ETLA Working Papers

No. 20

24 September 2013

Takao Kato
Antti Kauhanen
Julia Salmi

EMPIRICAL EVIDENCE ON THE DYNAMICS OF INCENTIVE PLANS

Suggested citation: Kato, Takao, Kauhanen, Antti & Salmi, Julia (24.9.2013). "Empirical Evidence on the Dynamics of Incentive Plans". ETLA Working Papers No 20. http://pub.etla.fi/ETLA-Working-Papers-20.pdf

ETLA ELINKEINOELÄMÄN TUTKIMUSLAITOS
THE RESEARCH INSTITUTE OF THE FINNISH ECONOMY

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

Empirical Evidence on the Dynamics of Incentive Plans

Takao Kato[a], Antti Kauhanen[*b], Julia Salmi[b]

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

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

Abstract

Using three different waves of a survey on compensation practices from 2005, 2007, and 2010, we study the dynamics of incentive plans. We describe that firms frequently discontinue incentive plans and often change performance measures and organizational levels of performance measurement. The results show that incentive plans are dynamic in nature, but that the dynamics are difficult to explain. The power of the leading theoretical explanations for these changes, i.e., turbulence in the operating environment and learning about performance measures, is not strong.

[*] Corresponding author. Tel. + 358 9 609 90224, Fax + 358 9 601 753. Address: Lönnrotinkatu 4 B, 00120 Helsinki, Finland.

E-mail addresses: tkato@colgate.edu (T. Kato), antti.kauhanen@etla.fi (A. Kauhanen), julia.salmi@helsinki.fi (J. Salmi).

Funding from Tekes—the Finnish Funding Agency for Technology and Innovation—is gratefully acknowledged.

Introduction

There have been many recent advances in theoretical research on performance measurement in incentive systems (Baker 2002, Raith 2008). This recent theoretical literature has tended to address real-world questions, such as distortions caused by performance pay, better than the previous models (Holmstrom and Milgrom 1991, Baker 1992, Baker 2002, Raith 2008). Empirical research on performance measures used in incentive plans is relatively scarce but largely supportive of the economic theories (Ittner and Larcker 2002, Gibbs et al. 2009, Kauhanen and Napari 2012b).

One aspect of incentive systems that has received little attention in the economics literature—and in other streams of literature, such as operations management and accounting—has been the dynamics of incentive systems (Bourne et al. 2000, Henri 2010). Dynamics here refer to changes in performance measures or discontinuance of incentive systems. Adoption of these systems has been studied in the economics literature for a long time (see e.g. Kruse 1996).

The research on the dynamics of incentive systems has focused on the public sector. Courty and Marschke (2003) offer a theoretical analysis of dynamic changes in performance measures and discuss the evolution of performance measurement in the context of a U.S. employment and training program. Heinrich and Marschke (2010) discuss broader evidence from public sector incentive plans. A common theme is that problems in performance measurement are learned over time and that this explains the dynamics. Outside the economics literature, Henri (2010) studies the changes in incentive plans in a sample of Canadian manufacturing firms and Bourne et al. (2000), Kennerley and Neely (2002) offer case studies that tend to emphasize changes in the operating environment of firms as a source of dynamism in performance measurement.

We study the dynamics of incentive systems using three waves of survey data from the Finnish private sector. The survey thus allows us to follow firms over time and observe changes in their incentive plans. We discuss the relevant theories of the dynamics of incentive plans and study the incidence and determinants of the following: i) discontinuing incentive plans, and ii) changes in performance measures and in organizational levels of performance measurement.

Conceptual framework

Why do firms introduce incentive pay plans? The main reason is to improve firm performance, although there may also be other goals (see Kruse 1996). Theoretical work shows that incentive pay may increase the level of

effort and improve the direction of effort when there are multiple tasks (Holmstrom 1979, Holmstrom and Milgrom 1991, Baker 1992, Baker 2002, Raith 2008). There is ample empirical evidence to show that incentive pay increases productivity[1].

Recent theoretical work emphasizes the importance of performance measurement in the success of incentive plans. The early literature was interested in the role of uncontrollable risk for optimal incentives and emphasized the trade-off between risk and incentives. Other difficulties in performance measurement entered into the literature with the contributions of Baker (1992) and Holmstrom and Milgrom (1991). Baker studied optimal incentive plans when the objective of the principal is not contractible and the employee has a better idea of correct actions given the state of the world, whereas Holmstrom and Milgrom (1991) studied the implications of multiple tasks for performance measurement.

In the model of Holmstrom and Milgrom (1991), differences in uncontrollable risk among tasks is reflected in the optimal weights received by each measure and severe problems in performance measurement in one task may lead to fixed wages being the optimal compensation plan. In Baker's model, an important determinant of the intensity of incentives is the correlation of the marginal value of effort with respect to the performance measure and the principal's objective (goal of the organization. The higher this correlation, the higher are the optimal incentives. In both cases, the bottom line is clear: problems in performance measurement lead to lower incentives. Thus, with imperfect performance measures, the intuition is that incentive pay may lead to distortions or unwanted behavior.

