | 0.09 | 96 | 0.32 | 284 | **0.52** |
| Quality of service or products | 114 | 0.31 | 96 | **0.77** | 284 | **0.66** |
| Productivity | 114 | 0.09 | 96 | **0.68** | 284 | **0.64** |
| Lead time | 114 | 0.01 | 96 | **0.47** | 284 | 0.34 |
| Inventory quantity or value | 114 | 0.00 | 96 | 0.13 | 284 | 0.21 |
| Development goal | 114 | 0.02 | 96 | 0.28 | 284 | 0.43 |
| Sales target | 114 | 0.02 | 96 | 0.20 | 284 | 0.17 |
| Profitability | 114 | **0.45** | 96 | 0.00 | 284 | **1.00** |
| Enhanced utilization of capital | 114 | 0.00 | 96 | 0.02 | 284 | 0.11 |
| Turnover | 114 | 0.00 | 96 | 0.14 | 284 | 0.26 |
| Market share | 114 | 0.00 | 96 | 0.05 | 284 | 0.10 |
| Other | 114 | 0.02 | 96 | 0.34 | 284 | **0.47** |
| % of total | 23 % | | 19 % | | 57 % | |
| Number of measures on average | | 1.00 | | 3.41 | | 4.91 |
| | *Panel B. Clerical* | | | | | |
| | Only one measure | | Multiple measures, all narrow | | Multiple measures, including a broad measure | |
| | Obs | Mean | Obs | Mean | Obs | Mean |
| Cost savings | 138 | 0.06 | 80 | 0.58 | 506 | **0.47** |
| Quality of service or products | 138 | 0.01 | 80 | **0.83** | 506 | **0.67** |
| Productivity | 138 | 0.01 | 80 | **0.67** | 506 | **0.58** |
| Lead time | 138 | 0.00 | 80 | 0.51 | 506 | 0.24 |
| Inventory quantity or value | 138 | 0.00 | 80 | 0.01 | 506 | 0.23 |
| Development goal | 138 | 0.04 | 80 | **0.65** | 506 | **0.58** |
| Sales target | 138 | 0.07 | 80 | **0.61** | 506 | 0.32 |
| Profitability | 138 | **0.72** | 80 | 0.00 | 506 | **1.00** |
| Enhanced utilization of capital | 138 | 0.00 | 80 | 0.02 | 506 | 0.16 |
| Turnover | 138 | 0.05 | 80 | 0.07 | 506 | 0.34 |
| Market share | 138 | 0.00 | 80 | 0.04 | 506 | 0.16 |
| Other | 138 | 0.04 | 80 | 0.30 | 506 | 0.45 |
| % of total | 19 % | | 11 % | | 70 % | |
| Number of measures on average | | 1.00 | | 4.29 | | 5.20 |
| | *Panel C. Upper White-Collar* | | | | | |
| | Only one measure | | Multiple measures, all narrow | | Multiple measures, including a broad measure | |
| | Obs | Mean | Obs | Mean | Obs | Mean |
| Cost savings | 137 | 0.00 | 54 | 0.25 | 553 | **0.52** |
| Quality of service or products | 137 | 0.01 | 54 | **0.56** | 553 | **0.62** |
| Productivity | 137 | 0.01 | 54 | **0.55** | 553 | **0.61** |
| Lead time | 137 | 0.00 | 54 | 0.35 | 553 | 0.25 |
| Inventory quantity or value | 137 | 0.00 | 54 | 0.00 | 553 | 0.22 |
| Development goal | 137 | 0.06 | 54 | **0.62** | 553 | **0.64** |
| Sales target | 137 | 0.06 | 54 | 0.06 | 553 | 0.45 |
| Profitability | 137 | **0.80** | 54 | 0.00 | 553 | **1.00** |
| Enhanced utilization of capital | 137 | 0.00 | 54 | 0.06 | 553 | 0.23 |
| Turnover | 137 | 0.01 | 54 | 0.10 | 553 | 0.42 |
| Market share | 137 | 0.00 | 54 | 0.03 | 553 | 0.22 |
| Other | 137 | 0.04 | 54 | 0.73 | 553 | 0.37 |
| % of total | 18 % | | 7 % | | 74 % | |
| Number of measures on average | | 1.00 | | 3.31 | | 5.53 |

Note: The figures are weighted by employment and sampling probability.

