EVIDENCE ON INCENTIVE EFFECTS OF SUBJECTIVE PERFORMANCE EVALUATIONS

AXEL ENGELLANDT AND REGINA T. RIPHAHN*

The authors investigate the effect of managerial performance evaluation styles on employee work effort. Using panel data on 4,080 employees in a Swiss unit of an international company for the period 1999–2002, they test two hypotheses using paid and unpaid overtime work as effort indicators. The company applies two performance-based remuneration mechanisms: an individual "surprise" bonus and one in which salary is affected by the extent to which an individual has reached personalized targets. The authors hypothesize that effort is higher in departments in which individual performance evaluations are more flexible over time as well as when surprise bonuses are used more frequently. Both hypotheses are supported, and the estimated effects are substantial. Increases in rating flexibility or bonus payments by one standard deviation above the mean increase average overtime hours by more than 20%. The findings are robust and suggest that surprise bonuses and flexible performance evaluations over time provide effective incentives.

Contracts that determine salary based on subjective evaluations have incentive effects that differ from those of other types of performance pay such as piece rates or commissions. This is important since subjective performance measures are increasingly used in organizations in which job complexity is high, individual productivity is hard to measure, and compliance with the norms of firm culture is considered to be part of individual performance. Such compensation schemes by nature have relational features and raise new issues: supervisors' and employees' perceptions of performance may not correspond and favoritism may arise. Only since the early 2000s have these characteristics of compensation schemes attracted considerable interest in the theoretical literature (see MacLeod 2003; Levin 2003).

In this paper we evaluate employees' effort response to managers' different styles of applying subjective evaluations. We provide an intra-firm test of the relevance of management styles (Bertrand and Schoar 2003) and test whether negative performance effects result from the discriminatory application of subjective evaluations (MacLeod 2003).

The question of whether employees provide more effort when it pays has been addressed by prior studies using various methods. Some investigate aggregate outcomes, comparing for example the performance of companies and entire industries

* Axel Engellandt is a professional statistical programmer in the field of validation and design of credit rating procedures. Regina Riphahn is Professor of Economics at the University of Erlangen-Nuremberg in Germany.

The authors thank Michael Gerfin, Jennifer Hunt, and Bentley MacLeod. They also thank the participants of the following conferences/gatherings: the Ausschuss für Bevölkerungsökonomie in Vienna; the third Transatlantic Meeting of Labor Economists in Buch; the Annual Meeting of the European Society of Population Economists (ESPE) in Bergen; the personnel economics workshop in Konstanz; and the seminar participants in Tübingen and at IZA for helpful comments. The authors gratefully acknowledge the company that provided the data (which are proprietary). Programs can be made available by writing to the authors at axelen@gmx.de or at regina.riphahn@wiso.uni-erlangen.de.

Industrial and Labor Relations Review, Vol. 64, No. 2 (January 2011). © by Cornell University.
0019-7939/00/6402 $05.00

with different human resource practices.[1] Survey-based studies examine whether incentive pay is correlated with higher productivity and wages (Booth and Frank 1999; Parent 1999; Pekkarinen and Riddell 2008). A few analyses point to sorting and effort responses when individual firms replace fixed wages by piece-rate payments (Lazear 2000; Shearer 2004; Haley 2003). Bandiera et al. (2005, 2006) demonstrated the relevance of social preferences for employee effort when they compared a system of relative performance evaluation to piece-rate pay. Many modern day work environments, however, are not amenable to systems of performance pay designed for easily observable output (Prendergast 1999). Instead, companies have to rely on subjective performance evaluations, which heretofore have been rarely studied empirically, nor with reliable data.[2]

The contributions of our study are threefold. First, we provide an analysis of the effectiveness of work incentives that are based on supervisors' subjective performance evaluations. In particular, this involves an evaluation of the effects of favoritism, which may result from employee rent-seeking behavior (see also Prendergast and Topel 1993, 1996; and Milgrom and Roberts 2004). Second, we investigate the incentive effects of surprise bonus payments, which have not yet been analyzed. Third, and in contrast to prior studies that have concentrated on certain types of manual employees or on managers exclusively, we evaluate employee behavior across heterogeneous occupational categories. Our research is based on panel data from one company, which allows us to control for unobserved employee- and department-level heterogeneity.

Our data describe about 4,000 employees of a large international company. The employees comprise not only managers, who have received much attention in the pay-performance literature,[3] but also all employees of a Swiss unit of our company: production workers, researchers, and administrators. Using paid and unpaid individual overtime work as indicators of effort, we analyze objective, productivity-related outcomes. This eliminates the measurement problems that arise when employee evaluations are compared across supervisors (see Prendergast 1999; Schwab and Olson 1990; Baker 2000) or when employees' statements on their attitudes, efforts, and performance are used. Clearly, objective indicators of individual behavior provide a more reliable measure of effort than the aggregate firm or industry productivity measures applied in some of the literature.