Baker (2002) discusses the trade-off between risk and distortion. Distortion arises when the agent's actions have different impacts on the performance measure and the goal of the organization. He argues that there are basically two types of performance measures: narrow (distorted but low risk) and broad (not distorted but risky). Optimal incentive systems combine both broad and narrow measures and attempt to balance the trade-off between risk and distortion by weighting broad and narrow measures optimally.

Combining narrow and broad measures is also justified when employees have specific knowledge (Raith 2008). Specific knowledge refers to knowledge that is costly to communicate (Jensen and Meckling 1992). When this type of knowledge is valuable to the employer, the incentive system should include a broad performance

[1] For case studies, see, e.g., Paarsch and Shearer (1999), Lazear (2000), Paarsch and Shearer (2000), Shearer (2004), Bandiera et al. (2005), Bandiera et al. (2007); for studies using larger data sets, see, e.g., Bhargava (1994), Jones and Kato (1995)

measure that captures the value of such information. Narrower measures are used to diminish the risk that the employee faces.

Job design affects the trade-off between risk and incentives; thus, it affects optimal incentive systems. The broader the job is in terms of number of tasks and decision rights—and the more important specific knowledge is—the broader performance measurement should be. Broad measures are necessary because narrower measures might lead to significant distortions and the measures are not that risky because of broad decision rights. In a narrow job, narrower measures will do fine because they will not cause that many distortions and are not risky. (See Kauhanen and Napari 2012bfor discussion on the differences between incentive plans for white-collar and blue-collar workers).

Incentive systems are complex in the sense that they typically include multiple performance measures and multiple organizational levels of performance measurement (see e.g. Ittner et al. 1997, Ittner and Larcker 2002, Gibbs et al. 2009, Kauhanen and Napari 2012b). The reason for the complexity is that there is no single perfect performance measure; if there were, there would be no need for other measures. In practice, multiple measures are used to mitigate problems caused by others (Gibbs et al. 2009). Economic theory has emphasized the performance measures used and the weights received by each. Arriving at correct measurements is not enough to balance an incentive system, but the weights should also be correct. In practice, this is likely to be difficult.

Thus far, we have discussed incentive plans in terms of worker effort. However, the plans may also have other goals (Ittner and Larcker 2002). For example, the may be used to increase wage flexibility (to shift risks to employees) and to improve recruitment and retention. It is also possible that performance-related pay (PRP) is introduced as a management fad.

In the following, we use this framework of performance measurement in incentive plans to discuss why firms alter or discontinue their incentive plans.

Dynamics of incentive plans

Problems in performance measurement may lead to changing performance measures or to discontinuing the plans altogether. Courty and Marschke (2003) present a model in which the principal learns over time how distorted a performance measure is. They argue that, in practice, the designers of incentive systems do not know in advance how distorted a given measure is. In their model, learning over time leads to changes in performance measures: if a measure is more distorted than expected, it is changed. Heinrich and Marschke (2010) discuss in the public sector context how organizations respond to gaming of incentive systems.

Performance measures may also degrade over time. For example, employees may learn to manipulate the measure over time (Gibbs 2012). This would be one reason, in particular, to change narrow performance measures. Courty and Marschke (2003) also show that even a measure that would be quite good with no incentives (i.e., the marginal value of effort with respect to the performance measure and the goals of the organization are similar) is changed by the introduction of incentives because the employee's behavior and the alignment of the performance measure with the goals of the organization changes.

Incentive systems may also change if job design changes over time or other changes occur that affect the trade-off between risk and distortion. For example, technological change may affect job design by increasing or decreasing decentralization, which changes the incentive system.

Changes in incentive systems may also reflect wider changes within the firm. When complementarities are important, changes in certain organizational decisions lead to multiple simultaneous changes (Milgrom and Roberts 1995). Changes in a firm's strategy should be reflected in the incentive system. In the operations management and accounting literatures, the role of rapid changes in the competitive environment, internal structures, strategic directions, and technologies for the updating of performance measurement systems has been emphasized (Bourne et al. 2000, Henri 2010). In terms of the models discussed above, these changes would most likely address the choice of the narrow performance measures because the theories imply that the ultimate goal of the employer, and hence the broad measure, remains unchanged.

There is also research that suggests that firms should change their internal organization more often than the contingency theories predict. The idea proposed by Nickerson and Zenger (2002) is that firms cycle between formal organizational practices to affect the informal organization. This theory suggests that even in stable environments, firms would alter their organizational choices.

In summary, changes in incentive plans may be caused by problems in performance measurement or changes in the external and internal environments of the firm. Sometimes, problems may be so severe that they lead to firms discontinuing the incentive systems, such as when performance measurement leads to dismantling a plan, which is an example from Baker et al. (1994). It might also be that the costs of operating the plan exceed the benefits. Additionally, firms may make mistakes in implementing the plans that lead to discontinuation of the plan (Kauhanen 2011).

Data

To empirically study the dynamics of incentive plans we use three waves of the compensation surveys of the Confederation of Finnish Industries (EK). The EK is the central organization of employer associations and it has

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 survey has been conducted three times, i.e., in 2005, 2008 and 2011 (the three "waves"). Table 1 presents information about the sample and respondents. The sample size has varied from 2,676 to 3,204 firms and the response rates have been between 31% and 55%, which are good response rates for this type of survey. For example, the European Company Survey (which considers inter alia payment systems) that was conducted 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 approximately 34%.