The results in Table 3 are thus consistent with hypothesis 1. White-collar employees are more often than blue-collar workers in plans that combine both broad and narrow measures, but incentive plans that use only narrow measures are clearly more common for blue-collar workers.

Table 4 examines the breadth of performance measures in a regression framework, thus going beyond the results in Table 3. Here we control for firm characteristics such as industry (23 categories) and firm size (5 categories) and use dummies for family-owned firms, export-intensive firms, growth firms, subsidiaries of foreign firms, firms with the most employees abroad and multi-plant firms. Additionally, after many firm characteristics have been accounted for, white-collar employees are less likely to be compensated solely on inputs (columns 1 and 3) and are more often compensated for outputs (column 2). The last column shows that multiple measures are clearly more common with clerical and upper white-collar workers than with blue-collar workers. These results support hypothesis 1.

**Table 4 Breadth of Performance measurement**

| VARIABLES | **Narrow only** | **Broad only** | **Combination of narrow measures** | **Combination of narrow and broad measures** |
|---|---|---|---|---|
| Clerical | -0.116*** | 0.0360** | -0.0376 | 0.117** |
| | (0.042) | (0.015) | (0.039) | (0.054) |
| Upper white-collar | -0.0906*** | 0.0486** | -0.139*** | 0.181*** |
| | (0.031) | (0.024) | (0.029) | (0.043) |
| Observations | 1945 | 1945 | 1945 | 1945 |
| R-squared | 0.303 | 0.110 | 0.110 | 0.177 |

*** p<0.01, ** p<0.05, * p<0.1 An observation here is firm-employee group. The standard errors are robust to clustering of employee groups within firms. The estimation uses employment and sampling probability weights.

Table 5 considers the organizational level of performance measurement and thus hypothesis 2. The results show that the number of measurement levels increases with the organizational level. For upper white-collar employees, there are on average 3.5 different measurement levels in place, whereas the performance of blue-collar workers is measured using an average of 2.5 different

**Table 5 Level of Performance Measurement**

*Panel A. Workers*

| | Only one measure, narrow | | Only one measure, broad | | Multiple measures, all narrow | | Multiple measures, including broad measure | | Total | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Obs | Mean | Obs | Mean | Obs | Mean | Obs | Mean | Obs | Mean |
| Own work | 39 | **0.82** | 69 | 0.19 | 89 | 0.37 | 281 | 0.28 | 478 | 0.33 |
| Team | 39 | **0.41** | 69 | 0.04 | 89 | **0.47** | 281 | **0.48** | 478 | 0.41 |
| Establishment | 39 | 0.22 | 69 | 0.02 | 89 | 0.25 | 281 | 0.28 | 478 | 0.23 |
| Department | 39 | 0.04 | 69 | 0.03 | 89 | **0.42** | 281 | 0.33 | 478 | 0.28 |
| Project | 39 | 0.00 | 69 | 0.00 | 89 | 0.11 | 281 | 0.08 | 478 | 0.07 |
| Profit Center | 39 | 0.07 | 69 | 0.17 | 89 | 0.30 | 281 | **0.53** | 478 | 0.40 |
| Company | 39 | 0.07 | 69 | **0.57** | 89 | 0.21 | 281 | **0.63** | 478 | 0.50 |
| Group (of companies) | 39 | 0.00 | 69 | 0.34 | 89 | 0.06 | 281 | 0.43 | 478 | 0.31 |
| % of total | 8 % | | 14 % | | 19 % | | 59 % | | | |
| Number of levels on average | | 1.63 | | 1.35 | | 2.20 | | 3.04 | | 2.52 |

