Performance Measurement and Incentive Intensity

Alberto Bayo-Moriones
Jose E. Galdon-Sanchez
Sara Martinez-de-Morentin

July 2015

Performance Measurement and Incentive Intensity

Alberto Bayo-Moriones
Universidad Publica de Navarra

Jose E. Galdon-Sanchez
Universidad Publica de Navarra
and IZA

Sara Martinez-de-Morentin
Universidad Publica de Navarra

ABSTRACT

Performance Measurement and Incentive Intensity[1]

This study addresses the factors that determine the intensity of pay for performance schemes. The results indicate that the use of individual and group incentives boost intensity, whereas plant or firm pay for performance do not seem to affect the variable of interest. In addition, the adoption of measures of results, such as productivity or quality, has a significant positive effect on intensity. On the contrary, measures of human resource management outcomes, subjective measures and financial measures are not significant or have a negative effect on the intensity of pay for performance.

Corresponding author:

Jose E. Galdon-Sanchez
Department of Economics
Universidad Publica de Navarra
Campus de Arrosadia
Pamplona 31006
Spain
E-mail: jose.galdon@unavarra.es

Introduction

Performance measurement is a key issue both in the basic agency model and in subsequent theoretical extensions (for a review of theoretical work see, for example, Gibbons and Waldman, 1999; Prendergast, 1999; or Raith, 2008). More specifically, incentive theory establishes that the optimal intensity of incentives (that is, the proportion that variable pay represents in total compensation) depends on the properties of performance measures. For example, a body of research concludes that the design of incentives should be based on the informativeness principle. According to this principle any (costless) measure that adds information about workers' effort should be used in compensation contracts[1].

In the last few years, empirical work has addressed the influence of performance measurement on incentive intensity (see Zenger and Marshall, 2000; Bowens and Van Lent, 2006; Gibbs et al., 2009; Kauhanen and Napari, 2012, among others). Despite the advances made by these recent studies, empirical work on this issue is still scarce, and much remains to be learned about how the decision to implement low or high powered incentives is made. Consequently, it is still not clear how organizations actually design their incentive schemes. In addition, the way incentives are implemented does not always match the theoretical predictions, as occurs with the informativeness principle previously mentioned (see Raith, 2008).

One of the main difficulties faced by the study on the determinants of incentive intensity is the lack of comprehensive data on how organizations actually design their incentive systems. As a consequence, work on the topic present certain limitations. For example, some authors include the number of performance measures as a determinant of intensity, but they do not consider the influence of specific measures and their properties (see Zenger and Marshall, 2000). Others take into account only certain properties of performance measures, such as risk and distortion (see Bowens and Van Lent, 2006). Certain analyses focus on the examination of case studies (see for example Gibbs et al., 2009), while others do not make in-depth studies of the implications of performance measurement for incentive intensity. For example, Kauhanen and Napari (2012) have extensive information about performance measures and measurement levels, but they only focus on the descriptive analysis of such information.

Our aim here is to complement this empirical research and examine the influence that performance measurement exerts on the intensity of incentives. In order to do so, we address two specific questions. First, we examine how the measurement level influences incentive intensity. The choice of the measurement level

is one of the relevant dimensions employers have to consider when designing their incentive schemes (see Baker, 2002). As Zenger and Marshall (2000) suggest, individual and collective incentives have different properties in terms of noise, distortion, and other attributes. For example, the adoption of group or plant incentives could help to solve potential problems such as distorted behavior by the worker. Consequently, we expect to find differences in the intensity of incentives depending on the organizational level of performance measurement.

Second, we take into account a broad set of measures and examine their influence on intensity for individual incentives, group incentives, and plant or firm incentives. Many studies on the topic analyze more vague categories such as financial and non-financial measures (see Ittner and Larcker, 2002). In this work, our main aim is to consider a set of indicators that comprehensively represents the measurement options available to employers. We think that our group of measures captures the specific attributes of the measurement process with precision. Consequently, we are able to carry out a detailed analysis of the determinants of incentive intensity. As we have already mentioned, we perform this analysis for the three different measurement levels available to organizations: individual group and plant or firm. We are not aware of previous work that analyses the relationship between performance measurement and incentive intensity in such an exhaustive way. Since the design of incentives at different levels may be driven by different forces, we go a step further in the study of the practice and examine the determinants of incentive intensity from a comprehensive perspective.

The analysis is based on a unique data set that contains extensive information on the provision of incentives in manufacturing establishments. More specifically, it includes data about the use of pay for performance (from now on, PfP), and about the use of PfP measured at individual (*Individual PfP*), group (*Group PfP*) and plant or firm (*Plant or Firm PfP*) levels. Regarding incentive intensity, the data reports on the proportion of total pay that depends on performance for each PfP scheme. In addition, it specifies which measures are used to determine performance, and contain information about a set of 11 indicators. To our knowledge, our data set constitutes a unique source of information about the relationship between PfP intensity and the use of performance measures at the three broad levels at which PfP can be implemented.

In order to study this relationship, our empirical strategy is as follows. First, we classify performance measures into four categories: *Results Category*, human resource management *(HRM) Outcomes Category,*

Subjective Category and *Financial Category*. Then, we characterize each category in terms of five properties: *Noise, Controllable Risk*, *Impact on Firm Value*, *Distortion* and *Manipulability*. In addition, we describe the properties of *Individual PfP, Group PfP* and *Plant or Firm PfP*. Once we have completed our characterization of performance measures and PfP schemes, we perform OLS regressions of the determinants of incentive intensity.

Our findings show that results measures are the most common indicators of performance. In our sample, the intensity of incentives increases with the use of *Individual PfP* and *Group PfP*. On the contrary, there is no significant effect of *Plant or Firm PfP* use on the variable of interest. The use of the *Results Category* also has a positive effect on the intensity of incentives.

The remainder of the paper is organized as follows. Section 2 describes the factors that, according to the theory, determine the intensity of incentives. Section 3 deals with the level of performance measurement and its relationship with intensity. In Section 4, we analyze the indicators used to measure performance and their properties. The next section describes the data set used in this study, and Section 6 presents the results of the empirical analysis. In the last section we draw some conclusions.

The determinants of incentive intensity

From the literature on incentive provision and performance measure properties, we compile a set of factors that may influence incentive intensity. We begin by examining the agency model, which is the classic framework for understanding the provision of incentives in organizations (see Gibbons and Waldman, 1999). The central point of this model is the tradeoff between risk and incentives. The maximization problem for determining the optimal bonus rate concludes that intensity is negatively related to the agent's risk aversion, to the uncertainty in the production process and to the rate at which the marginal cost of effort increases. This framework reveals a relevant influence on incentive intensity for the purposes of our study: noise. When noise is high, the principal should increase wages to compensate workers for the risk they are exposed to (see Prendergast, 1999). From these ideas, we identify noise as the first relevant determinant of incentive intensity. More precisely, agency theory predicts a decrease of incentive intensity when noise in the production process is high.

The recent literature on incentive intensity makes a distinction between uncontrollable and controllable risk

(see Gibbs et al., 2009 and Kauhanen and Napari, 2012; among others). On the one hand, uncontrollable risk refers to noise, that is, those environmental factors that workers are unable to control. On the other hand, controllable risk refers to the actions a worker can take to respond to risk, which depend on the worker's specific knowledge or her private information (see Kauhanen and Napari, 2012). In other words, this factor alludes to the extent workers can respond to uncertainty (see Gibbs et al., 2009). The relevance of distinguishing between the two types of risk lies in their influence on incentive intensity. In the presence of controllable risk, incentives are useful because they motivate workers to use their private information when changes in the environment occur. Therefore, controllable risk should foster incentive intensity in order to enhance workers to act in the organization's best interest.