Table 1: Sample and respondents.

	Sample		Respondents		Respondents share	
	Firms	Employees	Firms	Employees	Firms	Employees
2005	3,204	681,000	958	434,000	30 %	64 %
2008	3,180	722,637	1,738	529,417	55 %	73 %
2011	2,676	729,097	1,204	501,080	45 %	69 %

Notes: In 2008 and 2011, 25% of firms with fewer than 100 employees were included in the sample, and 100% of firms with more than 100 employees. In 2005, the sampling rates for different firm sizes were as follows: 6-9 employees: 10%; 10-49 employees: 20%; 50-99 employees: 50%; >100 employees:100%.

The survey mainly addresses the prevalence and characteristics of PRP systems. Each question relating to PRP plans asks for separate answers for blue-collar workers, clerical employees and white-collar workers. Top management, e.g., CEOs and equivalents, were not included in the survey. The survey allows us to characterize PRP plans in terms of performance measures, organizational levels of performance measurement, and plan coverage. The surveys also ask about views related to the functioning of the current PRP plan and the possible effects that the plan may have had; we utilize this information later. The surveys also request certain basic information about the firm, such as industry, export status and ownership status.

To analyze changes in the incentive plans over time, we must observe firms in multiple waves. There are 2,895 unique firms in the three waves of the data. Of these, 2,112 appear only once, 565 twice, and 218 are in all three waves, which indicates that there are enough firms to conduct the analysis.

We link certain other data sources to the surveys. First, certain financial statistics are linked from balance sheet data (Asiakastieto). Second, to control for average level of education, work experience and tenure of the firms' employees, we link information from EK's wage statistics. These data are a linked employer-employee data with a rich content of information about the employees (for more details, see e.g. Kauhanen and Napari 2012a). We also control for foreign ownership, which was an important determinant of PRP in Finnish firms in the 1990s that continues to have influence today; this information is linked from Statistics Finland's business register.

Results

Incidence of PRP

To begin the analysis, Table 2 reports the percentage of firms using PRP. The first three rows of the table present results for the three employee groups, whereas the last two rows show figures for firms with PRP plans for all groups and for at least one group. In 2004, 34.6 % of firms had performance pay schemes for blue-collar workers. For white-collar workers and upper white-collar workers, the percentages were 52.2 % and 55.3 %, respectively, which is noticeably higher than for blue-collar workers. This is consistent with the argument in Kauhanen and Napari (2012b), for example, that these types of PRP plans should be more common for white-collar employees.

Table 2: Incidence of PRP.

	year			
	2004	2007	2010	Total
	%	%	%	%
Blue-collar PRP	34.6	32.5	33.1	33.2
White-collar PRP	52.2	46.9	50.4	49.3
Upper white-collar PRP	55.3	49.3	54.2	52.3
PRP for all employee groups	31.5	28.7	29.8	29.7
PRP for at least one employee group	59.2	53.6	57.7	56.2

In 2007, the percentage of firms using PRP dropped in all groups. This may be largely explained by sampling and survey design. The 2007 survey included more small companies than the other surveys, which means that there will be fewer firms that use PRP. Notwithstanding the sampling and survey design issues, PRP schemes have become less frequent over time; the percentages in 2010 are below the 2004 levels.

Adoption and discontinuation of PRP plans

The figures discussed above hide many instances of adoption and discontinuation of PRP plans. We have 907 consecutive firm-year observations that can be used to capture movements in and out of PRP schemes in Table 3. Adoption indicates that firm did not have a PRP plan at *t-1* but has one at *t* and discontinuing of a plan is defined analogously. We find that it is common to make comprehensive changes to performance pay by abandoning and introducing PRP schemes. The patterns are similar for all employee groups; the only major difference is that PRP schemes are generally more common for white-collar and upper white-collar workers. For example, 233 firms had no PRP plan for white-collar workers during either period, 89 firms abandoned their schemes, 106 introduced PRP, and 479 firms had a scheme during both periods. The in- and outflows cancel one another out for the most part, which keeps the total percentage relatively stable. More PRP schemes are introduced in Table 3 than are abandoned, which are contrary to the results from Table 2. This is because in Table 3, we follow firms who respond in at least two waves of the survey, which means that the samples are a somewhat different.

Table 3: Adoption and discontinuation of PRP plans.

	PRP for blue-collars t+1		
Blue-collar PRP	0	1	Total
0	394	110	504
1	99	304	403
Total	493	414	907
	PRP for white-collars t+1		
White-collar PRP	0	1	Total
0	233	106	339
1	89	479	568
Total	322	585	907
	PRP for upper white-collars t+1		
Upper white-collar PRP	0	1	Total
0	193	107	300
1	87	520	607
Total	280	627	907
	PRP for at least one employee group t+1		
PRP for at least one employee group	0	1	Total
0	179	90	269
1	82	556	638
Total	261	646	907

To investigate why firms abandon PRP, we regress abandoning of a PRP plan on explanatory variables that are measured at *t-1*.