*Panel B. Clerical*

| | Only one measure, narrow | | Only one measure, broad | | Multiple measures, all narrow | | Multiple measures, including broad measure | | Total | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Obs | Mean | Obs | Mean | Obs | Mean | Obs | Mean | Obs | Mean |
| Own work | 35 | 0.40 | 94 | 0.21 | 76 | **0.71** | 501 | **0.52** | 706 | 0.49 |
| Team | 35 | 0.27 | 94 | 0.04 | 76 | **0.69** | 501 | **0.56** | 706 | 0.49 |
| Establishment | 35 | 0.01 | 94 | 0.01 | 76 | 0.05 | 501 | 0.24 | 706 | 0.18 |
| Department | 35 | **0.41** | 94 | 0.05 | 76 | 0.29 | 501 | 0.45 | 706 | 0.38 |
| Project | 35 | 0.02 | 94 | 0.04 | 76 | 0.20 | 501 | 0.33 | 706 | 0.26 |
| Profit Center | 35 | 0.29 | 94 | 0.21 | 76 | **0.53** | 501 | **0.71** | 706 | 0.60 |
| Company | 35 | **0.47** | 94 | **0.58** | 76 | 0.24 | 501 | **0.65** | 706 | 0.59 |
| Group (of companies) | 35 | 0.03 | 94 | 0.35 | 76 | 0.04 | 501 | 0.40 | 706 | 0.34 |
| % of total | 5 % | | 13 % | | 11 % | | 71 % | | | |
| Number of levels on average | | 1.90 | | 1.49 | | 2.73 | | 3.87 | | 3.33 |

*Panel A. Upper White-Collar*

| | Only one measure, narrow | | Only one measure, broad | | Multiple measures, all narrow | | Multiple measures, including broad measure | | Total | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Obs | Mean | Obs | Mean | Obs | Mean | Obs | Mean | Obs | Mean |
| Own work | 34 | **0.39** | 99 | 0.30 | 51 | 0.20 | 547 | **0.68** | 731 | 0.58 |
| Team | 34 | 0.23 | 99 | 0.10 | 51 | 0.39 | 547 | 0.43 | 731 | 0.37 |
| Establishment | 34 | 0.02 | 99 | 0.01 | 51 | 0.21 | 547 | 0.19 | 731 | 0.16 |
| Department | 34 | 0.08 | 99 | 0.06 | 51 | **0.44** | 547 | 0.38 | 731 | 0.33 |
| Project | 34 | 0.08 | 99 | 0.00 | 51 | 0.34 | 547 | 0.37 | 731 | 0.30 |
| Profit Center | 34 | 0.21 | 99 | 0.22 | 51 | **0.62** | 547 | **0.73** | 731 | 0.63 |
| Company | 34 | **0.63** | 99 | **0.61** | 51 | **0.56** | 547 | **0.79** | 731 | 0.74 |
| Group (of companies) | 34 | 0.15 | 99 | **0.36** | 51 | 0.16 | 547 | **0.47** | 731 | 0.42 |
| % of total | 5 % | | 14 % | | 7 % | | 75 % | | | |
| Number of levels on average | | 1.81 | | 1.66 | | 2.91 | | 4.04 | | 3.53 |

Note: The figures are weighted by employment and sampling probability.

levels. Furthermore, this tendency for the number of measurement levels being larger higher in the organizational hierarchy does not depend on the type (narrow vs. broad) or on the number

(one vs. multiple) of performance measures in place. Table 5 also suggests that the importance of the group, the company or the profit center as a level of measurement tends to increase with the organizational level, although the differences between employee groups in this respect are fairly small, at least between blue-collar workers and clerical white-collar employees.

Table 6 examines the frequency of performance-related payment by explaining the probability of receiving PRP payments more often than once a year with the same set of control variables than Table 4. The regression results support hypothesis 3, which states that incentive payments for clerical and upper white-collar employees are paid less frequently than the payments for blue-collar workers. The results for the other variables show that narrow measures pay more often than the wider ones.

