Universität Trier

The Use of Performance Appraisal Systems: Evidence from Dutch Establishment Data

Uwe Jirjahn
Erik Poutsma

Research Papers in Economics
No. 2/11

The Use of Performance Appraisal Systems:

Evidence from Dutch Establishment Data

Uwe Jirjahn

Professor of Economics, University of Trier, Germany

Erik Poutsma

Professor of Business Administration, University of Nijmegen, Netherlands

Abstract: Using Dutch data, we examine the characteristics of establishments that operate performance appraisal systems. Our estimates provide a couple of interesting results: (1) Several indicators of multitasking are positively associated with the use of performance appraisal systems. There is also a positive relationship between teamwork and performance appraisal. (2) While a shared corporate culture, the sharing of information and a professional personnel management emerge as positive determinants, unions appear to play a negative role. (3) Employers fostering career development and promotion opportunities make greater use of performance appraisal. (4) Public sector affiliation, establishment size and establishment age also play a role.

JEL: J33, J50, M52.

Keywords: Performance Evaluation, Multitasking, Teams, Trust, Internal Labor Market.

Corresponding Author: Uwe Jirjahn, Universität Trier, Fachbereich IV, Lehrstuhl für Arbeitsmarktökonomik, Universitätsring 15, 54286 Trier, Email: jirjahn@uni-trier.de.

1. Introduction

Since the emergence of the economics of personnel, explorations on the determinants of performance pay schemes have been increasingly common. However, econometric studies have mainly focused on incentive schemes such as piece rates or profit sharing. Little attention has been paid to the determinants of incentives that are based on performance appraisal.[1] This is surprising as performance appraisal systems appear to be much more widely used than piece rates or profit sharing.[2] Performance appraisal reflects the employer's need for a comprehensive measurement of workers' performance. Such comprehensive measurement of performance typically involves evaluations that are based on judgments and opinions of superiors, peers or even subordinates (Jackson and Schuler 2003, Lewin and Mitchell 1995, Murphy and Cleveland 1991).

In this study, we use establishment data from the Netherlands to examine the characteristics of employers which operate performance appraisal systems. The estimates show that several indicators of multitasking are positively associated with the use of performance appraisal systems. This conforms to theoretical expectations. Complex and multifaceted tasks are often associated with dimensions of performance for which objective performance measures are not available. Hence, the employer has to rely on subjective evaluations in order to assess individual employee performance. However, the estimates also indicate that the positive relationship between multitasking and performance appraisal holds specifically for given jobs while frequent changes of jobs appear to limit the intensity of using performance appraisal.

Importantly, the variables for multitasking also capture teamwork. The estimates reveal that teamwork increases the probability of a high intensity of performance

appraisal. This result is of particular interest as there is a wide held view that teamwork makes it difficult to identify individual employee performance. Our finding suggests that employers respond to this difficulty with subjective evaluations of performance. Teamwork may even increase the need for monitoring individual employee behavior because of a high degree of interdependent worker productivity. The high degree of production interdependencies implies that the performance of a single team member does not only affect his or her own output but also the output of the entire team.

However, performance evaluations entail a substantial degree of subjectivity and discretion. This suggests that incentives based on performance appraisal are more effective if employees can trust that the employer will provide fair evaluations of their performance. Hence, circumstances contributing to trust and cooperation within establishments should foster the use of performance appraisal systems as they increase the incentive effects of performance evaluations. This hypothesis is supported by our empirical findings. Information sharing and reducing hierarchical barriers by a shared corporate culture are positive determinants of the use of performance appraisal systems. A professional personnel management is also a positive covariate.

Furthermore, the estimates show that performance appraisal plays a role in the functioning of internal labor markets. Employers fostering career development and promotion opportunities make greater use of performance appraisal. Promotion decisions require a comprehensive assessment of the employees' skills and abilities. Our results suggest that performance appraisal allows such comprehensive evaluation.

Unions appear to limit the intensity of using performance appraisal. They may oppose the extensive use of performance appraisal if they fear that performance appraisal

increases wage inequality and, hence, reduce their bargaining power. Moreover, establishments in the public sector use performance appraisal less extensively. This conforms to theories stressing that public sector employers have ambiguous performance goals and, hence, face difficulties in systematically evaluating their employees' performance. Finally, our estimates suggest that the size and the age of the establishment have nonlinear influences on the use of performance appraisal systems.

Previous research has paid very little attention to the establishment characteristics that are associated with the use of performance appraisal systems. Two exceptions are the studies by Brown and Heywood (2005) for Australia and Addison and Belfield (2008) for Britain. Those studies obtain extremely mixed results with respect to the role of work organization, unionization, HRM policy, and establishment size. This clearly calls for further research. Our examination for the Netherlands provides a valuable data point to test whether the findings for Australia or the findings for Britain hold up in another country. The Dutch results are broadly in line with the findings by Brown and Heywood (2005) as they confirm a positive role of multifaceted jobs and a negative role of unions. Moreover, our study brings a new twist into the scarce research on the use of performance appraisal. First, we provide evidence on the role of establishment characteristics such as teamwork or corporate culture that have not been considered by the two previous studies. Our study also provides new evidence of a positive link between internal labor markets and performance appraisal. Second, we add to the method of estimation by applying a generalized ordered probit model.[3] Our critical dependent variable is an ordered variable indicating the intensity of using performance appraisal. Estimating the determinants of the intensity of use with the standard ordered probit model

assumes a single coefficient vector for all outcome categories. In contrast, the generalized model provides more flexibility as it allows different coefficient vectors. While a couple of results do not depend on the method used, our examination demonstrates that the influence of specific key variables such as teamwork can only be uncovered by the generalized model.

The rest of the paper is organized as follows. Section 2 gives a detailed discussion of the theoretical background and reviews the sparse empirical evidence. Section 3 describes data, variables and method. Section 4 presents the results while Section 5 concludes.

2. Background Discussion

The hypotheses guiding the empirical analysis are organized around four large themes: nature of production, cooperation and trust, internal labor markets, and general establishment characteristics.

2.1 Nature of Production

Workers' effort, motivation and commitment depend on the structure of incentives created within the firm. The provision of incentives for workers requires the measurement of worker performance. Performance measures may be either objective or subjective (Baker et al. 1988). Objective measures such as sales or the quantity of produced output involve a relatively low degree of discretion and can be easily verified. In contrast, subjective performance measures, such as performance evaluations by superiors (e.g., evaluations of a worker's cooperativeness), entail a substantial degree of discretion and often are not verifiable.

If production is characterized by rather simple tasks, the firm may primarily rely on objective performance measures. For example, the employer can use piece rates to reward the quantity of produced output. The measurement of performance gets more complicated if jobs involve multiple types of worker effort. To some extent the employer may use multiple objective performance measures to reward the different types of effort (Moers 2005). However, more complex and multifaceted tasks are often associated with dimensions of performance which hardly can be verified. If production is characterized by multitasking, the inability to reward every type of productive worker activity distorts the allocation of effort across tasks. An emphasis on performance as measured by one or a few objective indicators causes workers to cut back on productive behaviors for which they are not rewarded. Piece rates are a classic example. While piece rates induce workers to exert effort increasing quantity, they provide no incentives to engage in productive activities such as helping colleagues, maintaining equipment, cultivating customer goodwill, striving for quality, and reducing chances of workplace injury (Brown 1990, Drago and Garvey 1998, Freeman and Kleiner 2005). This problem may be seen as what Kerr (1975) called “the folly of rewarding A while hoping for B”.