We construct variables to depict firms' satisfaction with PRP to measure success in performance measurement and plan implementation. "Impact of the plan" is comprised of evaluations of how PRP has affected the following ten selected aspects of work: interest in work, taking responsibility, openness, cooperation, information flow, flexibility, improvement of operations, productivity, quality, and profitability. The respondents estimated the size of the impact using a 6-point Likert scale in which options ranged from negative to positive. The impact of the plan is the average of the ten aspects.

The views regarding the “functioning” of PRP systems are also based on 6-point Likert scale evaluations that range from “bad” to ”good”. Here, the question was concerned with both the impact of the system and experiences with the system. Two dummies were constructed based on the answers. If the overall impact was estimated to be ‘good’, the variable “Functioning: Well” obtains a value of 1 (zero otherwise). If the impact was ‘quite good’ or ‘in-between’ the dummy “Functioning: OK, no change” obtains a value of 1 (zero otherwise).

The OLS regression results from are reported in Table 4. The decision to abandon a PRP scheme is explained with the estimated impact and functioning, ownership status, size, and profits of the firm. However, the reasons behind abandoning PRP schemes remain largely unclear. The decision to abandon PRP is not explained at all by how well the system was perceived to function. The impact of the plan only has a significant effect among blue-collar workers. The sign is as expected, i.e., the schemes that have a positive impact are less likely to be abandoned.

The regression reveals that state-owned companies are significantly less likely to terminate PRP schemes for white-collar workers. Otherwise, ownership status, size or changes in profits do not explain why firms abandon performance pay schemes. Moreover, the constant terms for all employee groups are estimated to be statistically significant, which implies that many PRP schemes might be temporary by nature.

These results tend to suggest that problems with performance measurement do not explain decisions to discontinue PRP plans. Thus, it is likely that other factors, such as changes in the operating environment of the firm, are driving the results. However, the nature of these changes remains unclear. The financial crisis, controlled by a year dummy, does not explain abandoning. It seems that either the crisis did not have an influence or the influence was mixed.

Table 4: Abandonment of PRP plans.

	(1)	(2)	(3)
	Worker	Clerical	Upper White-Collar
Impact of the plan (based on 10 variables)	-0.0871*	-0.0297	-0.0130
	(-1.708)	(-0.746)	(-0.356)
Firm size	-1.59e-05	-1.14e-05	-1.12e-05
	(-1.490)	(-1.494)	(-1.479)
Gross change in profits	0.000110	4.65e-05	0.000121
	(0.677)	(0.300)	(0.912)
State-owned	-0.0838	-0.116***	-0.105***
	(-0.912)	(-4.263)	(-3.909)
Community-owned	-0.0299	-0.0590	0.0324
	(-0.246)	(-0.669)	(0.279)
Foreign owner	0.0870	0.0499	0.0529
	(1.064)	(0.896)	(1.020)
Year 2007	0.0814	0.0257	0.0234
	(1.512)	(0.621)	(0.589)
Functioning: OK, no change	-0.0219	-0.0922	-0.0902
	(-0.222)	(-0.905)	(-0.758)
Functioning: Well	0.0223	-0.0947	-0.137
	(0.225)	(-0.961)	(-1.230)
Constant	0.166*	0.229**	0.252**
	(1.822)	(2.314)	(2.259)
Observations	217	298	314
R-squared	0.048	0.034	0.037

Table 5: Order of implementation of PRP plans.

PRP for workers (t-1) = 0		
	PRP for upper white-collars (both t-1 and t)	
PRP for workers (t)	0	1
0	82 %	73 %
1	18 %	**27 %**
n	336	209
PRP for white-collars (t-1) = 0		
	PRP for upper white-collars (both t-1 and t)	
PRP for white-collars (t)	0	1
0	75 %	32 %
1	25 %	**68 %**
n	308	56

Table 5 inspects patterns in implementing PRP systems. The table is split into two parts, one set for blue-collar workers and one for clerical workers. The data have been chosen such that, during the first of two consecutive periods/observations, the employee group of interest has no PRP system. To inspect whether the presence of an upper white-collar PRP system affects implementing PRP with other employee groups, the only pairs of observations that are included are those in which the pay systems of upper white-collar workers are held constant over time: either the firm does not have a PRP plan in the two consecutive observations or it has a plan in both instances.

We find that it is common to introduce a PRP scheme to upper white-collar workers first and later to expand PRP to other employee groups, particularly to other white-collar workers. By the second period, 27 % of these firms had implemented a PRP scheme for blue-collar workers, whereas the percentage was as high as 68 % for other white-collar workers.

In addition to plans that are transferred from one group to another, the similarity of performance measures for different groups must be addressed. Unreported results show that companies frequently use similar PRP systems among different employee groups in terms of different measures and measurement levels. The PRP

schemes for blue-collar workers have identical measures and measurement levels as those for upper white-collar workers in 35 % of companies. Blue-collar and clerical workers have similar systems in 47 % of the firms, and the PRP systems use identical measures and measurement levels for clerical and upper white-collar workers in 49 % of firms. This is surprising when considering the assumption that white-collar jobs involve a higher degree of decision making and more specific knowledge, which might, therefore, require a broader performance measurement.