**Table 6 Frequency of payment, OLS**

| VARIABLES | **More often than once a year** |
|---|---|
| Clerical | -0.158*** |
| | (0.050) |
| Upper white-collar | -0.231*** |
| | (0.061) |
| Broad only | -0.386*** |
| | (0.120) |
| Combination of narrow measures | 0.0127 |
| | (0.127) |
| Broad and narrow combination | -0.182 |
| | (0.117) |
| Observations | 1911 |
| R-squared | 0.250 |

*** p<0.01, ** p<0.05, * p<0.1 An observation here is firm-employee group. The standard errors are robust to clustering of employee groups within firms

Table 7 turns to the intensity of the incentives. In line with hypothesis 4, the incentives are stronger for white-collar employees. For upper white-collar workers, actual PRP payments account, on average, for 8.6% of regular earnings. For clerical and blue-collar workers, the

corresponding numbers are 5.4 and 3.8, respectively. Additionally, if we look at the maximum PRP payments that firms have reported, we notice that they increase in the organizational level. However, there are no notable differences in the shares of actual payments of the maximum: for all employee groups, the actual incentive payments are around 45% of the maximum. The share of firms that have caps on PRP payments can be calculated by dividing the number of observations to the question on maximum PRP payments by the number of observations to the question on realized PRP payments, which assumes that firms that did not answer to the question on maximum payments in fact do not have a cap on payments. Calculated this way, we see that roughly 85% of firms have restrictions on the maximum amount of PRP payments. Murphy (2000) also found that in roughly 80% of executive incentive contracts the bonuses are capped.

**Table 7 Incentive Intensity I**

| **Actual PRP payments as share of regular earnings** | Mean | P10 | Median | P90 | Obs |
|---|---|---|---|---|---|
| Workers | 3.81 | 0.3 | 4 | 7 | 367 |
| Clerical | 5.44 | 0.5 | 5 | 10 | 610 |
| Upper white-collar | 8.66 | 1.2 | 8 | 15 | 639 |
| | | | | | |
| **Maximum PRP payments as share of regular earnings** | Mean | P10 | Median | P90 | Obs |
| Workers | 8.50 | 2.5 | 8 | 15 | 318 |
| Clerical | 11.42 | 5 | 10 | 20 | 513 |
| Upper white-collar | 18.22 | 6 | 16 | 33 | 535 |

Notes. In the upper panel only firms that have a positive number of employees in PRP plans in the given employee group are included. In the second panel, firms that have reported a positive maximum amount are used.

Table 8 continues examining incentive intensity. The first two columns show that also after various firm characteristics have been accounted for, the main conclusions made from Table 7 hold: incentives are stronger at the upper end of the organization. Furthermore, as can be seen from columns three and four, this result does not depend on the breadth of the incentive plans. However, there are some differences in the incentive intensity between plans that differ in terms of their scope; incentive plans that include only a broad measure have paid more in PRP payments, whereas the maximum is highest for a combination of narrow measures.

**Table 8 Incentive Intensity II, OLS**

| VARIABLES | **Actual PRP payments as share of regular earnings** | **Maximum PRP payments as share of regular earnings** | **Actual PRP payments as share of regular earnings** | **Maximum PRP payments as share of regular earnings** |
|---|---|---|---|---|
| Clerical | 1.523*** | 2.443** | 1.292*** | 1.995* |
| | (0.356) | (1.050) | (0.369) | (1.022) |
| Upper white-collar | 5.318*** | 11.35*** | 5.379*** | 11.79*** |
| | (0.773) | (1.679) | (0.727) | (1.534) |
| Broad only | | | 2.804** | 4.622 |
| | | | (1.277) | (2.920) |
| Combination of narrow measures | | | 1.094 | 5.060* |
| | | | (1.031) | (2.614) |
| Broad and narrow combination | | | 1.540* | 3.917* |
| | | | (0.921) | (2.130) |
| Observations | 1616 | 1350 | 1568 | 1324 |
| R-squared | 0.297 | 0.255 | 0.322 | 0.262 |

*** p<0.01, ** p<0.05, * p<0.1 An observation here is firm-employee group. The standard errors are robust to clustering of employee groups within firms. The estimation uses employment and sampling probability weights.