Another extension to the basic agency model that is useful for the purposes of our analysis is that proposed by Baker (1992). This author pointed out the difference between the worker's measured performance and the worker's total contribution to firm value, which are assumed to be the same in the baseline model. According to this model, if measured performance is highly correlated with firm value, the principal should set strong incentives (high intensity). On the contrary, a low association between measured performance and organizational value renders low incentives optimal. Consequently, the link between measured performance and firm value should be a relevant determinant of incentive intensity. For example, the connection between workers' performance and firm value varies between jobs, since certain positions have a higher impact on organizational value than others. Ortin-Angel and Salas-Fumas (1998) show that this impact becomes less straightforward as we move down the organizational hierarchy, decreasing from top executives to middle managers. Hence, the intensity of incentives should be higher for those jobs whose performance has a high influence on firm value.

The literature on incentive provision also makes reference to the idea of distortion (see Kauhanen and Napari, 2012) or dysfunctional behavior (Prendergast, 1999). The problem of distortion emerges when the principal rewards the wrong behavior (see Holmstrom and Milgrom, 1991; Baker, 1992). This problem is particularly prevalent when multitasking is present, leading agents to focus on some tasks more than on others. For example, rewarding volume may incentivize workers to put relevant dimensions of performance, such as quality or cooperation with other workers, aside. As Baker et al. (1994) pointed out, "firms get what they pay for". Consequently, incentives should be carefully designed in order to limit the occurrence of

dysfunctional behavior. Regarding the size of variable pay, the possibility of distortion should diminish intensity.

Finally, incentive provision may suffer from a problem of manipulation of performance measures. This problem emerges when workers scam the incentive scheme, increasing their performance ratings without improving organizational value (see Gibbs et al., 2009). For example, they can manipulate performance standards concerning production time or volume. Due to the presence of information asymmetries, it might be difficult for the principal to verify these standards and set optimal incentives. Manipulation could also occur when subjective measures of performance are used in incentive schemes. In this case, the agent may bias the measurement process, thus earning the favor of the superior (see Prendergast, 1999). Consequently, potential manipulability should be negatively correlated with incentive intensity. Table 1 summarizes the effects of all the factors described on the intensity of incentives.

[TABLE 1 HERE]

The Organizational Level of Performance Measurement

A necessary condition for the use of pay for performance is the existence of measured performance (see Prendergast, 1999). Therefore, one of the first issues organizations face when designing compensation schemes concerns the measurement level (see Baker, 2002). In some cases, output is easily measured in individual terms, which enhances the use of schemes that reward individual performance. When performance measurement at the individual level is difficult or costly, collective measures may be used. As Zenger and Marshall (2000) point out, the adoption of collective incentive schemes allows organizations to provide high-powered incentives when individual schemes are difficult to use. Collective incentives could help to mitigate potential problems such as distorted behavior by the worker. In addition, the combination of individual and collective incentives may bring the design of incentives closer to its optimal configuration in a firm.

In this paper, we consider the use of the three levels at which performance can be measured: *Individual PfP*, *Group PfP*, and *Plant or Firm PfP*. More precisely, we make an attempt to assess each of the three schemes in terms of the degree of *Noise*, *Controllable Risk*, *Impact on Firm Value*, *Distortion* and *Manipulability*. At this point, it is worth noticing that rating each incentive type in terms of the aforementioned factors is not an

7

easy task. However, this classification is easier to construct in relative terms, that is, by describing the properties of each incentive scheme in relation to the others.

Regarding *Noise*, we expect that it increases with the level of performance measurement. As Zenger and Marshall (2000) point out, it is more difficult for the employee to control performance measured at high organizational levels in relation to performance determined at low levels. In collective incentive schemes, performance depends on the actions taken by all members of the group. Furthermore, collective incentives are implemented in settings where teamwork and cooperation are required (see Jirjahn, 2002). These working environments are characterized by the existence of interdependencies between workers, a wider variety of tasks and, in general, more complex and uncertain work settings. Hence, due to the complexity existing in these environments, workers are subject to risk since their compensation depends on factors they are unable to control (see Baker, 2002). Overall, the degree of uncontrollable uncertainty should increase from *Individual PfP* (lower *Noise*) to *Plan or Firm PfP* (higher *Noise).*

As we have previously mentioned, *Controllable Risk* should foster incentive intensity. Our intuition is that the workers' scope of action when facing uncertainty is higher if performance is measured at the individual level. On the contrary, a measure of collective performance, either at group, plant or firm level, depends on the specific knowledge and private information of multiple agents. Consequently, the set of actions an individual worker can take to respond to uncertain situations should be lower in comparison to individual incentive schemes. To sum up, we expect the degree of *Controllable Risk* to be higher in individual incentive schemes than in group or plant schemes.

Another relevant factor in the evaluation of the intensity of optimal incentives is the *Impact on Firm Value*, i.e. the impact of measured performance on the firm's goals. If this relationship is weak, incentives should be low powered (Gibbons and Waldman, 1999). Regarding the measurement level, we presume that the relationship between measured performance and the contribution to firm value increases from *Individual PfP* to *Plant or Firm PfP*. Intuitively, when aspects such as the volume of production or its quality are measured at the plant or firm level, these measures are a good representation of firm value. On the contrary, measurements of volume or quality at the individual level have a lower impact on organizational value.

As regards *Distortion*, the use of group incentive schemes may help to mitigate dysfunctional behavior responses. In particular, collective incentives may reduce specific problems associated with the use of

individual measures of performance. For example, teamwork and cooperation frequently help to improve performance. However, individual rewards reduce incentives to cooperate (see Baker, 2002). This dysfunctional behavior could be alleviated if collective incentives are introduced. Similarly, the use of pay linked to plant or firm results may help to avoid distortions associated with the provision of individual incentives. Overall, our idea is that *Distortion* decreases from *Individual PfP* to *Plant or Firm PfP*.

Finally, we expect the problem of *Manipulability* of performance measures to vary with the level of measurement. This problem arises when workers take advantage of information asymmetries to increase their performance ratings without improving firm value. We anticipate that the ability of a worker to manipulate game a measure will decrease from individual incentives to collective schemes. For example, it is easier for a worker to manipulate volume standards when production is measured individually. However, if collective incentives are used, it becomes implausible that workers can game these standards, or manipulate any other dimension of performance. All these ideas concerning the determinants of incentive intensity and the measurement level are summarized in the following table.

[TABLE 2 HERE]

According to this table, *Individual PfP* presents good properties in terms of *Noise* and *Controllable Risk*. *Plant or Firm PfP* have a higher *Impact on Firm Value*, as well as low levels of *Distortion* and *Manipulability*. Finally, *Group PfP* is at the center of our ranking regarding the five properties. This characterization of PfP schemes makes it difficult to draw conclusions about their influence on intensity. Overall, *Plant or Firm PfP* display better properties than *Individual PfP*. However, it may also be true that employers give more importance to some characteristics (e.g. reducing noise problems) than others. The empirical analysis will shed more light on the relationship between the measurement level and incentive intensity.