Changes in performance measurement

Next, we describe how common it is to change performance measures and the organizational levels of performance measurement. The theories discussed above suggest that firms may wish to change their PRP plans when their environments change or because of the need to learn about performance measures.

We look at changes in the *number* of measures and the *number* of levels of measurement and at changes in the measures and levels themselves.

The average number of performance measures remains relatively constant over time for all employee groups. Although the increase in the measures was statistically significant in 2010 for all groups, as observed in Table 6, and for upper white-collar workers in 2007, the overall increases have been small. In PRP schemes for blue-collar employees, for example, the average number of measures increased by 0.04 measures in the 2010 survey. In 2007, the average number of measures was 4 for blue-collar workers, 4.6 to clerical and 5 for upper white-collar workers (Kauhanen and Napari 2012b).

Table 6: Changes in the number of performance measures.

	(1)	(2)	(3)
	Workers	Clerical	Upper White-Collar
2007	0.0181	0.0215	0.0232*
	(1.211)	(1.584)	(1.670)
2010	0.0394**	0.0421***	0.0351**
	(2.555)	(2.682)	(2.132)
Observations	264	421	455
R-squared	0.031	0.024	0.017
Robust t-statistics in parentheses.			
*** p<0.01, ** p<0.05, and * p<0.1.			

The estimates might also be biased downward as a result of questionnaire updating. In the 2010 survey, two infrequently used measures were removed from the options; although this has been taken into account in calculating the change, some bias might remain. The different measures available are profitability, cost savings, quality, productivity, lead time, inventory quantity or value, development goals, sales targets, utilization of capital, turnover, market share and 'other' measures; however, inventory quantity or value and utilization of capital were only available in the 2004 and 2007 surveys. All these measures are classified as narrow except for profitability, which is classified as broad.

Performance is usually assessed with multiple measures. Table 7 describes changes in the combinations of these measures. As noted above, the total number of measures has been stable; thus, the results of Table 7 may be interpreted in terms of swapping one measure for another. For example, between 2004 and 2007, the measures for upper white-collar employees changed by 0.3. The number expressed the number of changes (adapting a measure or removing it) relative to the total number of measures available. Thus, assuming a constant number of measures, firms have changed 15 % of the performance measures, on average. PRP schemes for upper white-collar employees typically use approximately 40% of the available measures (i.e., 5 measures); thus, on average, firms change just over 1/3 of the measures (slightly under two). This is a significant number and cannot be explained by the small increase in the number of performance measures in use that are shown in Table 6.

Table 7: Changes in individual performance measures.

	(1)	(2)	(3)
	Workers	Clerical	Upper White-Collar
2007	0.253***	0.287***	0.297***
	(16.67)	(22.25)	(24.07)
2010	0.233***	0.284***	0.311***
	(17.99)	(22.35)	(25.26)
Observations	264	421	455
R-squared	0.695	0.704	0.729
Robust t-statistics in parentheses.			
*** p<0.01, ** p<0.05, and * p<0.1.			

Table 8 inspects changes in the number of levels in which performance is measured. The levels available in the survey include: own work, team, establishment, department, project, profit center, company, and the group (of companies). The results indicate that the number of levels has increased slightly in all employee groups but that this change has been mainly statistically insignificant. Upper white-collar workers in 2007 are an

exception; for this category, the number of levels increased significantly by 0.3. Thus, compared to 2004, the performance of every third upper white-collar worker was measured at an additional level in 2007.

Table 8: Changes in the number of levels of measurement.

	(1)	(2)	(3)
	Workers	Clerical	Upper White-Collar
2007	0.110	0.158	0.323**
	(0.903)	(1.295)	(2.564)
2010	0.127	0.191	0.100
	(0.917)	(1.610)	(0.796)
Observations	261	418	452
R-squared	0.006	0.010	0.016
Robust t-statistics in parentheses.			
*** p<0.01, ** p<0.05, and * p<0.1.			

As with the individual performance measures, firms also swap the levels of measurement. Table 9 shows that, between 2007 and 2010, the estimate for blue-collar workers was 0.25. Assuming that the number of levels remains constant—which is reasonable because the rise was not significant as reported in Table 8—firms have changed 12.5 % of the measurement levels, on average. There are eight options available, which indicates that one level is swapped, on average. The performance of a blue-collar worker is typically measured at 2.5 levels. For white-collar workers, the number is 3.3; for upper white-collar workers, that number is 3.5 (Kauhanen and Napari 2012b).

Table 9: Changes in individual levels of measurement.

	(1)	(2)	(3)
	Workers	Clerical	Upper White-Collar
2007	0.269***	0.288***	0.308***
	(11.48)	(14.49)	(14.83)
2010	0.253***	0.284***	0.292***
	(11.21)	(14.60)	(14.56)
Observations	261	418	452
R-squared	0.499	0.505	0.490
Robust t-statistics in parentheses.			
*** p<0.01, ** p<0.05, and * p<0.1.			