To identify the effect of subjective performance evaluations on behavior, we compare employees across company departments and over a period of time. Departments differ in the intensity with which the performance pay system is applied. Whereas some departments assign the same rating to a given person every year, possibly indicating favoritism, others are more flexible. Some spend their entire budget allocated to surprise bonuses, while others never use this tool. The validity of our identification strategy hinges on the exogeneity of employee assignment to departments, which we discuss below. The longitudinal nature of the data allows us to address issues of causality and to control for unobserved effects.

In complete contracts, every employee can be remunerated exactly according to effort. High job complexity and integration of behavioral norms in the employment objectives, however, contribute to the emergence

[1] Examples include Groves et al. (1994), who investigated the productivity consequences of managerial autonomy in Chinese industries; Jones and Kato (1995), who measured productivity effects of employee stock-ownership and bonuses for Japanese firms; Lee and Rhee (1996), who estimated similar models with South Korean time series on eight industries; and Morton (1998), who applied quarterly industry data from Taiwan; and Cable and Wilson (1989, 1990), who provided evidence on productivity enhancing effects of profit-sharing in the United Kingdom and Germany.

[2] Murphy and Oyer (2001) investigated the determinants of discretion in executive pay and Gibbs et al. (2004) focused on determinants and correlates of subjective pay for car dealers.

[3] Murphy (1999) surveyed the literature on the pay–performance relation for managers. For further evidence see Aggarwal and Samwick (2003) or Bandiera et al. (2006).

of non-contractible performance. Here, contracts are incomplete and third-party enforcement becomes increasingly costly. In such situations, performance pay is typically governed by relational contracts that are based on implicit long term agreements (MacLeod and Parent 1999). Obviously, aspects such as perceived fairness, trustworthiness, and absence of favoritism become crucial for the effectiveness of incentive provisions in these situations.[4] We identify and measure employees' responses to incentives set by departmental styles of applying the subjective evaluation system, thus revealing the aspects mentioned above in departmental compensation.

Institutional Background

We received personnel data on white- and blue-collar employees from an international company. The data cover one organizational unit in Switzerland between 1999 and 2002. During this time, the unit employed 6,425 individuals for most of whom we have repeated annual observations.[5]

The firm is organized into departments, of which we observe a total of 78 different ones in the raw data. Due to organizational changes, not all of these departments exist over all four years of our data. On average, there are 99 employees per department in our analysis sample; the department size varies between 4 and 670. Departments—the majority of which are concentrated at a given location—typically focus on specified tasks within the operation, which implies a clustering of similar occupations.

In this company, two performance-related remuneration mechanisms are in place: one is an individual surprise bonus in the range of €1,400–3,400, awarded for special achievements. Depending on employees' salary, the bonus can amount to between 10 and 100% of a gross monthly salary. This bonus payment is paid at the discretion of supervisors, who can spend a given annual budget for this purpose. On average, they apply this instrument about seven times a year, with significant heterogeneity across departments. About 15% of the departments did not pay out bonuses at all. The others paid on average 11 bonuses per 100 employees per year.[6] Employees should not be able to calculate the probability of receiving a bonus since they know neither the department bonus budget nor the amount of other bonuses. This surprise bonus system resembles Holmstrom and Milgrom's (1987) recommendation for an optimal compensation with linear bonus systems: in order to avoid a drop in employee effort after having reached specific goals of anticipated audits, the company institutes random performance audits to maintain in its employees the incentive to provide high levels of effort.

The other mechanism is a more complex performance pay system, in which—depending on the employee's level in the company—between 10 and 85% of the annual salary is determined by the outcome of an annual individual performance evaluation.[7] In a goal-setting session in the beginning of the year, supervisors apply a multi-dimensional scheme to describe the expected behaviors and achievements of the individual employee. At the end of the year, they rate individual performance in a condensed rank on a scale between 0 and 150% of the originally stated goals. The goals comprise to a substantial degree the compliance with behavioral norms. The ratings are cross-checked by the supervisors' managers. The absolute amount of the payout depends on

[4] See Fehr and Schmidt (2007) for experimental evidence on the relevance of trustworthiness, even for one shot relationships.

[5] The data exclude a very small fraction of top managers as well as some "social responsibility employees," such as disabled individuals, who are formally on the payrolls even though they no longer engage in the firm's productive activity. In addition, we disregard those employees for whom information on department number or performance rating is unavailable.

[6] We have no information on the individual bonus amounts (nor earnings), on the timing of payments during the course of the year, nor on the total bonus budget per department.

[7] The rate of merit pay for regular employees is between 10 and 13%; it reaches 20% for middle management and increases to as much as 70 to 85% for top management.

individual performance, the individual base salary, and the performance of the entire division. The ratio of performance pay relative to the base salary (PP) is calculated, based on the following formula:

(1) PP = individual performance rating × division performance × salary rate.[8]

Up until 2002, supervisors were required to adhere to a prescribed distribution of ratings within their department. About 10% of the employees were formerly grouped as low achievers; two-thirds were formerly grouped in the medium range; and about one-quarter were ranked in the group of top performers. This left supervisors substantial leeway in their rating decisions. After this prescribed distribution was abandoned, management clearly communicated the expectation that future rating distributions should be in the same range. Also in 2002, mid-year reviews were introduced to provide early feedback. Since the reforms did not change the general character of the evaluation system, they are disregarded in our analysis.[9] Performance-oriented pay was extended to all employees already in the mid 1990s, several years before our data were gathered.[10]

The expenditures for surprise bonus payments amount to 0.3% and those for evaluation-based performance pay amount to about 15% of the firm's base salary cost. The company also offers an employee stock ownership plan and a stock option plan for executives.