There appear to be two alternatives to alleviate this problem. On the one hand, the employer may simply use low-powered incentives to avoid a distortion in the allocation of worker effort (Holmstrom and Milgrom 1991). On the other hand, the employer may use subjective performance appraisals by superiors (Gibbons 1998, Prendergast 1999). While subjective performance measures entail a substantial degree of discretion, they allow making comprehensive judgments about the performance of the individual workers. The more holistic assessment of performance could evaluate helping colleagues,

learning new skills, participating effectively in groups, or cultivating good relations with customers. This suggests that formal performance appraisal systems should be more likely to be used in firms where production is characterized by a larger degree of multitasking. However, the sparse econometric studies appear to be inconclusive. Brown and Heywood (2005) provide Australian evidence of a positive association between multifaceted jobs and the use of formal performance appraisal systems. In contrast, Addison and Belfield (2008) find no link between worker influence over task allocation and performance appraisal.

In our empirical analysis, multitasking is captured by several indicators. First, an ordered variable for the delegation of responsibilities to lower layers of hierarchy is included. While the delegation of decisions may help reduce decision overload at higher layers of hierarchy, it entails more complex tasks and greater influence over the allocation of effort across tasks at lower layers of hierarchy. Hence, the employer may use performance appraisal not only for managers at higher layers of hierarchy but also for employees at lower layers. This implies that the delegation of responsibilities should be specifically associated with the use of more comprehensive performance appraisal systems. Second, a dummy variable for a capital intensive production is included. Incentives focusing solely on the quantity of produced output may cause workers to abuse machinery or reduce maintenance. If those counterproductive behaviors specifically hurt firms with capital intensive production, capital intensity should lead a firm to use performance evaluations as alternative incentive devices in order to induce both the production of output and the careful use of machinery. Third, a variable for employer-provided further training captures multiskilling. As multitasking presupposes

that workers comprehend important elements of the production process, multiskilling is closely related to multitasking (Morita 2005).

Moreover, we include a variable indicating the intensity of using teams. Teamwork also involves increased responsibility, expanded involvement in decision making, more complex tasks, and greater influence over task allocation. Yet, most importantly, it is characterized by a specific feature, namely by a high degree of interdependent worker productivity. As the performance of a single worker does not only affect the worker's own output but also the output of the entire team, it is crucial to provide incentives to the individual team members to exert high effort (Jirjahn and Kraft 2010). However, from a theoretical point of view, the relationship between team production and the monitoring of individual worker performance is ambiguous. On the one hand, interdependent worker productivity may make it difficult for the firm to identify individual contributions (Alchian and Demsetz 1972). On the other hand, particularly because individual worker output cannot be measured objectively, the firm may use subjective appraisal by superiors to evaluate the behavior of the individual members of a team. Moreover, the firm may even use mutual monitoring of peers to individually evaluate each team member's behavior (Marx and Squintani 2009). Case studies for Japan and the U.S. find that performance appraisal indeed includes teamwork elements (Shibata 2002). An econometric study by Gibbs et al. (2004) shows that department managers in car dealerships are more likely to receive subjective bonuses when there are strong interdependencies between departments. Altogether, the sparse available evidence indicates a positive relationship between teamwork and the use of performance appraisal. Of course, if it is more difficult for supervisors to monitor

individual worker behavior or mutual peer evaluations are required, the firm has to spend substantial effort to implement a comprehensive performance appraisal system. Hence, we do not simply anticipate a positive link between team production and the use of performance appraisal. We rather anticipate a specific association between team production and a high intensity of using performance appraisal.

We note that it can be important to distinguish between multitasking within given jobs and frequent changes of jobs that may make monitoring workers' performance too costly. First, we include a variable for job rotation. If too frequent moves of workers between jobs make evaluating workers' performance difficult, there should be a negative influence on the employer's use of performance appraisal. Finally, two variables capture the extent to which the establishment engages in product development. The establishment's innovativeness can have conflicting influences on using performance appraisal. One the one hand, new products may require workers' flexibility and the switch from one task to another. This suggests a positive influence. On the other hand, innovation activities may involve frequent changes in production processes and jobs making the evaluation of worker performance more difficult. This tends to limit the use of performance appraisal.

2.2 Cooperation and Trust

While performance appraisal may help provide incentives for multitasking, the difficult process of subjectively evaluating employees' performance generates its own problems. Focusing on the psychometric properties of performance appraisals, the early research was concerned with validity, reliability and freedom of bias (Milkovich and Newman 2002). Researchers identified common errors made by superiors, such as halo effects (a

favorable overall rating based on outstanding performance in only a single duty), central tendency bias (rating all employees close to the scale midpoint irrespective of performance), and recency effects (placing too much emphasis on recent performance).

However, added to the possibility of honest mistakes is the possibility that performance measures will be strategically manipulated (Gibbons 1998, Prendergast 1999). Subjective evaluations of workers' contributions are not verifiable by outsiders such as a court. This potentially provides an incentive for the employer to renege on the promises made to the employees. The employer may misreport employees' performance to save on wages. If workers anticipate the possibility of strategic manipulation of performance evaluation, they will withhold effort and cooperation. Under some circumstances, repeated games and reputation concerns can induce the employer to behave honestly. However, those mechanisms are far from being perfect. First, they fail in situations in which the employer overly discounts the future loss of trust and cooperation. Second, workers may distrust the employer's evaluations if there are ambiguities and information asymmetries. If a worker can only imperfectly observe the value his contribution has to the employer, he may fear that the employer does not correctly report that value. This may specifically hold true if (implicit) incentive contracts are incomplete and allow a recalibration of the performance standards to adjust to changes in the economic environment which may be less observable to the worker.

Furthermore, the discretionary nature of performance appraisal provides opportunities for favoritism at the various levels of hierarchy. Superiors' prejudices and preferences toward subordinates may enter performance appraisal (Prendergast and Topel 1993, 1996). Indeed, Elvira and Town (2001) find a race bias in performance

evaluations.[4] White supervisors rate black subordinates lower, while black superiors rate white subordinates lower.[5] Moreover, superiors can use their discretionary power to reward only subordinates who provide private services or goods (Laffont 1990). These services include flattery or loyalty to the superiors' career concerns. Alternatively, the supervisor may rate all subordinates highly to demonstrate to those further up the hierarchy his or her outstanding managerial skills. A less productive superior may even favor unproductive employees to protect him- or herself from being replaced by productive employees (Friebel and Raith 2004). Finally, subordinates may strategically engage in influence activities that result in a positive evaluation but not necessarily in increased performance (Milgrom and Roberts 1988). Thus, in Prendergast's (1993) theory of 'yes men', superiors favor proposals from their subordinates that mirror their own opinions.

This long list of 'perverse incentives' (Lewin and Mitchell 1995) limits the productive effectiveness of performance appraisal.[6] The employer is likely to be reluctant to use a performance appraisal system until those problems have been reduced. In our empirical analysis we control for several circumstances potentially mitigating the counterproductive incentives and, hence, supporting the implementation of performance appraisal. First, we control for the extent management provides information about the firm's economic situation to the workforce. Opening the books reduces information asymmetries and, hence, fosters workers' trust that the employer will honestly evaluate their performance. Second, the employer's attempt to implement a shared corporate culture that reduces barriers between the various layers of hierarchy is taken into account. A shared corporate culture provides standards of fair behavior at work allowing workers

to more easily assess the behavior of managers and supervisors. As clear standards generate transparency, they reduce the employer's incentive to renege on promises made to the employees (Kreps 1990). The standards may also reduce the discretionary power of supervisors and, hence, limit favoritism. Third, a variable for the existence of a personnel department indicates a professional HRM policy. For example, superiors may be trained to avoid the potential biases associated with subjective performance evaluations.