These results indicate that firms continuously change their incentive plans. However, the results do not show which measures firms change, and it is difficult to characterize changes in incentive plans. There are no measures that would be more likely to be abandoned or introduced. Table 10 clarifies the picture somewhat by providing information about the general direction of the change.

We find that there has been significant movement toward more comprehensive measurement systems. Following Kauhanen and Napari (2012b), we classify incentive plans according to the number of measures and their breadth. In Table 10, the dependent variable obtains the value of one when the measures become broader by moving up in the following system: the first stage consists of one narrow measure, the second stage utilizes one broad measure (profit), the third level contains several narrow measures and the fourth and final level employs a combination of narrow and broad measures. In all other instances, the dependent variable obtains the value of zero. With this construction, we observe that firms have been consistently moving toward broader evaluation of performance with each employee group. The change has been largest among workers, which is not entirely unexpected as they have typically had the narrowest measures (Kauhanen and Napari 2012b).

Table 10: Changing to broader measurement.

	(1)	(2)	(3)
	Workers	Clerical	Upper White-Collar
2007	0.496***	0.352***	0.299***
	(10.86)	(10.29)	(9.530)
2010	0.529***	0.362***	0.301***
	(12.47)	(10.52)	(9.471)
Observations	261	392	423
R-squared	0.514	0.357	0.300
Robust t-statistics in parentheses.			
*** p<0.01, ** p<0.05, and * p<0.1.			

Why do firms change their incentive plans?

The next two tables help explain what types of firms change the performance measures of their PRP systems and why. In Table 11, the degree of changes made to the performance measures is regressed against the

original reasons to adopt PRP[2]. The options available as original reasons to adopt PRP included the following: for purposes of compensation, as a tool of management, to increase wage flexibility, and because PRP was used with other employee groups or firms.

PRP systems that were adopted to compensate personnel or to increase wage flexibility showed little power to explain how much the system had been changed and why. Conversely, when PRP was implemented as an instrument of ordinary management, a significantly higher degree of changes were made to both types of white-collar employee PRP systems. When PRP was implemented because others used it—either within the firm or in other firms—the level of changes made to workers' PRP systems increased, whereas this same circumstance decreased changes made to upper white-collar workers' PRP systems.

The results are different between employee groups but can be largely explained in terms of the initial motives to adopt PRP because they reflect the objectives of the firm to a degree. "Because others use" is a passive reason, whereas PRP as part of management implies more active intensions to utilize performance pay. Given that upper white-collar workers usually obtain PRP systems first and that the systems are often similar among different employee groups, the systems of blue-collar workers that have been implemented in imitation of the systems of others might require larger adjustments whereas the systems of upper white-collar workers have been refined over the years. This could explain the difference in the signs of the estimates.

[2] The dependent variable is thus the same as in Table 9.

Table 11: Changes in performance measures and the reason for the adoption of PRP.

	Worker	Clerical	Upper White-Collar
Reason to use PRP:			
a) Compensation	0.0311	0.0154	0.0177
	(1.069)	(0.536)	(0.682)
b) Part of management	0.0271	0.0487**	0.0382*
	(1.204)	(2.342)	(1.769)
c) Wage flexibility	0.0151	0.0254	-0.00952
	(0.693)	(1.191)	(-0.461)
d) Because others use	0.0512*	0.0191	-0.0379*
	(1.750)	(0.714)	(-1.889)
Constant	0.188***	0.226***	0.266***
	(5.613)	(7.174)	(8.627)
Observations	264	422	456
R-squared	0.024	0.019	0.012
Robust t-statistics in parentheses.			
*** p<0.01, ** p<0.05, and * p<0.1.			

The level of changes made to the measures is explained in Table 12 in terms of broad industry classifications and their interactions with the year 2010, the estimated impact and functioning of the PRP plans and certain basic characteristics of the firms. Additionally, the regressions control for differences in internationality, ownership, and the mean experience, seniority and education levels of the employees.

We divide firms into the following four categories: 1) Technology industries, 2) Other manufacturing, 3) Business services, and 4) Other services. We make these classifications because the operating environment has changed most radically in the technology industries. Productivity growth has been rapid and restructuring in the industry has been strong (Hyytinen and Maliranta 2013). The recent financial crisis also hit these industries hardest. Productivity growth in services has been much slower and the role of restructuring is smaller. This classification offers the possibility of looking into the impact of changes in operating environment with respect to changes in incentive plans.

We find that firms operating in the technology industries make fewer changes, on average, to their performance measurement systems than firms in the manufacturing and processing industries. The manufacturing and processing industries (other than technology) are used as the point of comparison in the regressions. The technology industries dummy is significant for workers and clerical employees. The interaction term of technology industries with the year 2010, however, is significantly positive for all workers, which indicates that firms in the technology industry made significantly more changes for this employee group between 2007 and 2010. Their environment changed most rapidly and the results reflect this.

Firms in the business services sector also made significantly fewer changes in the case of upper white-collar employees to their PRP schemes compared to firms in the manufacturing and processing industries. Notably, the interaction term with the year 2010 is positive and significant for white-collar employees in business services. Changes made to PRP schemes in the rest of the service sector did not differ much from those in manufacturing and processing, and the results are mostly similar for all employee groups. However, firms in the other services made significantly fewer changes to the performance measures of clerical workers in 2010.