Hypotheses

In order to clarify the incentive mechanisms within this institutional framework, we formalize the expected incentive effects of performance pay—assuming risk neutrality—as follows: individual payoff W_t in any given year t consists of the fixed salary, S; a potential surprise bonus payment for the individual employed in department j; performance pay; and the cost of individual effort e, $C(e_t)$:

$$(2) \qquad W_t = S + Pr(Bonus)_j \times (Bonus\ value_j) + b\,(q_t - Q_t) - C(e_t).$$

The expected value of the surprise bonus in department j depends on the probability with which the surprise bonus is applied in that department j, $Pr(bonus)_j$, as well as the amount that is paid out, the $Bonus\ value_j$. Performance pay depends on the reward parameter b, which reflects the individual salary rate and the division performance (see equation (1)) and the difference between the realized rating, q_t, and the individual's performance goal in period t, Q_t. The realized rating q_t again is influenced by individual effort e_t as well as by stochastic factors u_{jt}, which might be specific to the department. For example, $q_t = e_t + u_{jt}$. The stochastic term u_{jt} might reflect unobserved rating determinants on the part of the supervisor, such as style, learning about employee characteristics, or favoritism, as well as purely random shocks.

With the exception of their fixed salary (S), employees do not know the parameters of the firm's incentive pay system. We assume that they use past observations to form expectations of their department's style of applying the system. Based on these observations, they are able to calculate an expected marginal effect of individual effort e_t on individual payoff W:

$$(3) \qquad \partial W_t / \partial e_t = Pr(Bonus_{jt}) \times \partial\,(Bonus\ value_{jt}) / \partial e_t + b\,(\partial q_t / \partial e_t) - \partial C(e) / \partial e_t .$$

[8] If, for example, the individual performance rating is 120%, the division performance is 105% and the salary rate 10%, this employee receives performance pay of 1.2 * 1.05 * 0.1 = 12.6% of base salary. For managers with identical personal and divisional ratings but a higher fixed salary rate of 20%, the payout would amount to 1.2 * 1.05 * 0.2 = 25.2% of the base salary. The employees in our data belong to divisions with at least 1,000 employees such that divisional results should be exogenous for the individual employee. We do not have information on individual salary levels.

[9] In addition, they are irrelevant for the more technical reason that we consider only lagged rating distributions which occurred within the regime prior to the reforms. Our data confirm that the rating distribution did not change in 2002.

[10] Taylor and Pierce (1999) and Kahn and Sherer (1990) have pointed out that a system's effects may differ depending on whether it was just introduced or had been in place for several years. In this respect, we consider our study as an evaluation of an ongoing system in which initial employee responses to its introduction already faded.

Equation (3) yields the determinants of individual effort: the probability of a surprise bonus payment in department j, which we consider to be exogenous to individual effort e_i; the sensitivity of the surprise bonus amount in department j to individual effort e_i; the sensitivity of individual ratings q_t with respect to effort e_t, which may vary across departments depending on favoritism; and the disutility of additional effort. This simple model yields testable hypotheses regarding the effectiveness of performance pay and of alternative "styles" of applying the remuneration system for employee effort.

We expect to see higher levels of effort among those individuals whose performance is more strongly reflected in their pay and for whom a given amount of effort yields a larger wage differential. For the incentive system in this company, this translates into two mechanisms. First, the benefit of effort is higher in departments with a high expected probability of surprise bonus payments, such as in departments that provide more surprise bonuses per employee. Therefore, we expect higher effort there.[11] Second, one would expect larger responses to larger bonus payments. However, because we do not observe bonus amounts, we cannot consider the expected value of the bonus payment in our empirical specification. Instead, we test whether the likelihood of receiving a positive payment affects effort. This provides a lower bound of the marginal effect of the bonus instrument on effort.

Second, the rating behavior of departments should matter for individual behavior. The connection between departmental rating and behavior can be derived from equation (3), in which performance incentives are stronger with higher values of $\partial q_t / \partial e_t$. This prediction also follows from the models presented by MacLeod (2003) and Levin (2003), in which a downward bias in principal evaluations and the perception of unfair evaluations cause lower performance and costly conflicts. For our scenario, this translates to an additional prediction. One can compare department rating policies based on individual-specific rating flexibility over time. In a low-variability scenario, employees receive about the same performance rating every year even if their performance varies, such as in a case of favoritism. In a high-variability scenario, individual ratings in one year have little predictive power for next year's rating, such as if they closely match actual performance. We expect that performance incentives are stronger in the high-variability scenario.

We test these two hypotheses regarding the sensitivity of employee effort to the department style of applying the remuneration system below.[12] Since data limitations do not allow us to separate the effect of effort on true performance and the effect of true performance on supervisor behavior, we can only consider the joint or reduced-form effects. Moreover, we do not model the behavior of supervisors but take their behavior as a given.