Properties of Performance Measures

In this section, we characterize a set of performance measures in terms of their degree of *Noise*, *Controllable Risk*, *Impact on Firm Value*, *Distortion* and *Manipulability*. A performance measure is an indicator, either

quantitative or qualitative, that evaluates employees' outcomes. We group performance measures into four categories: *Results Category*, *HRM Outcomes Category*, *Subjective Category* and *Financial Category*. In order to construct this classification, we draw on the work by Heneman et al. (1999). According to these authors, results measures are in an intermediate position regarding both controllability and the link with the firm's value. In relation to these measures, HRM outcomes (in the terminology used by Heneman et al., behavioral measures), have a higher motivational value because workers can control them to a great extent. On the other hand, financial measures are more closely related to the firm's value, but workers have only limited control over them. In addition to these three factors we incorporate subjective indicators (*Subjective Category*) in our classification. We think this category is worth considering when grouping performance measures. On the one hand, subjective measures are widely used by organizations because they help to solve the deficiencies in objective measurement, such as the presence of noise (see Bol and Smith, 2011). In addition, subjective measures display different properties from the other categories, since they involve discretion and judgment by supervisors. Hence, they should be included in an independent category when performing the empirical analysis. The *Results Category* includes the following measures: *Productivity*, *Volume* and *Quality*. The *HRM Outcomes Category* considers *Absenteeism*, *Punctuality* and *Injuries*. The *Subjective Category* includes *Subjective Evaluation*, *Customer Satisfaction* and *External Evaluation*. The *Financial Category* comprises *Profitability* and *Cost Savings*.

The literature on performance measurement has considered other classifications besides the one we employ in this study. Among the most frequently used we find the distinction between financial and non-financial measures, that between broad and narrow measures or between input and output measures (for a review of different classification schemes see, for example, Ittner and Larcker, 2002). However, we consider that our choice of the four different categories mentioned is a more exhaustive representation of measurement options. Instead of focusing on comparing two broad measurement schemes (financial vs. non-financial, or input vs. output) we go a step further and compare the effects on incentive intensity of four different categories of performance dimensions. Hence, we avoid the limitations associated with the use of more vague classification schemes, such as the failure to account for characteristics like the specific activities (productivity, absenteeism, customer satisfaction, etc.) that the measure accounts for (see Ittner and Larcker, 2002). In what follows, we describe the properties of each category of measures.

Noise

Focusing on the first property, the *Results Category* provides some sources of *Noise* that should be considered when designing incentive schemes. Measures such as *Productivity*, *Volume* or *Quality* depend on uncontrollable factors to a certain extent. For example, they are influenced by technology. The productivity of an employee also depends on the results of colleagues or the decisions taken by superiors. On the other hand, the measures included in the *HRM Outcomes Category* may be beneficial to reduce potential sources of *Noise* in performance measurement. When incentives are based on aspects such as absenteeism, punctuality or injuries, for which a lower level of uncontrollable risk is involved, the problem of *Noise* becomes less prevalent.

Measures included in the *Subjective Category* are claimed to provide a comprehensive picture of performance (see Prendergast, 1999). Since they are based on assessments of multiple dimensions, these measures reduce *Noise* in the evaluation process. More precisely, one of the main reasons for using subjective measures is the risk associated with objective indicators (Baker, 1994). This idea is also tested by Bol and Smith (2011). By means of an experiment, these authors show that supervisors use subjective measures to compensate for the deficiencies in objective measurement and, in particular, to offset the negative effects caused by uncontrollable factors. Consequently, we conclude that the introduction of subjective evaluation mitigates *Noise* problems in incentive contracts.

Finally, measures belonging to the *Financial Category*, such as *Profitability* or *Cost Savings* are highly exposed to factors outside the worker's scope. As Bowens and Van Lent (2007) suggest, indicators such as profits are more dependent on exogenous factors than non-financial measures and, consequently, they are noisier. These authors state that "the primary function of disaggregated and nonfinancial measures is to reduce the noise in aggregated financial measures". For example, financial measures may be influenced by uncertainty in the environment, driven by factors such as regulatory decisions that affect competition, or by economic conditions. They also depend on the firm's decisions such as investment, on which the worker has little or no influence. The limited control of the worker hinders the motivational value of such measures (see Heneman et al. 1999), thus affecting the intensity of the optimal incentive scheme.

Controllable Risk

Regarding *Controllable Risk*, Heneman et al. (1999) place the *Results Category* in an intermediate position in comparison with other categories. In particular, it is claimed that workers have high control over measures such as *Productivity* or *Volume*. As Ittner and Larcker (2002) point out, non-financial measures such as *Productivity* "are better at signalling the actions workers can take to improve overall performance and at isolating the contribution of particular workers or activities". Consequently, workers are informed about how they should behave to improve performance, and they can exert control over the actions leading to incentive increases.

As Heneman et al. (1999) also describe, employees can easily control behavioral measures. In fact, these measures are more controllable by workers than the rest of the criteria, since they can determine with precision their outcomes regarding, for example, *Absenteeism* or *Punctuality*. Overall, when employees clearly understand how their actions influence compensation, the degree of *Controllable Risk* increases.

Subjective evaluation commonly accounts for different dimensions of a worker's job, ranging from collaboration with other employees to the relationship with clients (see Jirjahn and Poutsma, 2013). All of these aspects are controllable by workers, so one should expect the degree of *Controllable Risk* to be higher in relation to other measure categories. However, the *Subjective Category* involves judgement and discretion by the person who performs the appraisal. Hence, there are aspects of the process workers are unable to control. Overall, we consider the *Subjective Category* to display a high degree of *Controllable Risk*, not forgetting that they may be subject to certain factors not controlled by employees.

Regarding the *Financial Category*, production employees hardly exert any control over measures such as *Profitability* or *Cost Savings* (we only include production workers in the analysis). Other workers, such as top managers, have higher control over these measures. This characteristic reduces the degree of *Controllable Risk* displayed by financial measures in relation to the results, HRM outcomes and subjective indicators.

Impact on Firm Value

Another relevant factor concerning incentive intensity is the measure of *Impact on Firm Value*. Looking at the *Results Category*, their impact is high in comparison with other performance criteria. For example,

improvements in workers' productivity will eventually have an impact on organizational performance. However, the correlation between the two variables is not perfect, since overall performance is a global result that depends on many aspects (Heneman et al., 1999). The *Results Category* focuses on particular activities (such as *Volume* or *Quality*) and, consequently, they do not entirely account for all the dimensions relevant to determining global performance.

On the contrary, the contribution to firm's objectives is blurred when it comes to the *HRM Outcomes Category*. Despite the validity of these measures as instruments to motivate workers, they are only weakly related to organizational performance. HRM outcomes such as *Punctuality* or *Absenteeism* are very task-specific, so they only reflect a narrow set of workers' actions. Hence, their *Impact on Firm Value* should be lower in comparison with other categories of indicators that represent a broader set of activities (Ittner and Larcker, 2002).

When management uses evaluation systems based on the perceptions of a supervisor, customers, or external agencies, it is difficult to interpret their ratings in terms of organisational performance. For example, there is no direct link between subjective perceptions of collaboration between employees, or the quality of relations with customers, and organisational performance. However, on many occasions supervisors are able to provide assessments that represent firm value more accurately than certain objective measures. As Baker et al. (1994) point out, when objective criteria are difficult to determine, subjective evaluations may improve the measurement of workers' contribution to firm value, even if such subjective criteria are imperfect.

The *Financial Category* adds a broad set of variables and actions, accounting for a variety of dimensions of a worker's performance. Consequently, it represents organizational performance more accurately than measures of results or HRM outcomes, which are narrower and more task specific as we have already described. Therefore, any incentive scheme that uses the Financial Category links compensation to the firm's value with more precision than schemes based on other type of indicators (Gibbons and Waldman, 1999).