We are not aware of any earlier studies examining performance measurement in incentive plans by comparing differences in performance measures between employee groups. Therefore, it is difficult to evaluate how our results exactly compare to those of the previous studies on the topic. However, it can be said that our results are generally in line with the earlier findings of the literature on performance evaluation in incentive plans. For example, our finding that PRP payments account for a higher share of total earnings for white-collar employees than for blue-collar workers supports the results by O'Shaughnessy (1998), who concluded that the relationship between the bonus to base ratio and performance is stronger higher in the organizational hierarchy. Furthermore, the results of Murphy (1999, 2000) show that in executive incentive contracts, typically i) the performance measures include a measure of profitability, ii) multiple measures are used and iii) bonuses are capped. Ittner and Larcker (2002) find that accounting- and quality-based performance measures are used in about half of the firms in their sample, while around 40% use cost measures and 25% use productivity. These findings are quite similar to ours, although productivity is more commonly used in our sample as a performance measure.

Finally, Gibbs et al. (2009) find that firms use several interrelated measures to mitigate weaknesses in individual measures. The authors also show that noise, risk, distortion and manipulability are all properties relevant to incentive plan design. Our results support these findings. We also observe that firms typically use multiple measures instead of relying on one particular measure. Furthermore, we report a significant variation in performance measurement between white-collar and blue-collar workers because white-collar and blue-collar jobs differ in some important respects. Because of these differences in job characteristics, the relevance of those measurement properties analyzed by Gibbs et al. (2009) is also likely to differ between white-collar and blue-collar employees.

## *6. Conclusion*

We investigated performance-related pay plans and especially performance measurement using a large survey. Despite significant theoretical interest in incentive plans and performance measurement, empirical research has been scarce. Most existing studies have concentrated on incentive contracts for top management, while the contracts for other employees have received little attention. Moreover, the prior studies typically have not considered specific performance measures but instead have considered aggregated performance measure classes. Recent theory suggests that optimal incentive plans depend heavily on job design, especially in terms of the breadth of tasks and decision rights and the importance of specific knowledge. We empirically explored the implications of this recent theoretical work on performance measurement. Based on prior research (Hopp et al. 2009), we argued that blue- and white-collar employees differ in terms of the relevance of specific knowledge and the breadth of tasks and decision rights, leading to differences in incentive systems. We found that performance measurement in incentive systems for white-collar is broader in terms of performance measures, the organizational level of performance measurement and the time horizon. The intensity of incentives is also stronger for

white-collar employees. All of these findings are consistent with the theory. Our results suggest that controllable risk (specific knowledge) and the trade-off between distortion and risk play a significant role in shaping incentive plans, as shown by theoretical work.

In addition to offering support to recent theories on performance measurement in incentive plans, we present evidence of what incentive systems look like. The most common performance measures focus on profitability, productivity, quality and, in the case of white-collar employees, development goals. Incentive plans typically use multiple performances, and the most common combination is to have profitability combined with other, narrower, measures. Performance is typically measured at multiple levels, ranging from the personal level to the company level. Incentive pay is usually paid once a year, and the incentive intensity is low, on average in the range of 4 to 8 per cent of regular earnings. Most of the plans also have a cap on the rewards. Many of these findings are in line with earlier studies, even though the studies deal with very different populations.

# *Appendix*

**Table 9 Sample and respondents by firm size**

| Firm size | Sampling rate | Sample | | Respondents | | Respondents share | |
|---|---|---|---|---|---|---|---|
| (Employees) | | Firms | Employees | Firms | Employees | Firms | Employees |
| 10 – 49 | 25 % | 1 488 | 31 709 | 718 | 15 774 | 48 % | 50 % |
| 50 – 99 | 25 % | 302 | 20 619 | 164 | 11 120 | 54 % | 54 % |
| 100 – 249 | 100 % | 777 | 119 176 | 442 | 67 635 | 57 % | 57 % |
| 250 – 999 | 100 % | 501 | 228 212 | 332 | 155 172 | 66 % | 68 % |
| 1000 - | 100 % | 112 | 322 921 | 83 | 279 716 | 74 % | 87 % |
| Total: | | 3 180 | 722 637 | 1 738 | 529 417 | 55 % | 73 % |

## *References*

Asplund, Rita. 2007. Finland: Decentralisation Tendencies within a Collective Wage Bargaining System. Discussion Papers No. 1077. Helsinki: The Research Institute of the Finnish Economy.

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

Baker, George P. 1992. 'Incentive Contracts and Performance Measurement.' Journal of Political Economy 100(3): 598-614.