The assessments of the functioning and impact of performance pay systems do not sufficiently explain many of the changes observed in the performance measures. A system that is perceived to function well with a positive impact is significant only with respect to PRP plans for lower white-collar workers. Otherwise, the perceived impact of the plan or the functioning of the plan do not explain changes in performance measures.

All in all, problems in performance measurement, as gauged by the functionality of the plan, do not seem to be important for changing performance measures.

Regarding the basic characteristics of a firm we find that larger and younger firms change their performance pay systems more than smaller and older firms. Age is a significant explanatory variable with lower white-collar workers whereas firm size is significant with both white-collar groups. Larger firms may have more resources with which to manage their incentive plans and older firms have had more time to learn about better performance measurement. The mean hourly basic wage is significant without the control variables; however, when differences in mean seniority, experience and education are accounted for, wage level becomes insignificant.

Again, we find that the results differ between employee groups—at least with respect to the significance of estimates. Changes made to the PRP plans of lower white-collar workers are mostly explained by the factors considered and reported in Table 12. This result might reflect the differences in the motives for implementing performance pay. For example, when performance pay schemes are expanded from one group to another, the

schemes are not necessarily well tailored for the new group. Thus, the characteristics of the PRP scheme might not reflect the relevant work environment of all the employees.

Table 12: Changes in performance measures.

	Workers	Clerical	Upper White-Collar
Technology industries	-0.0904*	-0.0891*	-0.0602
	(-1.877)	(-1.923)	(-1.063)
Business services	-0.0242	-0.0498	-0.101**
	(-0.342)	(-1.173)	(-2.390)
Other services	0.0486	0.0459	-0.00748
	(1.398)	(1.264)	(-0.226)
Technology industries*2010	0.145***	0.0754	0.0185
	(2.853)	(1.469)	(0.291)
Business services*2010	0.115	0.114*	0.166***
	(0.562)	(1.840)	(2.897)
Other services*2010	-0.0552	-0.0517	-0.0295
	(-1.444)	(-1.293)	(-0.814)
Firm size (1000)	0.0176*	0.0144**	0.0151**
	(1.693)	(2.178)	(2.378)
Age	-0.000963	-0.00151**	-0.000835
	(-1.313)	(-2.347)	(-1.278)
Impact: mean (standardized items)	0.0158	0.0109	0.0191
	(0.792)	(0.412)	(0.846)
Functioning: Well	0.0839**	0.0288	-0.00308
	(2.300)	(0.703)	(-0.0661)
Functioning: OK, no change	-0.000152	-0.00417	0.00198
	(-0.00446)	(-0.113)	(0.0443)
Constant	0.198***	0.277***	0.308***
	(5.681)	(6.944)	(6.320)
Observations	200	295	317
R-squared	0.150	0.084	0.063
Robust t-statistics in parentheses.			
*** p<0.01, ** p<0.05, and * p<0.1.			

The regressions also control for ownership.

Conclusion

Successful performance measurement is the key to a good incentive plan. Despite theoretical advances, there remains little empirical research on performance measurement in incentive plans. Recent studies have focused on performance measures, but there are few studies on changes in performance measurement. Theoretical research in economics, and studies in the fields of operations management and accounting, suggest that firms are likely to change their performance measures as the result of learning more information about the performance measures or about changes in their operating environment.

Using three waves of a survey on compensation practices, we study the dynamics of incentive plans. The first phenomenon we study is abandoning incentive plans. We find that a substantial share of firms discontinue their incentive plans, but that perceptions about the functioning or impact of the plan does not explain these decisions. This suggests that reasons other than problems in performance measurement are likely to causing these results.

Firms also change performance measures and organizational levels of performance measurement. The number of measures and levels is constant over time, but individual levels and measures change. We find some evidence that supports that both changes in operating environment and problems in performance measurement correlate with changes in performance measures.

The results show that incentive plans are dynamic in nature, but that the dynamics are difficult to explain. Changes in operating environments and learning about performance measures matter in the dynamics of PRP schemes, but their exact role is difficult to capture in the type of survey employed in this study. Case study approaches might be useful complements for larger scale studies.

References

Baker, George. 2002. "Distortion and Risk in Optimal Incentive Contracts." Journal of Human Resources, Vol. 37, No. 4, pp. 728-751.

Baker, George, Robert Gibbons and Kevin J. Murphy. 1994. "Subjective Performance Measures in Optimal Incentive Contracts." Quarterly Journal of Economics, Vol. 109, No. 4, pp. 1125-1156.

Baker, George P. 1992. "Incentive Contracts and Performance Measurement." Journal of Political Economy, Vol. 100, No. 3, pp. 598-614.

Bandiera, Oriana, Iwan Barankay and Imran Rasul. 2005. "Social Preferences and the Response to Incentives: Evidence from Personnel Data." Quarterly Journal of Economics, Vol. 120, No. 3, pp. 917-962.