The role of unions is also examined. On the one hand, unions may act as contract enforcers allowing firms to make promises that would otherwise not believed. They can facilitate communication and coordination among employees. To the extent coordinated actions result in a more severe punishment of employer opportunism, the employer's incentive to renege on implicit agreements is reduced (Hogan 2001). This suggests a positive role of unions in the use of performance appraisal systems. On the other hand, individual based incentives increase the dispersion of earnings within the firm. This in turn may undermine worker solidarity resulting in a weaker bargaining strength of the union. Hence, unions may oppose the implementation of a performance appraisal system. Empirical studies provide extremely mixed results. While Brown and Heywood (2005) find a negative link between unions and the use of performance appraisal for Australia, Addison and Belfield (2008) obtain a positive association for Britain. In our study for the Netherlands, the influence of unions is captured by the frequency the local works council of the establishment meets with unions.

2.3 Internal Labor Markets

The use of a performance appraisal system may also depend on how it fits into the firm's general HRM policy. Internal labor markets play an important role in the firm's HRM

activities with promotions being a key feature of these markets (Baker et al. 1994a, 1994b). Promotions serve different purposes (Baker et al. 1988, Prendergast 1999). On the one hand, promotions are used to sort employees on the basis of their abilities. Promotions take the form of job changes within the firm, in the sense that responsibilities increase with abilities. If employees differ in their abilities and jobs differ in the demand they place on skills and abilities, promotions are one way to match employees to the jobs for which they are best suited. On the other hand, promotions provide incentives for lower level employees as higher ranks in the firm are associated with higher wages.

The role performance appraisal plays in internal labor markets is ambiguous. Performance appraisal is often used to provide contemporaneous incentives (Brown and Heywood 2005). If contemporaneous performance pay and the delayed rewards associated with promotions are substitutes, performance appraisals may be less likely to be used in firms that rely on internal labor markets. However, promotion decisions require a comprehensive evaluation of the employees' abilities. To the extent objective performance measures do not allow such comprehensive evaluation, the employer has to use subjective measures (Gibbons 1998). This suggests that performance appraisal should be more likely to be used in firms placing high emphasis on internal promotion.

The scarce empirical evidence provides no clear answer. Brown and Heywood (2005) find for Australia no robust association between the extensive use of internal promotions and the implementation of a formal performance appraisal system. Moreover, Australian establishments with a high share of long-tenure employees appear to be less likely to implement such system. In contrast, using employee data from Germany, Grund and Sliwka (2009) find that performance appraisal plays an important role in promotions.

Given these mixed results, further evidence is certainly warranted. In our study, the importance of promotions is captured by a variable indicating the firm's effort to provide internal career development and promotion opportunities.

2.4 General Establishment Characteristics

General establishment characteristics are also likely to influence the use of performance appraisal systems. First, returns to scale may play a role (Brown and Medoff 1989). Implementing a performance appraisal system may involve a fixed cost, and the fixed cost per employee diminishes with the number of employees subject to performance appraisal. This, in turn, may increase the net benefits of a performance appraisal system. Hence, theoretical conjectures suggest a positive relationship between firm size and the use of performance appraisal. Yet, even the evidence on the role of firm size is mixed. The Australian study by Brown and Heywood (2005) confirms a positive link between firm size and performance appraisal systems while the British study by Addison and Belfield (2008) fails to find any significant association. In our study for the Netherlands, we include three dummy variables for establishment size.

Second, the age of the firm appears to play an important role in many dimensions of firm strategy and performance (Wagner 2006). However, findings by Thornbill and Amit (2003) suggest that the effect of firm age can be highly nonlinear. Younger firms are more likely to fail because of inadequacies in managerial knowledge. Older firms are more likely to fail because of the inability to adjust to environmental changes. Against this background, we hypothesize that firms of intermediate age are more likely to adopt a performance appraisal system. In very young firms, liability of newness may hinder the adoption. In very old firms, a more bureaucratic organization and established routines

may be obstacles to the implementation of a performance appraisal system. In our study, we include three dummy variables for establishment age.

Third, the use of performance appraisal may also differ between private and public sector firms. Theoretical work emphasizes that firms in the public sector deal with multiple principals and have unclear missions (Dixit 1997, Dewatripont et al. 1999). Hence, public sector firms have ambiguous performance goals. As a consequence, they face difficulties in using a coherent system of incentives (Burgess and Metcalfe 2000). Finally, we control for industry affiliation to account for basic differences in production technology and the economic environment.

3. Data, Descriptive Statistics and Method

The empirical investigation uses data from establishments in the Netherlands. With the exception of mining, agriculture and fishery all sectors of the Dutch economy are covered. In 1998, a questionnaire was sent to 1,765 works councils of a representative sample of establishments. According to Dutch law, works councils are mandatory in establishments with 35 or more employees. Works councils are a very valuable source of information as they provide a highly developed mechanism for establishment level participation (Visser 1995). Employers are obliged to provide works councils with comprehensive information on all relevant matters. Moreover, works councils have codetermination rights in social matters such as working hours, hiring and firing, training, job evaluation, method of pay, and promotions. They have consultation rights in matters such as investments, and expansion or reduction of business activities.

The questionnaire covered various aspects of establishment structure and establishment behavior with an emphasis on issues relating to personnel management and

industrial relations. 37 percent of the works councils participated in the survey. As the survey was based on voluntary participation, this response rate is not unusual. A non-response analysis indeed revealed no bias. The analysis was performed by surveying works councilors who did not respond to the survey. They were asked questions about the perceived influence of the works council, and the size, age and industry affiliation of the establishment. There were no significant differences between the response and the non-response sample.

After eliminating observations for which full information is not available, the empirical analysis is based on data from 542 establishments. *Table 1* and *Table 2* show the definitions of variables and their descriptive statistics. The critical dependent variable asks 'does the establishment conduct annual performance appraisals for all employees'. Importantly, the data provide also information on the intensity of use. 45 per cent of establishments report spending much time and energy on it and 39 per cent report spending some attention to it. That means that 84 per cent of establishments conduct annual performance appraisals.