Data and Measurement

To test our hypotheses, we evaluate the correlation between individual employee effort and the department-specific style of performance pay and test its statistical significance in a regression that controls for possible composition effects. Before describing our empirical approach, we explain as our key variables the measures of employee effort and of performance-pay incentives.

An ideal indicator of employee effort would closely mirror employee productivity. Unfortunately, such indicators are not available for employees as diverse as the ones we investigate here. Therefore, we apply two alternative indicators that measure the provision of paid and unpaid overtime. In this

[11] An exception of little practical relevance is the case when all employees receive a bonus, at which time reducing the probability of bonus payments might increase incentives. However, given the budgetary restrictions faced by supervisors in charge of the bonus instrument, this is merely a hypothetical case. In fact, no department ever paid bonuses for more than 62% of its employees, the average being 7%.

[12] Our data do not allow us to evaluate the effect of changes in effort on bonus amounts or the relevance of the cost of effort.

firm, regular overtime work is not remunerated financially but is used to substitute for working hours at a later time. Employees cannot, however, carry balances of more than 120 hours from one month to the next.[13] Our data describe the total number of overtime hours that have accumulated at the end of the year; numerous employees accumulate more than 120 hours of overtime, which end up being a gift to the company.

Our first effort measure takes the employee year-end balance as it is and indicates how many hours of overtime a given employee accumulated at the end of the year. This number is interpreted as a time credit that the employee grants the company. The lower an individual's motivation, the less likely the person may be to work beyond contractual requirements and the lower the balance on the individual overtime account. Accumulated overtime hours, therefore, can be a plausible indicator of employee motivation and effort. The measure has previously been applied in the literature.[14] Our second outcome measure, taken from the same database, is binary and indicates whether an employee accumulated more than 120 hours of overtime at the end of the year. Since additional hours of work should presumably enhance productivity, an employee's willingness to accumulate more than 120 hours of overtime—and thus to work without pay—appears to be a reliable indicator of effort.

One might argue that indicators of overtime work reflect labor and effort as demanded by the firm rather than the supply and effort of the individual worker. However, a number of circumstances render this scenario unlikely in our case. First, Swiss labor laws are extremely flexible, provide little employment protection, and allow the firm to accommodate demand shocks by adjustments in the number of employees instead of the number of hours worked per person (see Engellandt and Riphahn 2004). Second, we provide robustness tests to our results in which we control for time-varying unobserved heterogeneity for groups of departments, which should account for demand shocks. Finally, our company utilizes a separate category of overtime work, which is labeled "overtime by order of the supervisor." This would be the preferred response to demand shocks at the department level. Importantly, this type of ordered overtime is not counted in our effort measure because it is remunerated immediately and is not reflected in the individual overtime accounts.[15]

If our overtime measures are reliable effort indicators, they should be correlated with individual ratings and bonus payments. In simple test regressions of ratings and bonus payments on overtime hours and the voluntary provision of overtime, we find highly significant correlation patterns with t-values on all coefficients larger than ten. The number of overtime hours by itself explains 23.3% of the variation in ratings. This leaves no doubt as to the indicator value of our effort measures.

We observe an average number of 23.5 accumulated overtime hours at the end of the year in our sample and an overall propensity of 3% of all employees to provide overtime hours without pay. The individual correlation coefficient for overtime hours in neighboring periods is sizeable at about 0.49. A simple regression of overtime hours on department fixed effects explains about 48% of the variation, leaving only 52% of the heterogeneity in overtime hours to be explained by changes over time or by inter-individual heterogeneities.

[13] Employees who accumulate 140 hours of overtime at the end of one month, for example, will start out the next month with a surplus of only 120 hours. Twenty hours are deducted from their overtime account.

[14] Landers et al. (1996) discussed scenarios in which firms use hours of work as a signal of unobserved employee characteristics. Drago (1991) used the willingness to work an extra unpaid 20 minutes and the propensity to go to work even if not feeling well as performance indicators. Sousa-Poza and Ziegler (2003) considered overtime work as an indicator of employee productivity.

[15] In addition, the effect of time-varying heterogeneity, e.g. in unobserved demand shocks on overtime provision, is dampened by the use of lagged incentive indicators. Beyond that and given that we only analyze two years of data, we assume that any time-varying demand shock that hits department j in period t is uncorrelated with individual effort in period $t+1$.

Individual effort will respond to past and expected future remuneration responses. Because future supervisor evaluations are unknown, individuals base their current effort choice on expectations. Such expectations are informed by past departmental remuneration behavior as a plausible and likely the most reliable signal of the current departmental remuneration style that is available to the individual.[16] Therefore, we apply lagged indicators of past departmental remuneration behavior as indicators of the existing incentive system and of the expected marginal benefit of effort.[17]

We use two indicators of performance pay incentives. First, we hypothesize that rating flexibility over time matters. We therefore generate an indicator of the change in individual ratings over time by department. This measure is calculated in two steps. We obtain for every person the difference in performance ratings for two subsequent years; in any given department, the average of these differences should approximate zero. Next, we calculate the standard deviation of these person-specific changes by department as an indicator of departmental rating dispersion.[18] Second, we consider the department-specific number of bonus payments per year and per employee.