Davenport, Thomas H. and Laurence Prusak. 2002. Working Knowledge: How Organizations Manage What They Know. Boston, MA: Harvard Business School Press.

Eurofound. 2010. European Company Survey 2009: Overview. Luxembourg: Office for Official Publications of the European Communities.

Fama, Eugene F. 1991. 'Time, Salary, and Incentive Payoffs in Labor Contracts.' Journal of Labor Economics 9(1): 25-44.

Feltham, Gerald A. and Jim Xie. 1994. 'Performance Measure Congruity and Diversity in Multi-Task Principal/Agent Relations.' Accounting Review 69(3): 429-453.

Gibbs, Michael J., Kenneth A. Merchant, Wim A. Van Der Stede and Mark E. Vargus. 2004. 'Determinants and Effects of Subjectivity in Incentives.' Accounting Review 79(2): 409-436.

Gibbs, Michael J., Kenneth A. Merchant, Wim A. Van Der Stede and Mark E. Vargus. 2009. 'Performance Measure Properties and Incentive System Design.' Industrial Relations 48(2): 237-264.

Holmstrom, Bengt. 1979. 'Moral Hazard and Observability.' Bell Journal of Economics 10(1): 74-91.

Holmstrom, Bengt and Paul Milgrom. 1991. 'Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design.' Journal of Law, Economics, and Organization 7: 24-52.

Hopp, Wallace J., Seyed M.R. Iravani and Fang Liu. 2009. 'Managing White-Collar Work: An Operations-Oriented Survey.' Production and Operations Management 18(1): 1-32.

Hwang, Yuhchang, David H. Erkens and John H. Evans. 2009. 'Knowledge Sharing and Incentive Design in Production Environments: Theory and Evidence.' Accounting Review 84(4): 1145-1170.

Ittner, Christopher D. and David F. Larcker. 2002. 'Determinants of Performance Measure Choices in Worker Incentive Plans.' Journal of Labor Economics 20(2): S58-S90.

Jensen, Michael C. and William H. Meckling. 1992. 'Specific and General Knowledge, and Organizational Structure.' In Contract Economics, edited by L. Werin and H. Wijkander, pp. 251-274: Cambridge, Mass. and Oxford: Blackwell.

Lazear, Edvard P. and Michael Gibbs. 2009. Personnel Economics in Practice. Hoboken, NJ: Wiley.

Murphy, Kevin J. 1999. 'Executive Compensation.' In Handbook of Labor Economics, edited by O. C. Ashenfelter and D. Card, pp. 2485-2563: Elsevier.

Murphy, Kevin J. 2000. 'Performance Standards in Incentive Contracts.' Journal of Accounting and Economics 30(3): 245-278.

O'Shaughnessy, K. C. 1998. 'The Structure of White-Collar Compensation and Organizational Performance.' Relations Industrielles / Industrial Relations 53(3): 458.

Raith, Michael. 2008. 'Specific Knowledge and Performance Measurement.' RAND Journal of Economics 39(4): 1059-1079.

Shalley, Christina E. 1995. 'Effects of Coaction, Expected Evaluation, and Goal Setting on Creativity and Productivity.' Academy of Management Journal 38(2): 483-503.

Shi, Lan. 2009. Respondable Risk and Incentives for Ceo's: The Role of Information-Collection and Decision-Making: University of Washington.

[i] By using the term "controllable risk", we follow the terminology suggested by Gibbs et al. (2009); others have called controllable risk "specific knowledge" (Jensen and Meckling 1992, Raith 2008), or "respondable risk" (Shi 2009).

[ii] In Raith's model, "output" would correspond to a broad or "undistorted" measure, and "input" would correspond to a narrow or "distorted" measure. Indeed, Baker (2002) argues that what we usually call undistorted measures are "outputs" and what we usually call distorted measures are "inputs".

[iii] Such a definition for creativity can be found, e.g., in Shalley (1995).

[iv] Piece rates are based on the quantity of output, whereas reward rates may also be based on the quality of output or other measures of performance.

[v] For more discussion about classifying performance measures as inputs or outputs, see Raith (2008 pp. 1063), and for the relation between distortion and the input-output ranking, see Baker (2002 pp. 738-739).