Following previous studies, we start with a probit regression, which takes the dependent variable as a dichotomous indicator of the use of a performance appraisal system. Next, we estimate the determinants of intensity of use. Let y_i denote establishment *i*'s intensity of using performance appraisal (0 = *i* does little or nothing about it; 1 = *i* pays some attention to it; 2 = *i* spends much time and effort on it). First, we apply the widely used ordered probit model. The ordinal variable y_i is assumed to depend on a latent variable y_i^*:

$$y_i^* = \boldsymbol{\beta}' \mathbf{x}_i + \varepsilon_i, \tag{1}$$

where $\boldsymbol{x}_i$ is the vector of establishment characteristics, and $\boldsymbol{\beta}$ the vector of coefficients. The error term ε_i has a normal distribution with zero mean and a variance equal to one. Taking the latent variable into account, the intensity of use is

$$y_i = \begin{cases} 0 \text{ if } y_i^* \leq \mu_1, \\ 1 \text{ if } \mu_1 < y_i^* \leq \mu_2, \\ 2 \text{ if } y_i^* > \mu_2. \end{cases} \tag{2}$$

The threshold values μ_1 and μ_2 are estimated jointly with $\boldsymbol{\beta}$.[7] The probabilities are:

$$\Pr(y_i = 0) = \Phi(\mu_1 - \boldsymbol{\beta}' \boldsymbol{x}_i), \tag{3a}$$

$$\Pr(y_i = 1) = \Phi(\mu_2 - \boldsymbol{\beta}' \boldsymbol{x}_i) - \Phi(\mu_1 - \boldsymbol{\beta}' \boldsymbol{x}_i), \tag{3b}$$

$$\Pr(y_i = 2) = 1 - \Phi(\mu_2 - \boldsymbol{\beta}' \boldsymbol{x}_i), \tag{3c}$$

where $\Phi(.)$ is the standard normal distribution function. The marginal probability effects are:

$$\partial \Pr(y_i = 0)/\partial x_{ki} = -\phi(\mu_1 - \boldsymbol{\beta}' \boldsymbol{x}_i)\beta_k, \tag{4a}$$

$$\partial \Pr(y_i = 1)/\partial x_{ki} = \{\phi(\mu_1 - \boldsymbol{\beta}' \boldsymbol{x}_i) - \phi(\mu_2 - \boldsymbol{\beta}' \boldsymbol{x}_i)\}\beta_k, \tag{4b}$$

$$\partial \Pr(y_i = 2)/\partial x_{ki} = \phi(\mu_2 - \boldsymbol{\beta}' \boldsymbol{x}_i)\beta_k, \tag{4c}$$

where $\phi(.)$ is the density function, x_{ki} the establishment characteristic *k* of observation *i*, and β_k the coefficient. A potential shortcoming of the simple ordered probit model is that it has constant threshold values and only a single coefficient vector for all outcome categories of the dependent variable (single index assumption). Moreover, the marginal probability effects can change their sign only once as one moves from the smallest to the largest outcome. Fortunately, a generalized ordered probit model (generalized threshold model) stands as an alternative to the rather restrictive ordered probit model (Boes and

Winkelmann 2010).[8] The generalized model provides more flexibility as it doesn't treat the thresholds as constant but makes them dependent on the explanatory variables. As it allows for different coefficient vectors, the generalized model takes into account that the effects of the explanatory variables may vary with the categories of the dependent variable. Thus, it also allows for more flexibility in the estimation of the marginal probability effects than the simple ordered probit model.

The features of the generalized model are potentially very important in our context. For example, our hypotheses suggest that teamwork makes monitoring performance more difficult and, thus, requires substantial effort to evaluate workers' individual performance. This suggests that team production has no positive influence on a modest use of performance appraisal but only on an intensive use. Hence, in a further step, we estimate the generalized ordered probit model to obtain more flexibility. The basic idea of the model is to make the thresholds linear functions of the explanatory variables:

$$\mu_{ij} = \tilde{\mu}_j + \boldsymbol{\gamma}_j' \boldsymbol{x}_i \quad (j = 1, 2), \tag{5}$$

where $\tilde{\mu}_j$ is a constant term, $\boldsymbol{x}_i$ again the vector of establishment characteristics, and $\boldsymbol{\gamma}_j$ a vector of coefficients. Substituting μ_{ij} for μ_j in equations (3a) to (3c) yields:

$$\Pr(y_i = 0) = \Phi(\tilde{\mu}_1 - \boldsymbol{\beta}_1' \boldsymbol{x}_i), \tag{6a}$$

$$\Pr(y_i = 1) = \Phi(\tilde{\mu}_2 - \boldsymbol{\beta}_2' \boldsymbol{x}_i) - \Phi(\tilde{\mu}_1 - \boldsymbol{\beta}_1' \boldsymbol{x}_i), \tag{6b}$$

$$\Pr(y_i = 2) = 1 - \Phi(\tilde{\mu}_2 - \boldsymbol{\beta}_2' \boldsymbol{x}_i), \tag{6c}$$

where $\boldsymbol{\beta}_j = \boldsymbol{\beta} - \boldsymbol{\gamma}_j$ as we cannot identify $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}_j$ separately. Altogether, the coefficient vectors are now allowed to vary across outcomes and we have a specific vector $\boldsymbol{\beta}_j$ for

each threshold. The marginal probability effects are:

$$\partial \Pr(y_i = 0)/\partial x_{ki} = -\phi(\tilde{\mu}_1 - \boldsymbol{\beta}_1' \boldsymbol{x}_i)\beta_{1k}, \quad (7a)$$

$$\partial \Pr(y_i = 1)/\partial x_{ki} = \phi(\tilde{\mu}_1 - \boldsymbol{\beta}_1' \boldsymbol{x}_i)\beta_{1k} - \phi(\tilde{\mu}_2 - \boldsymbol{\beta}_2' \boldsymbol{x}_i)\beta_{2k}, \quad (7b)$$

$$\partial \Pr(y_i = 2)/\partial x_{ki} = \phi(\tilde{\mu}_2 - \boldsymbol{\beta}_2' \boldsymbol{x}_i)\beta_{2k}, \quad (7c)$$

where β_{1k} and β_{2k} are the coefficients on establishment characteristic *k*.

4. Empirical Results

4.1 Probit and Ordered Probit Estimates

Table 3 provides the initial probit and ordered probit estimates. The marginal effects calculated at the means of the explanatory variables are shown in *Table 4*. Many of the variables take statistically significant coefficients of the expected sign. Specifically, the estimates confirm that multitasking is positively associated with performance appraisal. Capital intensive production, employer provided further training, and the delegation of responsibilities to lower levels of hierarchy are positive determinants of both the use and the intensity of performance appraisal. The estimated effects are not only statistically but also economically significant. Capital intensive production is associated with a more than 18 percentage point higher probability of spending much time and energy on performance appraisal. The indicators reflecting training and the delegation of responsibilities are three-level categorical variables. An increase of one category in the establishment's effort to provide further training increases the probability of spending much time and energy on performance appraisal by almost 16 percentage points. A one point increase in the effort to delegate responsibility is associated with a more than 14 percentage point rise in that probability. These findings conform to the hypothesis that an increased complexity of

tasks requires a more comprehensive evaluation of workers' performance. However, team production does not emerge with a statistically significant coefficient in the initial estimates. Our theoretical discussion suggests that the influence of teams may vary with the categories of our dependent variable. Thus, the role of teams may remain obscured until the generalized ordered probit model has been applied.

Job rotation takes a negative coefficient that is statistically significant in the probit regression. This supports the notion that too frequent moves between jobs make evaluating workers' performance difficult. However, the two variables reflecting the establishment's innovativeness are positive covariates of the use of performance appraisal systems.

While the variable for a personnel department does not take a significant coefficient, the variables for information sharing and a shared corporate culture are statistically significant determinants of performance appraisal. Information sharing emerges with a significantly positive coefficient in the ordered probit regression. A one point increase in the effort to share information with the workforce is associated with a more than 8 percentage point rise in the probability of spending much time and effort on performance appraisal. The variable for a shared corporate culture is a significantly positive covariate of performance appraisal in both the probit and the ordered probit regression. An increase of one category in the establishment's effort to reduce hierarchical barriers by implementing a shared corporate culture is associated with a nearly 11 percentage point increase in the probability of using performance appraisal very intensively. These results confirm the notion that reducing information asymmetries and implementing shared values of fair behavior strengthen the incentive effects of

performance appraisal by creating trust and cooperation. This in turn fosters the employer's use and the intensity of performance appraisal.