Empirical Approach, Identification, and Sample

Given these incentive and effort indicators, we estimate the following specification using data on individual employee i who is assigned to department j in period t:

$$(4) \qquad Y_{it} = \alpha + \beta I_{jt-1} + \gamma X_{ijt} + \delta_j + \nu_t + \ _i + \varepsilon_{it}.$$

Here, Y_{it} indicates employee i's effort (overtime hours and voluntary work) in period t; I_{jt-1} measures the pre-period value of the performance pay incentives (that is, rating flexibility and bonus payments) in department j as implied by our hypotheses; and X represents characteristics of both employee and department. α, β, and γ are parameters to be estimated. δ_j, ν_t, and $_i$ represent department, year, and individual fixed effects, respectively. ε is a random error term.

The department fixed effects account for department differences with respect to technology or unobserved characteristics of supervisors (e.g. capriciousness in rating), for whom, unfortunately, personal information is not available. Moreover, they capture any permanent "cultural" differences among departments, such as those related to peer pressure, as well as demand shocks that permanently affect departments for the time that we observe them. To the extent that supervisors stay with departments over time—which we cannot observe—they capture supervisor styles. Department type fixed effects are necessary to strengthen the identification of the effects we are interested in, as previous literature has discussed (Holmstrom and Milgrom 1994) and shows (MacLeod and Parent 1999) that optimal compensation systems vary by type of job. If the jobs within a given department are similar in character, then we can use the heterogeneity in the application of incentives across departments and over time to identify the effect these incentives have on behavior. It is important to note that we take the behavior of supervisors as a given and do not attempt to explain any differences.

Calendar year-specific shifts in effort are captured by the fixed effect υ_t. Individual fixed effects $_i$ capture person-specific heterogeneity in otherwise unobserved attributes such as ambition or family obligations, which affect the willingness to work overtime.

Among our control variables (X), we consider as individual characteristics age and its square, sex, and marital status. To proxy individual human capital, we use indicators of job requirements. The measure is originally

[16] In fact, in our data the raw correlation between past and present rating flexibility amounts to 0.55.

[17] The use of lagged indicators corresponds to Levin's model (2003), in which employees respond *after* receiving a performance evaluation. This modeling framework is also consonant with the forward-looking nature of rational expectations.

[18] This excludes the possibility that simple mean shifts in effort, output, and rating cause a spurious correlation between overtime and the flexibility measure.

Table 1. Descriptive Statistics

Variable Group and Description	*Mean*	*Standard Deviation*
Dependent Variables (Employee Effort in t)		
Overtime hours	23.510	44.224
More than 120 overtime hours (0/1)	0.030	0.171
Incentive Indicators (Supervisor Styles measured in t-1)		
Standard deviation of department rating changes	16.641	2.713
Bonus payments per 100 employees	7.150	6.196
Human capital and hierarchical position in the firm		
Job level = 1 (0/1)	0.008	0.090
Job level = 2 (0/1)	0.053	0.225
Job level = 3 (0/1)	0.115	0.319
Job level = 4 (0/1)	0.124	0.329
Job level = 5 (0/1) (*reference group*)	0.169	0.375
Job level = 6 (0/1)	0.136	0.343
Job level = 7 (0/1)	0.070	0.256
Job level = 8 (0/1)	0.080	0.273
Job level = 9 (0/1)	0.102	0.303
Job level = 10 (0/1)	0.062	0.241
Job level = 11 or beyond (0/1)	0.079	0.270
Additional Control Variables Individual Characteristics		
Age	44.936	8.964
Age squared	2099.548	789.552
Male (0/1)	0.634	0.482
Married (0/1)	0.688	0.463
Department Characteristics		
Average age of employees	44.936	1.578
Share of male employees	0.634	0.159
Average job level of employees	6.204	1.644
Number of employees (in thousands)	0.242	0.202
Year Dummies		
Year = 2001 (0/1)	0.472	0.499
Year = 2002 (0/1) *(reference group)*	0.528	0.499
Department Type		
Research department (0/1) *(reference group)*	0.587	0.492
Administrative department (0/1)	0.063	0.244
Production department (0/1)	0.350	0.477

Note: The table describes all variables at the level of N = 7,335 person-year observations. Thus department characteristics are weighted by the number of employees per department.

available in twenty discrete categories (levels), which we combine to eleven indicators. These covariates control for differences in individual effort that may be correlated with individual hierarchical position and human capital. The department is characterized by the number of employees, their average age, job level, and the share of male employees. We can distinguish production, administration, and research departments. Descriptive statistics are presented in Table 1.

We use least squares to estimate the model for both effort indicators. Of key interest are estimates of the β-vector, which indicate the association between departmental performance incentives and individual effort responses. Several conditions need to be met for β to reflect an unbiased estimator of the causal effect of incentives on effort.