 Distortion

The presence of *Distortion*, that is, rewarding the wrong behavior, is also a relevant factor influencing incentive intensity (see Holmstron and Milgrom, 1991; Baker, 1992). The *Results Category* may enhance an undesired allocation of effort across tasks. More specifically, when workers perform various activities, they

may put more effort into the activity that most influences their compensation. This problem is likely to arise if measures of results are used in compensation contracts. For example, when the size of incentives depends on the number of units produced by workers, they may be temped to ignore *Quality*. However, it is worth noticing that the problem of *Distortion* could be avoided or mitigated if several indicators of results are combined, so that different dimensions of performance are rewarded.

The *HRM Outcomes Category* also focuses on the deployment of specific tasks or activities. Consequently, it can induce distortionary behavior, leading workers to put relevant actions aside. However, we infer that the degree of *Distortion* is lower in comparison to the use of the *Results Category*. In order to support this argument, we draw from Holsmtrom and Milgrom (1991), and assume total effort is constant and should be allocated to different tasks. When activities are substitutes (such as quality and quantity), agents have to choose how to distribute effort. However, this problem is less likely to arise if activities are complementary. Hence, workers may work on improving several HHRR indicators at the same time, such as *Punctuality* and *Absenteeism*. Then, *Distortion* is less likely to appear.

In order to solve or mitigate distortionary behavior, the theory on incentive provision suggests the use of subjective assessments substituting or complementing objective measures (see Baker et al., 1994). The *Subjective Category* allows managers to account for different dimensions of a worker's job, thus limiting the focus of workers on particular actions. Hence, their inclusion in incentive schemes could reduce the problem of *Distortion*. Similarly, the fact that the *Financial Category* does not reward particular tasks leaves less scope for *Distortion*. As we have already mentioned, financial indicators provide aggregate measures of performance. The idea is similar to the one presented for the *Subjective Category*. In both cases, measures represent a broad picture of workers' performance, which avoids workers emphasizing only particular tasks.

Manipulability

Since workers possess asymmetric information about their results, they can manipulate performance standards concerning production time or volume, and it might be difficult for managers to verify these standards. Hence, the employer has to take into account that the use of incentive schemes based on the *Results Category* is subject to manipulation. On the other hand, the measures included in the *HRM Outcomes Category* are based on particular tasks. Due to the specificity of these measures, they can be easily controlled

by a supervisor. Therefore, it is not likely that a worker can use asymmetric information to bias these dimensions of performance. Hence, we anticipate that *Manipulability* is not an important issue in indicators of HRM outcomes.

Despite the virtues of using subjective measures in incentive schemes, the adoption of these measures is not exempt from potential problems. Specifically, there is what Prendergast (1999) calls "rent-seeking activities". This problem concerns any actions workers can take to influence the ratings given by supervisors. In other words, workers who want to win the favor of their superiors could play with incentive schemes based on the *Subjective Category*. This idea supports the concept of potential manipulation by workers in subjective evaluation. Regarding the *Financial Category*, non-managerial workers, like blue-collar ones, have little specific potential to determine outcomes such as *Profitability* or *Cost Savings*. Consequently, we infer the degree of *Manipulability* of these measures is low in relation to other categories such as results and subjective measures.

Table 3 summarizes all the ideas presented below and describes the four categories of measures in terms of their degree of *Noise*, *Controllable Risk*, *Impact on Firm Value*, *Distortion* and *Manipulability*. In order to construct this table, and as we have previously mentioned, we focus on the comparison of each category of measures with the rest of groups. However, we think this procedure does not compromise the purposes of the study. On the contrary, our final objective is to compare the incentive intensity associated with the different alternatives of performance measurement available to employers.

[TABLE 3 HERE]

From the examination of this table, it is not easy to come to definitive conclusions about the relationship between the categories of measures we have defined and incentive intensity. Each category has its strengths and weaknesses in terms of the properties described in this study. However, one interesting conclusion can be drawn from the analysis developed in this section. With the exception of *Distortion*, the use of the *Results Category* does not display what the theory considers "bad" properties in terms of the factors that affect intensity. As we hypothesized, the levels of *Noise*, *Controllable Risk*, *Impact on Firm Value*, and *Manipulability* are all moderate for results measures. On the contrary, the *HRM Outcomes*, *Subjective* and

Financial Categories display one or more properties that, according to theory, should lead to a low incentive intensity. Then, it is possible that the use of the *Results Category* promotes incentive intensity. However, it is also true that other categories present better characteristics than result measures in terms of certain properties. For example, the *Financial Category* has, in relative terms, a high *Impact on Firm Value*, and it ranks low in terms of *Distortion* and *Manipulation*. The empirical analysis will shed light on these issues.

As a final remark, one should bear in mind that incentive schemes may include several measures of a particular category. For example, *Productivity* may be combined with *Volume* or *Quality* measures. This is a relevant issue when analyzing incentive intensity, because some of the problems of a category of measures might be solved if several indicators are jointly implemented. According to the Informativeness Principle, a compensation contract should include all measures that provide information about workers' performance. However, it has been shown that this principle is not observed in many firms (Prendergast, 1999; Raith, 2008). In relation to this idea, Zenger and Marshall (2000) hypothesize that increasing the number of indicators in group incentives generates problems for organizations, such as an inefficient allocation of workers' effort across several tasks. In order to avoid such problems, managers might decide to decrease incentive intensity as the number of indicators increases. However, the authors do not find empirical support for this idea.

Data and Variables

The data were collected through personal interviews with managers in Spanish manufacturing plants with 50 or more employees, and represent a unique source of information about a range of human resource practices in Spanish firms. Information was gathered at the plant level, as this is the unit at which decisions about the implementation of the HRM practices of interest are taken. Furthermore, we expected knowledge of the issues included in the questionnaire to be greater at plant level and, as a consequence, believed the data obtained would be more reliable.

The process of development of the database was as follows. Once we defined the objectives and scope of our study, we conducted a thorough examination of the literature related to our purpose in order to properly design the questionnaire. With the information gathered, members of the research group and the firm in charge of the fieldwork jointly drew up a first draft of the questionnaire. We pre-tested the questionnaire in

nine plants and then modified it in several ways to come up with the final version.

Most of the information on HRM refers exclusively to blue-collar workers, that is, workers involved directly in the production process. The reason for restricting the analysis to this category of employees is the existence of a range of different internal labor markets with different features within the same organization. Limiting the study to manual workers makes comparisons across establishments easier.

The data were drawn from personal interviews with one of the managers at the plant. We thought that questions should be addressed to the general manager or to the human resource manager. In practice we interviewed the human resource manager most frequently. The range of potential survey respondents comprised all Spanish manufacturing establishments that had 50 or more employees in 2005. After stratification by sector, size, and location, we used 2005 data from the Spanish Central Directory of Firms (Directorio Central de Empresas, DIRCE) of the Spanish National Statistics Institute (Instituto Nacional de Estadistica, INE) to make a random selection of workplaces.

The interviews with managers who agreed to respond to our questionnaire were performed by professionals with specialized training in computer-assisted telephone interviews (CATI). The establishments were first approached by letter or email, indicating the goals of the survey and including a copy of the questionnaire. We contacted a total of 2,933 establishments and this effort yielded 1,001 valid interviews. This final number of interviews matched expectations regarding the size of the data set, representing a response rate of 34.1%.