The two indicators on the influence of unions are significantly negative determinants of performance appraisal in the ordered probit regression. Our hypotheses discussed earlier suggest that there may be two opposing effects of unions. On the one hand, unions may act as contract enforcers creating the trust and cooperation that helps implementing performance appraisal systems. On the other hand, unions may oppose the implementation of performance appraisal systems if they fear that those systems increase wage inequality and undermine worker solidarity. The empirical results suggest that the latter effect dominates.

The variable for promotions takes a significantly positive coefficient in the ordered probit regression. A one point increase in the establishment's effort to provide internal career development and promotion opportunities increases the probability of spending much time and energy on performance appraisal by more than 8 percentage points. The positive association between promotions and performance appraisal supports the hypothesis that career development and promotion decisions require a comprehensive evaluation of the employees' skills and abilities.

Turning to the general establishment characteristics, the variable for public sector establishments takes a significantly negative coefficient in the ordered probit estimation. This finding is consistent with the hypothesis that establishments in the public sector have ambiguous performance goals and, hence, face difficulties in systematically evaluating their employees' performance. Furthermore, the estimates provide some evidence that an intermediate age of the establishment is positively associated with

performance appraisal. In the ordered probit regression, establishments with an age of 5 to 25 years have a higher probability of using performance appraisal very intensively. Younger establishments may have a lower probability of a very intensive use due to inadequacies in managerial knowledge. Older establishments may have a more bureaucratic organization and established routines making a very intensive use of performance appraisal more difficult. Interestingly, the estimates provide also some evidence that an intermediate establishment size is positively associated with performance appraisals. There may be two opposing effects of establishment size. On the one hand, returns to scale suggest a positive influence of size. On the other hand, very large establishments may be characterized by a high degree of division of labor and inflexibility limiting the use of performance appraisal.

4.2 Generalized Ordered Probit Estimates

Table 5 presents the generalized ordered probit estimates. The marginal effects calculated at the means of the explanatory variables are shown in *Table 6*. A likelihood ratio test can be used to compare the general model with the standard ordered probit regression. The test rejects the hypothesis of equal coefficients for all outcome categories of the dependent variable at the 5 percent level ($\chi^2 = 40.33$). Thus, the generalized ordered probit regression is the preferred model. While it confirms many of the significant relationships obtained by the standard model, the general model provides new evidence on several variables that did not emerge with significant coefficients in the simple ordered probit estimation.

Importantly, conforming to our theoretical expectations, the generalized ordered probit estimation shows the role of teamwork in a differentiated light. The first

coefficient on teams is statistically insignificant whereas the second one is significantly positive. The implied marginal effects show that an increase of one category in the intensity of teamwork is associated with a 3.3 percentage point higher probability of spending much time and energy on performance appraisals. While this marginal effect appears to be smaller than those for several of the other variables, it is nonetheless economically significant. For an establishment that would otherwise have the mean probability of 45.4, this would be a 7.3 percent increase in the probability of using performance appraisal very intensively. Altogether, the results confirm the hypothesis that there is a specific positive association between teamwork and a very intensive use of performance appraisal. While interdependent worker productivity makes monitoring workers' performance more difficult, it increases at the same time the necessity to monitor performance as each worker's effort does not only influence his or her own output but also the output of the entire team. As a consequence, the establishment is more likely to conduct performance appraisals with high intensity.

The generalized ordered probit model provides also new insights into the role of innovation activities. In the initial estimates, our two indicators of innovation activities did not emerge with significant coefficients in the standard ordered probit but only in the probit regression. The generalized ordered probit estimation shows that both variables for innovation activities are positively associated with an intermediate intensity of using performance appraisal. This may indicate two opposing effects. On the one hand, innovation activities require flexibility and multitasking. This should have a positive influence on the necessity of using performance appraisal. On the other hand, those activities may involve frequent changes in production processes and jobs making the

evaluation of worker performance more difficult. As a result, innovative establishments appear to choose an intermediate intensity of using performance appraisal. Furthermore, the generalized model qualifies the negative influence of job rotation. This influence appears to be more pronounced with respect to an intermediate intensity of use.

For the other variables reflecting the nature of production, the pattern of results is similar to that obtained by the standard ordered probit estimation. The delegation of responsibilities is positively associated with a very intensive use of performance appraisal. The effect is very sizable. An increase of one category in the establishment's effort to delegate responsibilities to lower levels of hierarchy increases the probability of using performance appraisal very intensively by nearly 18 percentage points. While both coefficients on employer provided further training take significantly positive coefficients, the marginal effects show that employer provided further training specifically increases the probability of spending much effort and energy on performance appraisal.[9] A one point increase in the establishment's effort to provide further training is associated with an almost 17 percentage point rise in that probability. The variable for capital intensive production emerges with a similar pattern of results. While both coefficients are significantly positive, the marginal effects imply a positive influence specifically on the very intensive use of performance appraisal. Establishments with capital intensive production have a 13 percentage point higher probability of using performance appraisal very intensively.

The general model also confirms the roles of internal labor markets, information sharing and a shared culture. If the employer increases the effort to provide promotion opportunities by one point, the probability of a very intensive use rises by 9 percentage

points. A one point increase in the employer's effort to share information is associated with a 13 percentage point rise in that probability. An increase in one point of the effort to implement a shared corporate culture increases the probability by 10 percentage points. Interestingly, the variable for a personnel department now emerges for the first time as a statistically significant determinant. Establishment with a personnel department have a nearly 12 percentage point higher probability of spending much time and effort on performance appraisal. This finding conforms to the notion that a professional personnel management contributes to an increased effectiveness of performance appraisals. This in turn stimulates the employer's incentive to spend much time and effort on it.

Furthermore, the estimates show that the influence of unions reduces the probability of a high intensity of use but not the probability of an intermediate intensity of use. Similarly, establishments in the public have a lower probability of using performance appraisal very intensively while the probability of using them with an intermediate intensity is not reduced. An establishment age of 5 to 15 years is positively associated with an intermediate intensity and an establishment age of 16 to 25 with a high intensity of use. Finally, there is a positive link between an intermediate size class and an intermediate intensity of use and a negative link between the largest size class and a very intensive use of performance appraisal.

5. Conclusions

Performance appraisal plays an important role from both a practical and a theoretical point of view. It is widely used by establishments and has two specific characteristics that deserve particular attention. On the one hand, performance appraisal entails a high degree of discretion and subjectivity. On the other hand, it allows a more comprehensive

measurement of individual worker performance. However, empirical research on the determinants of performance appraisal is very scarce.

This study contributes to the literature by examining the employer characteristics that are associated with the use of performance appraisal in the Netherlands. Our results show that employers use performance appraisal more intensively if production is characterized by a substantial degree of multitasking. Interestingly, there is also a positive relationship between teamwork and performance appraisal. Moreover, our estimates suggest that employers use performance appraisal to a greater extent under circumstances that contribute to trust and cooperation. Furthermore, internal labor markets are positively associated with performance appraisal while unions and public sector affiliation are negative determinants.