Bandiera, Oriana, Iwan Barankay and Imran Rasul. 2007. "Incentives for Managers and Inequality among Workers: Evidence from a Firm-Level Experiment." Quarterly Journal of Economics, Vol. 122, No. 2, pp. 729-773.

Bhargava, Sandeep. 1994. "Profit-Sharing and the Financial Performance of Companies: Evidence from U.K. Panel Data." Economic Journal, Vol. 104, No. 426, pp. 1044-1056.

Bourne, Mike, John Mills, Mark Wilcox, Andy Neely and Ken Platts. 2000. "Designing, Implementing and Updating Performance Measurement Systems." International Journal of Operations & Production Management, Vol. 20, No. 7, pp. 754-771.

Courty, Pascal and Gerald Marschke. 2003. "Dynamics of Performance-Measurement Systems." Oxford Review of Economic Policy, Vol. 19, No. 2, pp. 268-284.

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

Gibbs, Michael. 2012. "Design and Implementation of Pay for Performance". Forschungsinstitut zur Zukunft der Arbeit Institute for the Study of Labor **6322**.

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, Vol. 48, No. 2, pp. 237-264.

Heinrich, Carolyn J. and Gerald Marschke. 2010. "Incentives and Their Dynamics in Public Sector Performance Management Systems." Journal of Policy Analysis and Management, Vol. 29, No. 1, pp. 183-208.

Henri, Jean-François. 2010. "The Periodic Review of Performance Indicators: An Empirical Investigation of the Dynamism of Performance Measurement Systems." European Accounting Review, Vol. 19, No. 1, pp. 73-96.

Holmstrom, Bengt. 1979. "Moral Hazard and Observability." Bell Journal of Economics, Vol. 10, No. 1, pp. 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, Vol. 7, No., pp. 24-52.

Hyytinen, Ari and Mika Maliranta. 2013. "Firm Lifecycles and Evolution of Industry Productivity." Research Policy, Vol. 42, No. 5, pp. 1080-1098.

Ittner, Christopher D. and David F. Larcker. 2002. "Determinants of Performance Measure Choices in Worker Incentive Plans." Journal of Labor Economics, Vol. 20, No. 2, pp. S58-S90.

Ittner, Christopher D., David F. Larcker and Madhav V. Rajan. 1997. "The Choice of Performance Measures in Annual Bonus Contracts." Accounting Review, Vol. 72, No. 2, pp. 231.

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

Jones, Derek C. and Takao Kato. 1995. "The Productivity Effects of Employee Stock-Ownership Plans and Bonuses: Evidence from Japanese Panel Data." American Economic Review, Vol. 85, No. 3, pp. 391-414.

Kauhanen, Antti. 2011. "The Perils of Altering Incentive Plans: A Case Study." Managerial and Decision Economics, Vol. 32, No. 6, pp. 371-384.

Kauhanen, Antti and Sami Napari. 2012a. Career and Wage Dynamics: Evidence from Linked Employer-Employee Data. In Solomon W. Polachek and Konstantinos Tatsiramos. Research in Labor Economics, pp 35-76: Emerald Group Publishing Limited.

Kauhanen, Antti and Sami Napari. 2012b. "Performance Measurement and Incentive Plans." Industrial relations: a journal of economy and society, Vol. 51, No. 3, pp. 645-669.

Kennerley, Mike and Andy Neely. 2002. "A Framework of the Factors Affecting the Evolution of Performance Measurement Systems." International Journal of Operations & Production Management, Vol. 22, No. 11, pp. 1222-1245.

Kruse, Douglas L. 1996. "Why Do Firms Adopt Profit-Sharing and Employee Ownership Plans?" British Journal of Industrial Relations, Vol. 34, No. 4, pp. 515-538.

Lazear, Edward P. 2000. "Performance Pay and Productivity." American Economic Review, Vol. 90, No. 5, pp. 1346-1361.

Milgrom, Paul and John Roberts. 1995. "Complementarities and Fit: Strategy, Structure, and Organizational Change in Manufacturing." Journal of Accounting and Economics, Vol. 19, No. 2-3, pp. 179-208.

Nickerson, Jack A and Todd R Zenger. 2002. "Being Efficiently Fickle: A Dynamic Theory of Organizational Choice." Organization Science, Vol. 13, No. 5, pp. 547-566.

Paarsch, Harry J. and Bruce Shearer. 2000. "Piece Rates, Fixed Wages, and Incentive Effects: Statistical Evidence from Payroll Records." International Economic Review, Vol. 41, No. 1, pp. 59-92.

Paarsch, Harry J. and Bruce S. Shearer. 1999. "The Response of Worker Effort to Piece Rates: Evidence from the British Columbia Tree-Planting Industry." Journal of human resources, Vol. 34, No. 4, pp. 643-667.

Raith, Michael. 2008. "Specific Knowledge and Performance Measurement." RAND Journal of Economics Vol. 39, No. 4, pp. 1059-1079.

Shearer, Bruce. 2004. "Piece Rates, Fixed Wages and Incentives: Evidence from a Field Experiment." Review of Economic Studies, Vol. 71, No. 2, pp. 513-534.