The data set contains exhaustive information about the use of PfP, which enables an in-depth analysis of the practice. Specifically, the questionnaire inquired about the use of general PfP, as well as about three particular schemes: *Individual PfP*, *Group PfP*, and *Firm or Plant PfP*. Regarding incentive intensity, the data informs on the proportion of total pay that depends on performance. In addition, it specifies the proportion of pay that depends on individual performance, on group performance, and on plant or firm performance. In additions, questionnaire respondents were asked about the use of 11 measures of performance: *Productivity, Volume, Quality, Absenteeism, Punctuality, Injuries, Subjective Evaluation, Customer Satisfaction, External Evaluation, Profitability and Cost Savings*. The database specifies whether each of these measures is used in *Individual PfP*, in *Group PfP* and in *Plant or Firm PfP*.

As we have already described, we construct a classification of measures consisting of four different categories. The *Results Category* includes measures of *Productivity*, *Volume* and *Quality*. The *HRM*

Outcomes Category refers to *Absenteeism*, *Punctuality* and *Injuries*. The *Subjective Category* considers *Subjective Evaluation, Customer Satisfaction* and *External Evaluation*. The *Financial Category* comprises *Profitability* and *Cost Savings*.

Finally, the analysis includes a set of four controls. The first is the variable *Size*, which is measured as the number of employees in the establishment (see Zenger and Marshall, 2000). This variable takes a mean value of 188.01 and a standard deviation of 467.87. The second one is *Age*, which represents the number of years the plant has been in operation (see Heywood and Jirjahn, 2014). The mean value of *Age* is 44.89, and its standard deviation equals 29.37. In the regression analysis, we include the logarithms of both the Size and Age variables to avoid the influence of extreme values. *Multinational* is a dichotomous variable that takes value one if the firm has foreign ownership (see Heywood and Jirjahn, 2014). It takes a mean value of 0.21, and a standard deviation of 0.41. Finally, *Trade Union* captures the influence of trade unions in the plant, and takes values from one (low influence) to five (high influence), (see Barth et al., 2008). The mean value of *Trade Union* is 2.91, and the standard deviation equals 1.15.

Results

Preliminary Analysis

Before performing the regression analysis, we examine the incidence of PfP schemes and performance measures in detail. Regarding the use of PfP schemes, 53.15 per cent of the plants in our sample report using PfP. As for the adoption of each particular scheme, *Individual PfP* is the most widespread (33.57 per cent of plants), followed by *Group PfP* (20.58 per cent) and *Plant or Firm PfP* (16.18 per cent).

Table 4 shows the incidence of performance measures. The first column represents the frequency of plants that adopt each measure for at least one incentive scheme (individual, group, or plant or firm). The other three columns include frequencies for each particular scheme. In the first column, we observe that *Productivity* is the most widespread measure in our sample (380 plants), followed by *Quality* (197 plants), *Volume* (128 plants), and *Absenteeism* (101 plants). At the lower end of the classification we find *Customer Satisfaction* (20 plants), *Cost Savings* (15 plants) and *External Evaluation* (12 plants), which are only implemented if *Plant or Firm PfP* is used.

18

[TABLE 4 HERE]

Regarding *Individual PfP* (column 2), the most common indicators of performance are *Productivity* (256) and *Quality* (121). Looking at the *HHRR Outcomes Category*, the use of *Absenteeism* predominates over the implementation of *Punctuality* and *Injuries*. Finally, *Subjective Evaluation* is adopted in 45 plants, that is, 13.68 per cent of establishments using *Individual PfP*. As far as *Group PfP* is concerned, the results reproduce the pattern described for individual incentives. *Productivity* and *Quality* are the preferred measures in this particular scheme, whereas *Punctuality* is the less widespread indicator. However, when we analyze the results for *Plant or Firm PfP*, some differences emerge in relation to the other two incentive schemes. This might be explained by the fact that there is a higher number of measures available to firms. Again, *Productivity* leads the ranking (around 50 per cent of plants), but it is used with a lower frequency compared to *Individual PfP* and *Group PfP*. The use of this measure is followed closely by *Profitability* (35 per cent of plants), the most common indicator of the *Financial Category*.

Table 5 displays the number of performance measures by incentive scheme. In the three schemes, around half of the plants in the sample use only one measure. Between 21 and 26 per cent of establishments implement two measures in their incentive schemes, and three measures are only adopted by approximately 19 per cent of plants. Eight per cent of employers make use of four different measures in *Plant or Firm PfP*, but this percentage is even lower in *Individual PfP* and *Group PfP*. The percentage of plants including five or more measures is almost negligible. Overall, the results in Table 5 show that, despite the variety of performance measures available, most employers base their PfP schemes on one or two measures only.

[TABLE 5 HERE]

Table 6 depicts the incidence of each category of measures. More precisely, the table shows the number and percentage of plants using at least one measure of the category. In the first column, we see that over 90 per cent of the sampled plants use the *Results Category*. *Individual PfP* and *Group PfP* schemes match this pattern. For *Plant or Firm PfP*, the percentage is lower but still noteworthy: 65 per cent of plants with this

scheme adopt at least one results measure. The second most popular category of measures in *Individual PfP* and *Group PfP* is the *HRM Outcomes Category*. But the frequencies are far below those observed in the first category: 22 per cent in individual incentives and 28 in group incentives. *Plant or Firm PfP* displays a different picture: the use of measures in the *Financial Category* is more widespread than the use of measures in the *HRM Outcomes Category*, being used in 38 per cent of the plants. For the three PfP schemes, the measures in the *Subjective Category* come last, although they are more frequent in *Plant or Firm PfP*.

[TABLE 6 HERE]

Table 7 displays the number of categories in our sample of establishments. We observe that, for all the incentive schemes, more than 60 per cent of plants use only one category of measures. This figure is considerably high for *Individual PfP*, with 77.51 per cent of plants using only one category of measures. Between 17 per cent (*Individual PfP*) and 28 per cent (*Plant of Firm PfP*) of plants use two categories of measures. The use of three categories is restricted to a small percentage of organizations, ranging from 5.5 per cent to 6.5 per cent. Finally, only three per cent of the total number of plants in the sample implement measures of the four categories.

[TABLE 7 HERE]

The number of performance measures per categories is shown in Table 8. In the *Results Category*, a significant proportion of establishments use more than one measure (approximately 45 per cent of plants). This figure is similar for the *HRM Outcomes Category*, whereas it decreases significantly when it comes to the *Subjective* and *Financial Categories*. Regarding the two latter categories, more than 80 per cent of establishments use only one performance measure.

[TABLE 8 HERE]

Regression Results

We now focus on examining the results concerning the intensity of incentives. PfP represents, on average, 18.49 per cent of total pay in our sample of plants. The intensity of PfP decreases from *Individual PfP* (17.36 percent of total pay) to *Group PfP* (16.08) and *Plant or Firm PfP* (11.86).

Table 9 depicts OLS estimations of the total intensity of PfP. The table presents five regression models, each including a different set of explanatory variables. The five models add the group of controls we have previously described: the size of the plant, its age, whether it belongs to a foreign company, and the influence of trade unions[2]. Model 1 focuses on the analysis of how the measurement level affects incentive intensity. The coefficients for both *Individual PfP* and *Group PfP* are positive and highly significant. Hence, according to our estimations, the adoption of any of these schemes promotes incentive intensity. The magnitude of the coefficient is slightly higher for *Group PfP*. On the contrary, the use of *Plant or Firm PfP* does not seem to exert any significant influence on intensity.