We end this article with suggestions for further research. First, our study shows that the influence of specific establishment characteristics can only be uncovered by applying a generalized ordered probit model. It would be interesting to apply this method to the British and Australian data that have been used in previous studies. Second, further studies on the determinants of performance appraisal systems are certainly warranted. However, it would be helpful not only to use data from further countries but also to use more than one data set for the same country. This would allow examining if mixed results are due to different institutional frameworks or rather due to different data sets. Third, our study, like the other two studies, has not examined the influence of performance appraisal on establishment performance. Examining this influence stands as important future research now that the determinants of performance appraisal have been examined. As stressed by Brown and Heywood (2005), it would be specifically interesting to analyze

the role of moderating factors as the establishment characteristics that influence the use of performance appraisal systems may also moderate the performance effects of those systems.

Table 1: Distribution of Performance Appraisal Systems

Annual Performance Appraisal for All Employees	*Percent*
Establishment does little or nothing about it	15.68
Establishment pays some attention to it	38.93
Establishment spends much time and energy on it	45.39

N = 542

Table 2: Descriptive Statistics of the Explanatory Variables (N = 542)

Variable	*Definition (Mean, Std.Dev.)*
Teams	Ordered variable indicating the use of teams (0 = no use, 1 = modest use, 2 = strong use) (1.031, .784)
Job rotation	Ordered variable for the practice of moving employees between different jobs within the establishment (0 = no use, 1 = modest use, 2 = strong use) (.788, .738)
Training	Ordered variable indicating the provision of further training for all employees (0 = no provision, 1 = modest provision, 2 = strong provision) (.958, .688)
Responsibility	Ordered variable indicating the delegation of responsibilities to lower levels of hierarchy (0 = no delegation, 1 = partial delegation, 2 = strong delegation) (.961, .669)
Capital intensive production	Dummy variable equal to 1 if management characterizes production as capital-intensive (.070, 256)
Complete product development	Dummy variable equal to 1 if the establishment provides products or services completely developed by the establishment (.513, .500)
Partial product development	Dummy variable equal to 1 if the establishment provides products or services partially developed by the establishment (.328, .470)
Public sector	Dummy variable equal to 1 if the establishment is a public sector establishment (.386, .487)
Information sharing	Ordered variable indicating information provision by management to employees (0 = no information provision about the situation of the establishment, 1 = partial provision of information, 2 = comprehensive provision of information) (1.031, .587)
Shared corporate culture	Ordered variable indicating the establishment's effort to reduce barriers between the various layers of hierarchy by implementing a shared corporate culture (0 = no effort, 1 = some effort, 2 = strong effort)(.515, .631)
Regular meetings with unions	Dummy variable equal to 1 if the works council of the establishment regularly meets with unions (.284, .451)
Occasional meetings with unions	Dummy variable equal to 1 if the works council of the establishment occasionally meets with unions (.542, .499)
Promotions	Ordered variable indicating the establishment's effort to provide internal career development and promotion opportunities (0 = no effort, 1 = some effort, 2 = strong effort) (.559, .660)
Personnel department	Dummy variable equal to 1 if the establishment has a personnel department or personnel manager (.878, 327)
Establishment age < 5	Dummy variable equal to 1 if the age of the establishment is less than 5 years (.105, .307)
Establishment age 5 – 15	Dummy variable equal to 1 if the age of the establishment is between 5 and 15 years (.140, 348)
Establishment age 16 – 25	Dummy variable equal to 1 if the age of the establishment is between 16 and 25 years (.085, .279)
Establishment size 100 – 199	Dummy variable equal to 1 if the establishment has between 100 and 199 employees (.238, .426)
Establishment size 200 – 999	Dummy variable equal to 1 if the establishment has between 200 and 999 employees (.400, .490)
Establishment size ≥ 1000	Dummy variable equal to 1 if the establishment has 1000 or more employees (.113, .316)
Industry dummies	Four industry dummies for manufacturing, banking and insurance, health care and social services, and transport and logistics are included.

Table 3: Initial Estimates

	Use of Performance Appraisal Method: Probit	*Intensity of Use Method: Ordered Probit*
Constant	-.8829 (0.21)	
Teams	.0400 (0.33)	.0545 (0.76)
Job rotation	-.2404 (6.17)***	-.1027 (1.59)
Training	.4416 (4.70)***	.3984 (6.35)***
Responsibility	.3521 (1.73)*	.3648 (3.45)***
Capital intensive production	.8407 (3.78)***	.4685 (4.11)***
Complete product development	.4617 (2.68)***	.2013 (1.48)
Partial product development	.4602 (2.02)**	.2760 (1.23)
Public sector	.1512 (0.80)	-.1126 (1.71)*
Information sharing	.0620 (0.46)	.2129 (1.85)*
Shared corporate culture	.2190 (2.15)**	.2676 (2.88)***
Regular meetings with unions	-.3676 (0.85)	-.3272 (2.17)**
Occasional meetings with unions	-.4665 (1.42)	-.3664 (2.70)***
Promotions	.0619 (0.45)	.2080 (2.11)**
Personnel department	.2602 (1.32)	.2577 (1.57)
Establishment age < 5	.0885 (0.61)	-.0278 (0.17)
Establishment age 5 – 15	.4460 (5.43)***	.2748 (1.87)*
Establishment age 16 – 25	.2955 (1.38)	.2319 (2.85)***
Establishment size 100 – 199	-.1116 (0.65)	-.0447 (0.54)
Establishment size 200 – 999	.3047 (3.76)***	.2186 (1.98)**
Establishment size $\geq$ 1000	-.0321 (0.15)	-.2247 (1.50)
Industry dummies	Included	Included
Pseudo R^2	.1689	
N	542	542

The table shows the estimated coefficients. T-statistics are in parentheses. Standard errors are adjusted for intra-industry correlation of random influences by using the Huber-White sandwich variance estimator. *** Statistically significant at the one percent level; ** at the five percent level; * at the ten percent level.

Table 4: Initial Estimates; Marginal Effects

	Use of Performance Appraisal	*Intensity of Use*		
		Establishment Does Little or Nothing about Performance Appraisal	*Establishment Pays Some Attention to Performance Appraisal*	*Establishment Spends Much Time and Energy on Performance Appraisal*
Teams	.0072	-.0099	-.0117	.0216
Job rotation	-.0435	.0186	.0220	-.0406
Training	.0800	-.0722	-.0855	.1578
Responsibility	.0638	-.0661	-.0783	.1445
Capital intensive production	.0948	-.0653	-.1193	.1846
Complete product development	.0847	-.0367	-.0429	.0796
Partial product development	.0758	-.0472	-.0623	.1095
Public sector	.0268	.0207	.0237	-.0445
Information sharing	.0112	-.0386	-.0457	.0843
Shared corporate culture	.0397	-.0485	-.0575	.1060
Regular meetings with unions	-.0735	.0648	.0628	-.1276
Occasional meetings with unions	-.0829	.0653	.0791	-.1445
Promotions	.0112	-.0377	-.0446	.0824
Personnel department	.0531	-.0526	-.0475	.1001
Establishment age < 5	.0153	.0051	.0059	-.0110
Establishment age 5 – 15	.0657	-.0439	-.0654	.1093
Establishment age 16 – 25	.0457	-.0371	-.0551	.0923
Establishment size 100 – 199	-.0210	.0082	.0094	-.0177
Establishment size 200 – 999	.0532	-.0386	-.0480	.0866
Establishment size $\geq$ 1000	-.0059	.0453	.0422	-.0876

The estimated coefficients shown in Table 3 are used to calculate the marginal effects at the means of the explanatory variables.