First, the incentive measure must be exogenous. Thus, we need to exclude the fact that employees self-select into departments

based on departments' style of providing performance pay. From private conversation we know that the human resources management of our company considers it extremely unlikely that moves occur in response to departmental performance pay policies. The firm does not publish the performance ratings at the departmental level, so employees can learn about department differences only by personal experience or by word of mouth. In order to test whether endogenous employee sorting affects our results, we perform three robustness tests. (a) We compare the coefficient estimates that are obtained with and without controls for individual fixed effects. This should account for endogenous employee sorting across departments to the extent that the relevant unobserved employee characteristics do not vary according to time. (b) In additional estimations, we consider only employees who are not observed to move between departments. If endogenous sorting is behind the incentive–effort correlation, the estimated correlation should decline if those who sorted themselves are dropped from the sample. (c) The final robustness test re-estimates our model separately for those with long and short tenure. Since employees who are just hired from the external labor market should have less information about department characteristics, they are less likely to be able to sort themselves into a department of their choice. If the incentive–effort correlation patterns are robust for this subsample as well, they are unlikely to be subject to self-selection bias.

The second condition to be met concerns identification of the incentive effect: this requires that the unobserved determinants of the incentive measures are uncorrelated with the effort indicator. We control for any omitted variables at the department level using fixed effects. Therefore, permanent department-specific heterogeneity does not bias our estimates. We present a number of additional robustness tests to corroborate the reliability of our results.

Third, we have no information on whether and how demand for overtime varies across departments. Because our key explanatory variables vary only at the department-year level, it is not possible to fully control for unobserved heterogeneity at the department-year level or for department-specific serial correlation in the error term. Any permanent differences between departments, however, are controlled for by the department fixed effects.[19]

Finally, we have to acknowledge that our data do not allow us to control for the self-selection of employees into this company. To the extent that these workers systematically differ from the population, we may not be measuring the overall effect of the firm's incentive mechanisms. However, we certainly identify the treatment of effect of the incentives on those who self-selected into the treatment; that is, we identify the average treatment effect on the treated.

We restrict our sample to include only full-time employees who were employed with the firm year-round and for whom a department indicator is available. We drop a few observations with missing or extreme performance ratings (values 0 and 1 or beyond 200), which reflect individual circumstances that are not necessarily correlated to actual performance. In addition, only those employees are sampled who were employed with the firm continuously for three years, because we intend to test the relevance of past experience for current behavior. Since our data contains performance indicators only for those employed year round, we do not allow new employees of the firm to enter our sample during the course of a year.

We lose observations in departments that are observed for less than three periods because we can measure the lagged incentive indicators only if individuals are employed in departments that exist continuously over the course of at least three periods.[20] Since it is unlikely that reorganizations of the firm respond to the heterogeneity of departmental ratings or the frequency of bonus payments,

[19] Below we discuss robustness tests, which consider time-varying controls for department groups. We discussed this point more generally in the previous section.

[20] To generate the lagged standard deviation in individual rating changes by department we need the difference between ratings in periods t-1 and t-2.

Table 2. Effort Outcomes by Incentive Intensity

Average characteristics of employees in departments with	*Overtime Hours in t*	*Propensity to Provide More than 120 Hours Overtime in t*
Standard deviation of department rating changes in t–1		
≤ Median	21.83	0.022
> Median	24.57***	0.035***
Bonus payments per employee in t–1		
≤ Median	26.44	0.036
> Median	20.03***	0.023***

Note: The median incentive intensities are calculated based on comparisons across departments and years (N = 74), while the average overtime outcomes reflect the characteristics of the 7,335 person-year observations.
***Statistically significant at the .01 level.

we consider the selection based on department stability as an exogenous criterion. Since we require indicators of rating dispersion, all departments must have at least three employees at any point in time to enter the sample.

Our analysis sample consists of 7,335 employee-year observations covering 4,080 different individuals. These employees are employed in 42 different departments of which 32 are observed over the course of two periods, 2001 and 2002.[21]

Results

First Evidence

Table 2 describes the average effort outcomes for individuals in departments in which performance incentives are above or below the department median. We would expect more effort—more overtime hours and a higher probability of working for free—in departments with flexible ratings, that is, large inter-temporal rating changes and with many bonus payments per employee. However, not all of the table's entries confirm our expectations.

The first row corroborates hypothesis one: we find more overtime hours and a higher propensity to work for free in departments in which individual performance ratings are flexible over time. The second row yields no support for hypothesis two. Individuals who work in departments using bonus payments frequently put in less overtime and have a significantly smaller propensity to work for free. This hardly supports the hypothesis that greater incentives are correlated with more effort. Next, we investigate in multivariate regression analyses whether these outcomes are robust to controls for potential composition effects.

Our study is motivated by the idea that supervisors differ in their behavior and that such differences in style can have real consequences for employee behavior. As a first test of whether there are indeed behavioral differences between supervisors and not just responses of identical supervisors to different employee groups, we evaluate the correlation between the distribution of efforts in a given department and the flexibility of ratings over time. If managerial styles did not differ, a given change in the distribution of effort in a department should call forth the same change in the rating distribution across all departments. We find that departmental effort distributions are uncorrelated with supervisor rating distributions. Neither is the

[21] We have only two observation years in the final sample because the definition of our main independent variables requires information on two prior periods and our panel covers four years overall. With only two annual observations, it is not possible to control for serial correlation or correlation patterns between departments. Unfortunately our data are not available at a monthly level. As a result, dynamic effects within a department over the course of the year cannot be investigated.