[TABLE 9 HERE]

Model 2 includes the 11 performance measures as explanatory variables in the estimated equation. *Productivity* and *Cost Savings* are the only measures with a positive and significant effect on the strength of incentives. On the other hand, there are four measures that display negative coefficients. In particular, two of the measures belonging to the *HRM Outcomes Category* exert a negative effect on PfP intensity: *Absenteeism* and *Injuries*. In addition, the use of *External Evaluation* as a subjective measure of performance also reduces intensity according to our estimations. Finally, *Profitability* emerges as a negative determinant of incentive intensity in our sample of plants.

In the third model, the regressors are the four categories of measures defined in the third section. These variables represent whether at least one measure of the category is used to determine performance. What does the empirical analysis reveal about the effect of the categories of indicators? Only the *Results Category* emerges as a significant predictor of incentive intensity, and it displays a positive sign. The other three categories of measures have negative coefficients, but they are not statistically significant. As we have already described, this outcome could be explained by the moderate properties of the *Results Category*. In

addition, the problems associated with the use of particular measures of the *Results Category* can be mitigated or eliminated if other indicators with better properties are added. As Table 8 reveals, a significant proportion of establishments use more than one measure of results, whereas this frequency is lower for other categories.

In Model 4, both the three PfP schemes and the 11 performance measures are included. Regarding the measurement level, this estimation reproduces the results of Model 1: a positive and significant effect of *Individual PfP* and *Group PfP* on incentive intensity, and no significant effect of *Plant or Firm PfP*. Regarding the performance measures, the findings are also similar to those obtained in the second model. The main differences concern *Productivity*, which is no longer a significant predictor of intensity, and *Subjective Evaluation*, which is now negatively correlated with the variable of interest. Finally, Model 5 includes the PfP schemes and the four categories of variables. For the first set of regressors, the results are no different than the ones described in the previous models. Regarding the second set, the *Results Category* has a positive incidence on incentive intensity. The magnitude of this effect is, however, smaller than the one obtained in Model 2. Again, the inclusion of *HHRR Outcomes*, *Subjective* or *Financial Categories* does not contribute to explaining the strength of PfP.

Table 10 presents OLS estimations of incentive intensity for *Individual PfP*, *Group PfP* and *Plant or Firm PfP*. For each scheme, we estimate two regression models with a different set of predictor variables. The first contains the 11 performance measures individually, whereas the second considers categories of measures. The first model offers evidence supporting the influence of several measures on *Individual PfP* intensity. In particular, intensity increases when *Productivity* is used as an indicator of workers' performance. On the contrary, it decreases when the incentive scheme is based on *Injuries* or *Subjective Evaluation*. According to Model 2, the intensity of *Individual PfP* increases if at least one measure of results is implemented. On the contrary, the use of *Subjective Evaluation* exerts the opposite effect on the magnitude of this PfP scheme.

[TABLE 10 HERE]

The results concerning *Group PfP* show that our sets of regressors are limited when attempting to explain the intensity of this incentive scheme. Model 3 shows that intensity is only related to the use of *Punctuality* and

Injuries. The first measure correlates positively with the dependent variable, whereas the second exerts a negative effect. The results of Model 4 are even poorer, since none of the three categories of measures for which there are observations seem to exert any influence on *Group PfP* intensity.

The last two models concern the intensity of *Plant or Firm PfP*. If we look at the influence of particular measures (Model 5), both *Absenteeism* and *Profitability* emerge as negative determinants of intensity. In contrast, *Cost Savings* promotes the strength of this scheme of incentives, with a coefficient of significant magnitude. Finally, Model 6 illustrates the positive impact of the *Results Category* on intensity, matching the results obtained for total intensity as well as *Individual PfP* intensity. But in contrast with these models, the use of at least one measure of the *HRM Outcomes* category is negatively and significantly related to *Plant or Firm PfP intensity*.

Conclusions

In this study, we have examined the impact of the measurement level and the measures used to determine performance on PfP intensity. In order to do so, we have analyzed both the three measurement levels and the four measure categories in terms of the five properties that the literature on incentives identifies as important determinants of intensity. The study makes use of an exhaustive data-set that contains information at the plant level about the use of 11 performance measures in three PfP schemes: *Individual PfP*, *Group PfP*, and *Plant or Firm PfP*. In addition, we have information on the total intensity of PfP in each plant, as well as on the intensity of the three PfP schemes. To our knowledge, our data-set constitutes a unique source of information about PfP intensity and the use of performance measures at the three broad levels at which PfP can be implemented.

The analysis of the data reveals interesting patterns of performance measurement by employers. Results measures are, by far, the most widely used in our sample of firms. More specifically, a high number of PfP schemes are based on indicators of *Productivity* and *Quality*. *Profitability* is also a popular measure in *Plant or Firm PfP*. Due to the scarcity of similar data, it is difficult to compare these results with those of other studies. For example, from the analysis of the Consortium for Alternative Reward Strategies Research data (see McAdams and Hawk 1994), Ittner and Larcker (2002) we find that 52 per cent of firms use accounting measures, 48 per cent use quality measures, and only 25 cent adopt productivity measures (each plant in the

sample may use more than one performance measure). Using a sample of Finnish companies, Kauhanen and Napari (2012) observe that profitability is the predominant measure for blue collar workers, followed by quality and productivity. However, it is noteworthy that these studies refer to plants in various industries, and consider a mixture of schemes that cover varied organizational units, from the entire company to subsidiaries, divisions, departments, small teams, etc.

In addition, our data reveals that *Individual PfP* is the most common scheme, followed by *Group PfP* and, finally, *Plant or Firm PfP*. Around fifty per cent of firms use only one measure of performance in their PfP schemes, and the adoption of more than three measures is very infrequent (around 10 per cent of firms). Hence, despite the variety of indicators available, employers prefer to focus on very few criteria when designing their incentive schemes. As we have already mentioned, the most popular measures are those that belong to the *Results Category*. These findings suggest that the Informativeness Principle, the idea that any measure that adds information to incentive schemes should be used, is not supported by our analysis. Other studies have drawn different conclusions. For example, Kauhanen and Napari (2012) find that incentive schemes for blue-collar workers use, on average, four performance measures.

According to our estimation, the scheme that has the greatest impact on intensity is the one that does not present significant problems regarding any of the properties considered, that is, *Group PfP*. The adoption of *Individual PfP* also promotes incentive intensity. On the contrary, there is no significant effect of *Plant or Firm PfP* use on the intensity of incentives. These results suggest that, when designing incentive schemes, managers give more weight to those schemes that reduce *Noise* and display high levels of *Controllable Risk* (*Individual PfP*) versus those schemes generating *Noise* and reducing *Controllable Risk* (*Plant or Firm PfP*). According to theory, the effectiveness of collective PfP schemes depends on their intensity (see Zenger and Marshall, 2000). In our sample, the adoption of *Plant or Firm PfP* does not seem to be linked to the use of high powered incentives, so this may limit their effectiveness for increasing workers' efforts, attracting more able workers and, consequently, enhancing organizational performance. This result is also relevant because *Plant or Firm PfP* display good properties in terms of their *Impact on Firm Value*, *Distortion* and *Manipulability*.

With the exception of *Group PfP*, the use of at least one measure of the *Results Category* promotes incentive intensity in all the schemes analyzed. The remaining categories of measures are either not significant or have

a negative effect on the magnitude of incentives. As we have already described, the *Results Category* displays moderate levels of *Noise*, *Controllable Risk*, *Impact on Firm Value*, and *Manipulability*. On the contrary, the *HRM Outcomes*, *Subjective* and *Financial Categories* present problems regarding some of the properties. A plausible interpretation of our findings is that employers choose performance measures that display, on average, good characteristics when using high-powered incentive schemes. These results are related to those obtained in Gibbs et al. (2009), who conclude that "the more than a measure is flawed along any of these dimensions (noise, controllable risk, distortion and manipulation), the less weight is given to that measure for explicit incentives".