Table 5: Generalized Ordered Probit Estimation of the Intensity of Use

	β_1	β_2
Constant	-.1714 (.39)	-1.451 (7.31)***
Teams	.0323 (0.25)	.0841 (2.10)**
Job rotation	-.2039 (3.01)***	-.0368 (0.39)
Training	.4501 (4.11)***	.4276 (6.39)***
Responsibility	.3194 (1.50)	.4319 (4.89)***
Capital intensive production	.8652 (4.04)***	.3334 (2.39)**
Complete product development	.5384 (2.77)***	-.0341 (0.18)
Partial product development	.5211 (2.37)**	.0673 (0.23)
Public sector	.1435 (0.39)	-.3049 (2.56)**
Information sharing	.0132 (0.10)	.3372 (2.36)**
Shared corporate culture	.1958 (1.98)**	.2605 (2.68)***
Regular meetings with unions	-.3378 (0.79)	-.3175 (1.80)*
Occasional meetings with unions	-.3936 (1.21)	-.3696 (3.02)***
Promotions	.0546 (0.34)	.2312 (1.78)*
Personnel department	.3013 (1.45)	.3110 (2.66)***
Establishment age < 5	.1150 (0.71)	-.1439 (0.61)
Establishment age 5 – 15	.4645 (6.13)***	.1616 (0.67)
Establishment age 16 – 25	.3331 (1.50)	.2297 (1.71)*
Establishment size 100 – 199	-.1547 (0.94)	-.0161 (0.15)
Establishment size 200 – 999	.3078 (2.63)***	.1437 (0.74)
Establishment size $\geq$ 1000	-.0467 (0.18)	-.3842 (1.95)*
Industry dummies	Included	Included
N	542	

The table shows the estimated coefficients. T-statistics are in parentheses. Standard errors are adjusted for intra-industry correlation of random influences by using the Huber-White sandwich variance estimator. *** Statistically significant at the one percent level; ** at the five percent level; * at the ten percent level.

Table 6: Generalized Ordered Probit Estimation of the Intensity of Use; Marginal Effects

	Establishment Does Little or Nothing about Performance Appraisal	*Establishment Pays Some Attention to Performance Appraisal*	*Establishment Spends Much Time and Energy on Performance Appraisal*
Teams	-.0060	-.0273	.0333
Job rotation	.0376	-.0231	-.0146
Training	-.0831	-.0860	.1691
Responsibility	-.0590	-.1118	.1708
Capital intensive production	-.0985	-.0339	.1324
Complete product development	-.1009	.1144	-.0135
Partial product development	-.0865	.0598	.0267
Public sector	-.0260	.1455	-.1195
Information sharing	-.0024	-.1309	.1333
Shared corporate culture	-.0362	-.0669	.1030
Regular meetings with unions	.0682	.0554	-.1236
Occasional meetings with unions	.0714	.0742	-.1456
Promotions	-.0101	-.0813	.0914
Personnel department	-.0637	-.0561	.1198
Establishment age < 5	-.0201	.0764	-.0564
Establishment age 5 – 15	-.0693	.0051	.0642
Establishment age 16 – 25	-.0515	-.0399	.0914
Establishment size 100 – 199	.0300	-.0237	-.0064
Establishment size 200 – 999	-.0548	-.0021	.0569
Establishment size $\geq$ 1000	.0088	.1377	-.1465

The estimated coefficients shown in Table 5 are used to calculate the marginal effects at the means of the explanatory variables.

References

Alchian, A. and H. Demsetz. 1972. "Production Costs, Information and Economic Organization," *American Economic Review* 62: 777 – 795.

Addison, John T. and Clive R. Belfield. 2008. "The Determinants of Performance Appraisal Systems: A Note (Do Brown and Heywood's Results for Australia Hold Up for Britain?)," *British Journal of Industrial Relations* 46: 521 – 531.

Baker, G., M. Gibbs and B. Holmstrom. 1994a. "The Internal Economics of the Firm: Evidence from Personnel Data," *Quarterly Journal of Economics* 109: 881 – 919.

Baker, G., M. Gibbs and B. Holmstrom. 1994b. "The Wage Policy of a Firm," *Quarterly Journal of Economics* 109: 921 – 955.

Baker, G., M.J. Jensen and K.J. Murphy. 1988. "Compensation and Incentives: Practice vs. Theory," *Journal of Finance* 18: 593 – 616.

Boes, Stefan and Rainer Winkelmann. 2010. "Income and Happiness: New Results from Generalized Threshold and Sequential Models," *Social Indicators Research* 95: 111 – 128.

Breuer, Kathrin, Petra Nieken and Dirk Sliwka. 2010. "Social Ties and Subjective Performance Evaluations: An Empirical Investigation," IZA Discussion Paper No. 4913, Bonn.

Brown, Charles. 1990. "Firms' Choice of Method of Pay," *Industrial and Labor Relations Review* 43: 165 – 182.

Brown, C. and J. Medoff. 1989. "The Employer Size-Wage Effect," *Journal of Political Economy* 97: 1027 – 1059.

Brown, Michelle and John S. Heywood. 2005. "Performance Appraisal Systems: Determinants and Change," *British Journal of Industrial Relations* 42: 659 – 679.

Burgess, Simon and Paul Metcalfe. 2000. "The Use of Incentive Schemes in the Public and Private Sectors: Evidence from British Establishments," CMPO Working Paper No. 00/15.

Dewatripont, M., I. Jewitt and J. Tirole. 1999. "The Economics of Career Concerns Part II: Applications to Missions and Accountability of Government Agencies," *Review of Economic Studies* 66: 199 – 217.

Dixit, A. 1997. "Power of Incentives in Public vs. Private Organizations," *American Economic Review (Papers and Proceedings)* 87: 378 – 382.

Drago, Robert and Gerald T. Garvey. 1998. "Incentives for Helping on the Job: Theory and Evidence," *Journal of Labor Economics* 16: 1 – 25.

Drago, Robert and John S. Heywood. 1995. "The Choice of Payment Schemes: Australian Evidence," *Industrial Relations* 34: 507 – 531.

Elvira, Marta and Robert Town. 2001. "The Effects of Race and Worker Productivity on Performance Evaluations," *Industrial Relations* 40: 571 – 590.

Engellandt, Axel and Regina Riphahn. 2011. "Evidence on Incentive Effects of Subjective Performance Evaluations," *Industrial and Labor Relations Review* 64, Forthcoming.

Fang, Tony and John S. Heywood. 2006. "Output Pay and Ethnic Wage Differentials: Canadian Evidence," *Industrial Relations* 45: 173 – 194.

Fang, Tony and John S. Heywood. 2010. "Immigration, Ethnic Wage Differentials and Output Pay in Canada," *British Journal of Industrial Relations* 48: 109 – 130.

Freeman, Richard B. and Morris M. Kleiner. 2005. "The Last American Shoe Manufacturers: Decreasing Productivity and Increasing Profits in the Shift from Piece Rates to Continuous Flow Production," *Industrial Relations* 44: 307 – 330.

Friebel, G. and Raith, M. 2004. "Abuse of Authority and Hierarchical Communication," *RAND Journal of Economics* 35: 224 – 244.

Gibbons, Robert. 1998. "Incentives in Organizations," *Journal of Economic Perspectives* 12: 115 – 132.