Table 3a. Linear Regression of Overtime Hours

	1	*2*	*3*	*4*
Indicators of Performance Pay Incentives in *t*–1:				
Std. Deviation of Rating Change	2.098	2.089	0.406	2.066
	(0.397)***	(0.380)***	(0.406)	(0.361)***
	(0.299)***	(0.297)***	(0.267)	(0.299)***
	(1.306)	(1.296)	(0.564)	-
Bonus Payments	0.725	0.749	0.684	0.856
	(0.158)***	(0.151)***	(0.184)***	(0.147)***
	(0.130)***	(0.130)***	(0.164)***	(0.141)***
	(0.428)*	(0.425)*	(0.408)*	-
Year Fixed Effect (1)	yes	yes	yes	yes
Individual Characteristics (14)	-	yes	yes	-
Department Characteristics (4)	-	-	yes	-
Individual Fixed Effects	-	-	-	yes
Department Fixed Effects (42)	yes	yes	yes	yes

*Statistically significant at the .10 level; **at the .05 level; ***at the .01 level.

distribution of changes in departmental efforts associated with the distribution of changes in supervisor rating behavior. This supports the presumption of supervisor heterogeneity and styles.

Tables 3a and 3b provide the results of least squares estimations of four specifications. Each of these controls for the two incentive indicators as well as for a year indicator and department fixed effects. These account for permanent differences across departments such as culture, and type of work, or the characteristics and style of a supervisor. As a result, the effects of the incentive measures are identified based on their changes within departments over time as well as based on individuals who move between departments. We present the unadjusted standard errors of the least squares estimator, those obtained when clustering at the individual and at the department-year level.[22]

The first specification in column 1 considers the two lagged incentive measures: the standard deviation of person-specific rating changes between period *t*-1 and *t*-2 and the number of bonuses paid per 100 employees in *t*-1. The estimated coefficients on both incentive measures are of the expected sign: the more flexible performance ratings are, the more overtime is provided on both dimensions. Employees in departments with more frequent bonus payments also provide significantly more effort.[23] The coefficient estimates are highly significant.[24]

In order to account for potential composition effects, we add covariates to the model in subsequent columns of Tables 3a and 3b. In specification 2, we consider 14 individual-specific covariates plus a year dummy and in specification 3, we add four time-varying measures that describe department characteristics. Specification 4 then considers individual-specific fixed effects instead of individual and department characteristics. Individual fixed effects control for

[22] We also estimate all models separately by gender. Since the coefficients on the incentive effects differed neither in substance nor in a statistically significant manner for the two subsamples, we present the joint models only.

[23] The change in the correlation between number of bonus payments and overtime work from Table 2 to Table 3 suggests that there are systematically heterogeneous patterns of bonus provision and overtime work either over time or across departments.

[24] The marginal effects of the incentive measures are of the same direction and similar levels of significance when probit or logit estimators are applied instead of the linear probability model for the dichotomous overtime indicator.

Table 3b. Linear Regression of Providing more than 120 Overtime Hours (0/1)

	1	*2*	*3*	*4*
Indicators of Performance Pay Incentives in *t*–1:				
Std. Deviation of Rating Changes	0.0050	0.0050	0.0014	0.0048
	(0.0016)***	(0.0015)***	(0.0016)	(0.0015)***
	(0.0012)***	(0.0012)***	(0.0012)	(0.0012)***
	(0.0033)	(0.0030)	(0.0017)	-
Bonus Payments	0.0014	0.0015	0.0015	0.0018
	(0.0006)**	(0.0006)**	(0.0007)**	(0.0006)***
	(0.0006)**	(0.0006)**	(0.0008)*	(0.0007)***
	(0.0010)	(0.0010)	(0.0008)**	-
Year Fixed Effect (1)	yes	yes	yes	yes
Individual Characteristics (14)	-	yes	yes	-
Department Characteristics (4)	-	-	yes	-
Individual Fixed Effects	-	-	-	yes
Department Fixed Effects (42)	yes	yes	yes	yes

Note: The tables present estimated coefficients and standard errors in parentheses. All models are estimated on 7,335 person-year observations. The tables show three different standard errors, first the uncorrected least squares ones, then those clustered by individual employee, and finally those clustered by department and year (these cannot be calculated for the person-specific fixed effects estimator in col. 4). The individual and department characteristics contain those listed in Table 1. The number of estimated parameters for each group of indicators is provided in parentheses. The department-specific fixed effects were controlled for using a set of 42 indicator variables.
*Statistically significant at the .10 level; **at the .05 level; ***at the .01 level.