Our study reflects certain caveats such as those inherent to the use of cross-section survey data. In addition, since most theoretical insights refer to the weights given to performance measures, it would be interesting to develop data sets containing such information. In this study, as in Kauhanen and Napari (2012), Hwang et al. (2009), and Ittner and Larcker (2002), we only observe whether a measure is used or not in an incentive scheme. Another limitation of this work is that we have developed our own characterization of measurement levels and categories of measures, but we cannot directly determine their properties. Despite these limitations, our results support the idea that the intensity of PfP is significantly related to the way performance is measured. We hope this work serves to launch further empirical research on the topic, and to develop new theoretical insights on the relationship between performance measurement and incentive intensity.

Notes

1. For a more detailed description of the Informativeness Principle, see Holmstrom (1979).

2. The results of the control variables are not displayed in the table, but they are available from the authors upon request.

References

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

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

Baker, George P., Robert Gibbons and Kevin J. Murphy. 1994. "Subjective Performance Measures in Optimal Incentive Contracts." *Quarterly Journal of Economics* 109: 881-919.

Barth, Erling, Bernt Bratsberg, Torbjørn Hægeland, and Oddbjørn Raaum. 2008. "Who Pays for Performance?" *International Journal of Manpower* 29(1): 8-29.

Bol, Jasmijn C., and Steven D. Smith. 2011. "Spillover Effects in Subjective Performance Evaluation: Bias and the Asymmetric Influence of Controllability." *Accounting Review* 86(4): 1213-1230.

Bowens, Jans and Laurence Van Lent. 2006. "Assessing the Performance of Business Unit Managers." *Journal of Accounting Research* 45(4): 667-697.

Gibbs, Michael, Kenneth Merchant, Wim Van der Stede, and Mark Vargus. 2009. "Performance Measure Properties and Incentive Plan Design." Industrial Relations 48(2): 237-264.

Gibbons, Robert, and Michael Waldman. 1999. "Careers in Organizations: Theory and Evidence." In *Handbook of Labor Economics*, edited by Orley Ashenfelter and David Card. North Holland: Elsevier Science.

Heneman, Robert L., Gerold E. Ledford Jr., and Maria T. Gresham. 1999. "The Effects of Changes in the Nature of Work on Compensation." In *Compensation in Organizations. Current Research and Practice*, edited by Sara L. Rynes and Barry Gerhart. San Francisco: Jossey Bass.

Holmstrom, Bengt. 1979. "Moral Hazard and Observability." *The Bell Journal of Economics* 10: 74-91.

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

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

Heywood, John S., and Uwe Jirjahn. 2014. "Variable Pay, Industrial Relations and Foreign Ownership:

Evidence from Germany." *Industrial Relations* 52(3): 521-522.

Ittner, Christopher D., and David F. Larcker. 2002. "Determinants of Performance Measure Choices in Worker Incentive Plans." *Journal of Labor Economics* 20(2): 58-90.

Jirjahn, Uwe. 2002. "The German Experience with Performance Pay." In *Paying for Performance: An International Comparison,* edited by Michelle Brown and John S. Heywood. Armonk, NY: M.E. Sharpe Publishers.

Jirjahn, Uwe, and Erik Poutsma. 2013. "The Use of Performance Appraisal Systems: Evidence from Dutch Establishment Data." *Industrial Relations* 52(4): 801-827.

Kauhanen, Antti and Napari, Sami. 2012. "Performance Measurement and Incentive Plans." *Industrial Relations: A Journal of Economy and Society* 51(3): 645-669, 2012.

McAdams, Jerry L., and Elizabeth J. Hawk. 1994. *Organizational Performance and Rewards: 663 Experiences in Making the Link*. Phoenix: American Compensation Association.

Ortin-Angel, Pedro, and Vicente Salas-Fumas. 1998. "Agency-Theory and Internal Labor Market Explanations of Bonus Payments: Empirical Evidence from Spanish Firms." *Journal of Economics and Management Strategy* 7(4): 573–613.

Prendergast, Canice. 1999. "The Provision of Incentives in Firms." *Journal of Economic Literature* 37(1): 7-63.

Raith, Michael. 2008. "Specific Knowledge and Performance Measurement." *The RAND Journal of Economics* 34(4): 1059-1079.

Zenger, Todd R. and C. R. Marshall. 2000. "Determinants of Incentive Intensity in Group-Based Rewards." *Academy of Management Journal* 43(2): 149-163.

Table 1. Measure Properties and Incentive Intensity

Property	Incentive Intensity
Noise	-
Controllable Risk	+
Impact on Firm Value	+
Distortion	-
Manipulability	-

Table 2. Measure Properties and Measurement Level

Property	Level		
	Individual	*Group*	*Plant/Firm*
Noise	Low	Medium	High
Controllable Risk	High	Medium	Low
Impact on Firm Value	Low	Medium	High
Distortion	High	Medium	Low
Manipulability	High	Medium	Low

Table 3. Properties of Categories of Measures

Property	Categories			
	Results Category *Productivity* *Volume* *Quality*	**HRM Outcomes Category** *Absenteeism* *Punctuality* *Injuries*	**Subjective Category** *Subjective Evaluation* *Customer Satisfaction*[+] *External Evaluation*[+]	**Financial Category**[+] *Profitability* *Cost savings*
Noise	Medium	Low	Low	High
Controllable Risk	Medium	High	Medium	Low
Impact on Firm Value	Medium	Low	Medium	High
Distortion	High	Medium	Low	Low
Manipulability	Medium	Low	High	Low

NOTE: [+]*Only used in plant or firm incentive programs*

Table 4. Use of Performance Measures

Performance Measure	Plants with any PfP scheme	Plants with Individual PfP	Plants with Group PfP	Plants with Plant or Firm PfP
Productivity	380 78.03%	256 77.81%	159 78.71%	82 52.90%
Volume	128 26.28%	67 20.36%	51 25.25%	35 22.58%
Quality	197 40.45%	121 36.78%	76 37.62%	44 28.39%
Absenteeism	101 20.74%	56 17.02%	45 22.28%	29 18.71%
Punctuality	59 12.11%	38 11.55%	20 9.90%	15 9.68%
Injuries	48 9.86%	21 6.38%	21 10.40%	23 14.84%
Subjective	68 13.96%	45 13.68%	28 13.82%	17 10.97%
Customer Satisfaction[+]	20 4.59%	n.a.	n.a.	20 12.90%
External Evaluation[+]	12 2.75%	n.a.	n.a.	12 7.74%
Profitability[+]	54 12.39%	n.a.	n.a.	54 34.84%
Cost Savings[+]	15 3.44%	n.a.	n.a.	15 9.68%
Number of Observations	487	329	202	155

NOTE: [+]*Only used in plant or firm incentive programs*

Table 5. Number of Performance Measures

Number of Performance Measures	Plants with Individual PfP	Plants with Group PfP	Plants with Plant or Firm PfP
1	186 56.53%	98 48.51%	80 51.61%
2	74 22.49%	53 26.24%	33 21.29%
3	40 12.16%	32 15.84%	16 10.32%
4	14 4.26%	7 3.47%	13 8.39%
5	5 1.52%	4 1.98%	3 1.94%
6	1 0.30%	4 1.98%	3 1.94%
7	9 2.74%	4 1.98%	0 0.00%
8	n.a.	n.a.	1 0.65%
9	n.a.	n.a.	1 0.65%
10	n.a.	n.a.	5 3.23%
11	n.a.	n.a.	0 0.00%
Number of Observations	329	202	155