Gibbs, M., K.A. Merchant, W.A. van der Stede and M.E. Vargus. 2004. "Determinants and Effects of Subjectivity in Incentives," *The Accounting Review* 79: 409 – 436.

Greene, W.H. 2003. *Econometric Analysis*. 5th ed. Upper Saddle River, NJ: Prentice Hall.

Grund, Christian and Dirk Sliwka. 2009. "Individual and Job-Based Determinants of Performance Appraisal: Evidence from Germany," *International Journal of Human Resource Management* 20: 2049 – 65.

Heywood, John S. and Patrick O'Halloran. 2005. "Racial Earnings Differentials and Performance Pay," *Journal of Human Resources* 40: 435 – 452.

Heywood, John S., W.S. Siebert and Xiandong Wei. 1997. "Payment by Results Schemes: British Evidence," *British Journal of Industrial Relations* 35: 1 – 22.

Hogan, Chad. 2001. "Enforcement of Implicit Contracts through Unions," *Journal of Labor Economics* 19: 171 – 195.

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

Jackson, S. and R.S. Schuler. 2003. *Managing Human Resources through Strategic Partnership* (8th edition). Canada: Thompson.

Jirjahn, Uwe and Kornelius Kraft. 2007. "Intra-firm Wage Dispersion and Firm Performance – Is There a Uniform Relationship?" *Kyklos* 60: 231 – 253.

Jirjahn, Uwe and Kornelius Kraft. 2010. "Teamwork and Intra-Firm Wage Dispersion among Blue-Collar Workers," *Scottish Journal of Political Economy* 57: 404 – 429.

Jirjahn, Uwe and Gesine Stephan. 2004. "Gender, Piece Rates and Wages: Evidence from Matched Employer-Employee Data," *Cambridge Journal of Economics* 28: 683 – 704.

Kerr, Steven. 1975. "On the Folly of Rewarding A while Hoping for B," *Academy of Management Journal* 18: 769 – 783.

Kreps, D.M. 1990. "Corporate Culture and Economic Theory," in J.E. Alt and J.A. Shepsle (eds.), *Perspectives on Positive Political Economy*, Cambridge University Press: 90 – 143.

Laffont, J.J. 1990. "Analysis of Hidden Gaming in a Three-Level Hierarchy," *Journal of Law, Economics and Organization* 6: 301 – 324.

Lewin, D. and Mitchell, D.J.B. 1995. *Human Resource Management: An Economic Approach* (2nd editon), Cincinnati, OH: South-Western.

Marx, L.M. and F. Squintani. 2009. "Individual Accountability in Teams," *Journal of Economic Behavior and Organization* 77: 260 – 273.

Milgrom, Paul and John Roberts. 1988. "An Economic Approach to Influence Activities in Organizations," *American Journal of Sociology* (supplement) 94: S154 – S179.

Milkovic, G.T. and Newman, J.M. 2002. *Compensation* (7th ed.), Boston: McGraw-Hill Irwin.

Moers, F. 2005. "Discretion and Bias in Performance Evaluation: The Impact of Diversity and Subjectivity," *Accounting, Organizations and Society* 30: 67 – 80.

Morita, Hodaka. 2005. "Multi-Skilling, Delegation and Continuous Process Improvement: A Comparative Analysis of US-Japanese Work Organizations," *Economica* 72: 69 – 93.

Murphy, K.R. and J.N. Cleveland. 1991. *Performance Appraisal: An Organizational Perspective*. Boston: Allyn and Bacon.

Pfeifer, Christian and Thomas Cornelissen. 2010. "The Impact of Participation in Sports on Educational Attainment – New Evidence from Germany," *Economics of Education Review* 29: 94 – 103.

Prendergast, C. 1993. "A Theory of Yes Men," *American Economic Review* 757 – 770.

Prendergast, C. 1999. "The Provision of Incentives in Firms," *Journal of Economic Literature* 37: 7 – 63.

Prendergast, C. and R.H. Topel. 1993. "Discretion and Bias in Performance Evaluation," *European Economic Review* 37: 355 – 65.

Prendergast, C. and R.H. Topel. 1996. "Favoritism in Organization," *Journal of Political Economy* 104: 958 – 978.

Shibata, Hirmichi. 2002. "Wage and Performance Appraisal Systems in Flux: A Japan-United States Comparison," *Industrial Relations* 41: 629 – 652.

Thornbill, S. and R. Amit. 2003. "Learning about Failure: Bankruptcy, Firm Age and the

Resource-Based View," *Organization Science* 14: 497 – 509.

Visser, Jelle 1995. "The Netherlands: From Paternalism to Representation," in Joel Rogers and Wolfgang Streeck, eds., *Works Councils: Consultation, Representation, and Cooperation in Industrial Relations*. Chicago, IL: Chicago University Press.

Wagner, Joachim. 2006. "Firmenalter und Firmenperformance: Empirische Unterschiede zwischen jungen und alten Firmen in Deutschland," in Lutz Bellmann and Joachim Wagner (eds.), *Betriebsdemographie*, Nuremberg: BeitrAB: 83 – 111.

Endnotes

[1] Of course, there exists an extensive body of psychological research on the psychometric properties of performance appraisal (Milkovich and Newman 2002). Yet, little research has been undertaken to examine what types of employers make use of performance appraisal systems.

[2] In Australia and Britain, about 70 percent of the establishments operate a formal system of appraisal (Brown and Heywood 2005, Addison and Belfield 2008). In contrast, the share of establishments using piece rates is only 33 percent in Australia (Drago and Heywood 1995) and 22 percent in Britain (Heywood et al. 1997). The share using profit sharing is 9 percent for Australia and 24 percent for Britain.

[3] Addison and Belfield (2008) have only applied a probit model to estimate the determinants of performance appraisal. Brown and Heywood (2005) have additionally used the standard ordered probit model to estimate the determinants of the share of workers who are covered by performance appraisal.

[4] Breuer et al. (2010) provide further evidence of favoritism. They find that superiors give more lenient ratings to employees who have worked for them for a longer period of time.

[5] Interestingly, the evidence from the U.S. and Canada shows that ethnic minority wage differentials are smaller when pay is based on objective performance measures (Fang and Heywood 2006, 2010; Heywood and O'Halloran 2005). Relatedly, Jirjahn and Stephan (2004) show for Germany that the gender wage gap is reduced when workers are paid piece rates. These findings conform to the hypothesis that objective performance measures provide less scope for discrimination.

[6] Indeed, Jirjahn and Kraft (2007) provide evidence that the productivity effect of differential rewards is stronger when there less discretion in performance measurement. Furthermore, a study by Engellandt and Riphahn (2011) indicates that subjective performance evaluations only provide positive incentive effects if superiors' evaluations are responsive to changes of individual worker performance over time.

[7] Note that the simple ordered probit model is typically estimated without including a constant term. If the model includes a constant term, the threshold value μ_1 is normalized to be equal to zero (Greene 2003: p. 736). However, most importantly, the estimated coefficients for the other variables do not depend on whether or not a constant term is included.

[8] See also Pfeifer and Cornelissen (2010) for an application of the generalized ordered probit model.

[9] Equation (7b) implies the marginal probability effect of training on an intermediate use can be negative even though the first coefficient is positive as the second coefficient works in the opposite direction. Moreover, the constant terms $\tilde{\mu}_1$ and $\tilde{\mu}_2$ influence the weights given to both coefficients.