Table 6. Categories of Performance Measures

Categories	Plants with any PfP scheme	Plants with Individual PfP	Plants with Group PfP	Plants with Plant or Firm PfP
Results Category	448 91.99%	303 92.10%	191 94.55%	101 65.16%
HRM Outcomes Category	131 26.90%	73 22.19%	57 28.22%	44 28.39%
Subjective Category	87 17.86%	45 13.68%	28 13.86%	39 25.16%
Financial Category	59 11.92%	n.a.	n.a.	59 38.06%
Number of Observations	487	329	202	155

NOTE: The *Results Category* includes *Productivity*, *Volume* and *Quality*; the *HRM Outcomes Category* includes *Absenteeism*, *Punctuality* and *Injuries*; the *Subjective Category* includes *Subjective Evaluation*, *Customer Satisfaction* and *External Evaluation*; the *Financial Category* includes *Profitability* and *Cost Savings*.

Table 7. Number of Categories

Number of Categories	Plants with any PfP scheme	Plants with Individual PfP	Plants with Group PfP	Plants with Plant or Firm PfP
One Category	321 65.91%	255 77.51%	140 69.31%	93 60.00%
Two Categories	108 22.18%	56 17.02%	50 24.75%	44 28.39%
Three Categories	44 9.03%	18 5.47%	12 5.94%	10 6.45%
Four Categories	14 2.87%	n.a.	n.a.	8 5.16%
Number of Observations	487	329	202	155

Table 8. Number of Performance Measures per Categories

Number of Measures	*Results Category*	*HRM Outcomes Category*	*Subjective Category*	*Financial Category*
1	248 55.36%	74 56.49%	62 83.78%	49 83.05%
2	143 31.92%	37 28.24%	11 14.86%	10 16.95%
3	57 12.72%	20 15.27%	1 1.35%	n.a.
Number of Observations	448	131	74	59

NOTE: The *Results Category* includes *Productivity*, *Volume* and *Quality*; the *HRM Outcomes Category* includes *Absenteeism*, *Punctuality* and *Injuries*; the *Subjective Category* includes *Subjective Evaluation*, *Customer Satisfaction* and *External Evaluation*; the *Financial Category* includes *Profitability* and *Cost Savings*.

Table 9. Determinants of PfP Intensity, OLS Regressions

Variable	Model 1	Model 2	Model 3	Model 4	Model 5
Constant	14.453**	16.838*	12.438*	11.889	10.119
	(6.975)	(1.577)	(6.994)	(7.310)	(7.386)
Individual PfP	4.723***	-	-	5.578***	4.873***
	(1.701)			(1.800)	(1.701)
Group PfP	5.482***	-	-	6.494***	5.307***
	(1.589)			(1.950)	(1.660)
Plant or Firm PfP	-1.220	-	-	1.393	.226
	(1.670)			(2.071)	(2.067)
Productivity	-	2.888*	-	1.430	-
		(1.577)		(1.559)	
Volume	-	.447	-	0.310	-
		(1.541)		(1.557)	
Quality	-	2.250	-	2.044	-
		(1.563)		(1.542)	
Absenteeism	-	-3.403**	-	-3.546**	-
		(1.640)		(1.709)	
Punctuality	-	3.093	-	3.476	-
		(2.382)		(2.388)	
Injuries	-	-5.185***	-	-6.196***	-
		(1.191)		(1.969)	
Subjective	-	-1.389	-	-3.776**	-
		(2.013)		(1.907)	
Customer Satisfaction	-	-.307	-	.829	-
		(3.591)		(3.980)	
External Evaluation	-	-10.532***	-	-10.699***	-
		(3.993)		(3.254)	
Profitability	-	-3.386**	-	-3.685*	-
		(1.670)		(2.120)	
Cost Savings	-	15.303**	-	16.304**	-
		(6.639)		(6.923)	
Results Category	-	-	8.132***	-	4.910**
			(2.040)		(1.701)
HRM Outcomes Category	-	-	-1.686	-	-1.874
			(1.662)		(1.680)
Subjective Category	-	-	-.605	-	-2.077
			(1.811)		(1.786)
Financial Category	-	-	-1.199	-	-.412
			(2.082)		(2.456)
Controls	Yes	Yes	Yes	Yes	Yes
R^2	0.06	0.11	0.06	0.14	0.08
F	7.80***	4.67***	5.41***	5.50***	6.37***
N	456	394	449	394	447

NOTES:

***$p < 0.01$, **$p < 0.05$, *$p < 0.10$.

The *Results Category* includes *Productivity*, *Volume* and *Quality*; the *HRM Outcomes Category* includes *Absenteeism*, *Punctuality* and *Injuries*; the *Subjective Category* includes *Subjective Evaluation*, *Customer Satisfaction* and *External Evaluation*; the *Financial Category* includes *Profitability* and *Cost Savings*.

Table 10. Determinants of PfP Intensity by Scheme, OLS Regressions

Variable	Individual PfP		Group PfP		Plant or Firm PfP	
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Constant	9.620	9.272	30.804***	30.804**	9.913	4.562
	(7.607)	(7.419)	(10.103)	(12.118)	(9.219)	(9.194)
Productivity	3.261*	-	-1.428	-	-.175	-
	(1.907)		(2.372)		(2.877)	
Volume	3.383	-	.116	-	.090	-
	(2.620)		(2.244)		(2.378)	
Quality	2.714	-	2.892	-	2.467	-
	(1.701)		(2.160)		(3.688)	
Absenteeism	-.574	-	-.390	-	-4.565*	-
	(2.578)		(3.083)		(2.727)	
Punctuality	2.857	-	9.142*	-	2.252	-
	(3.412)		(4.707)		(3.148)	
Injuries	-7.349**	-	-6.903**	-	-2.143	-
	(2.837)		(3.395)		(2.702)	
Subjective	-4.360**	-	-3.809	-	-.083	-
	(2.085)		(3.207)		(3.334)	
Customer Satisfaction	-	-	-	-	-3.435	-
					(3.496)	
External Evaluation	-	-	-	-	-3.553	-
					(4.349)	
Profitability	-	-	-	-	-3.709*	-
					(2.213)	
Cost Savings	-	-	-	-	11.999**	-
					(6.479)	
Results Category	-	4.741***	-	.234	-	3.973*
		(1.682)		(5.219)		(2.307)
HRM Outcomes Category	-	-.284	-	-.628	-	-3.471*
		(2.177)		(2.651)		(1.990)
Subjective Category	-	-3.616**	-	-1.985	-	1.050
		(1.739)		(2.749)		(2.227)
Financial Category	-	-	-	-	-	-.491
						(2.183)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.10	0.07	0.08	0.03	0.13	0.08
F	4.95***	7.00***	1.91**	1.07	1.27	2.00*
N	292	292	181	181	137	137

NOTES:
***$p < 0.01$, **$p < 0.05$, *$p < 0.10$.
The *Results Category* includes *Productivity*, *Volume* and *Quality*; the *HRM Outcomes Category* includes *Absenteeism*, *Punctuality* and *Injuries*; the *Subjective Category* includes *Subjective Evaluation*, *Customer Satisfaction* and *External Evaluation*; the *Financial Category* includes *Profitability* and *Cost Savings*.