person-specific heterogeneity, such as intrinsic motivation and leisure preference.[25] Overall, the nature of the correlation between the incentive and effort measures does not vary substantially when different specifications are considered, and only in a few cases is the precision of the estimates affected. Since the results of column 1 hold even when controlling for individual fixed effects (see column 4), the estimated incentive effect is not due to endogenous selection of employees into departments. The results are robust across specifications and support the hypotheses regarding the intertemporal variability in individual ratings and the effort enhancing effects of bonus payments.[26]

The magnitude of the impact of rating flexibility and bonus payments on overtime hours implied by the coefficients in Tables 3a and 3b is substantial: an increase in rating flexibility by one standard deviation above the mean raises overtime hours by 5.6 hours (2.713×2.066) or 24% relative to the mean (see column 4 of Table 3a). The probability of providing more than 120 hours overtime of any given month—or free labor—increases by 0.013 percentage points (2.713×0.0048) or 43% relative to the mean (see column 4 of Table 3b). The response to changes in the number of bonus payments by one standard deviation is similar in

[25] An alternative interpretation of our results could be that rating flexibility does not indicate the absence of favoritism but instead the capriciousness of a supervisor. Moreover, if individual effort is considered to be a permanent characteristic, it might actually be a constant rating over time, which is free from favoritism. Both interpretations, however, can be refuted based on the results of specification four. Even controlling for supervisor characteristics such as capriciousness, and even controlling for worker fixed effects such as general effort level, our results are robust. In additional regressions, we estimate the effect of rating flexibility in departments with below-average effort variability. Even here, the effect of rating flexibility was positive. Tournament theory would suggest that effort declines in departments whose supervisor ratings are randomly volatile. It seems that the positive effect of supervisor rating flexibility on individual effort would be implausible if it were to reflect mere supervisor capriciousness.

[26] These results also are robust to considering only one of the incentive measures at a time.

Table 4a. Linear Regression Adding Interaction Effects of Effort Indicators in *t*-1: Coefficients of Interaction Effects

	Overtime Hours		*Propensity to Provide More than 120 Hours of Overtime*	
	Std. Dev. of Rating Changes	*Bonus*	*Std. Dev. of Rating Changes*	*Bonus*
Males (vs. Females)	−0.436	0.046	−0.0027	0.0001
	(0.603)	(0.257)	(0.0025)	(0.0013)
Performance below (vs. above) median	−0.121	−0.377*	−0.0005	−0.0002
	(0.137)	(0.212)	(0.0006)	(0.0011)
Department size below (vs. above) median	0.074	1.220***	0.0001	0.0008
	(0.206)	(0.249)	(0.0008)	(0.0011)
Tenure below (vs. above) median	0.347	−0.292	−0.0005	0.0003
	(0.233)	(0.241)	(0.0011)	(0.0012)
Manager (vs. employee)	−1.047	−1.210***	−0.0020	−0.0015
	(0.724)	(0.231)	(0.0026)	(0.0011)
Research (vs. production) department	−10.686***	−3.371***	−0.0261***	−0.0077***
	(0.773)	(0.413)	(0.0033)	(0.0028)
Administration (vs. production) department	−7.817***	−1.206***	−0.0177***	−0.0031
	(0.722)	(0.440)	(0.0029)	(0.0030)

**Statistically significant at the .05 level; ***at the .01 level.

magnitude to the above effect of changes in rating flexibility.[27]

Robustness Tests

We investigate whether the incentive effects found so far are robust to changes in sample, variable definition, and specification. Specifically, we re-estimate the models in Tables 3a and 3b by adding interaction terms and using only specific subsamples to investigate whether certain employee subgroups respond differently to incentives provided by performance-based pay. In Table 4a, we present the coefficients of interaction effects, which describe differences in incentive effects by employee sex, for those with performance ratings above and below the median, for those working in large and small departments, and for those with tenure of more and less than the median. Because hardly any of these coefficient estimates are significantly different from zero, our conclusions regarding the direction of the effects appear to be robust for these subsamples.

There are two types of subsamples with systematically different response patterns: managers vs. regular employees and three department types. The bottom rows of Table 4a reveal that the coefficients of all incentive interaction terms for managers are negative, which means that their effort response to incentives provided by rating flexibility and bonus payments is generally smaller than that of other employees. This result confirms what MacLeod and Parent (1999) have suggested—that optimal employment contracts should vary by type of job. In addition, it is intuitively plausible first that overtime hours provide a poorer proxy for the effort of managers than employees and second that short term rating and bonus incentives are less relevant for managers who are more

[27] We confirm an incentive effect of the number of bonus payments. It is possible that in reality the size of the bonus payments interacts with their frequency and affects the incentive mechanism. However, since we have no information on the amount of bonus payments we restrict our test to the hypothesis as derived above and show that already the mere number of payments appears to modify behavior.

likely motivated by promotion-based incentives and other financial incentive programs. Therefore, the weaker response intensity among managers is not surprising. The overall incentive effect for managers remains positive, even when the model is estimated on the subsample of managers only.

Similarly, the significant negative interaction effects in the last two rows of Table 4a indicate that among the three types of departments in the company, the incentive mechanisms yield the largest effects among production employees. A potential explanation for the effectiveness of incentive measures for this group of employees is that supervisors here use easily observable objective performance measures. This connection also implies that favoritism and unfairness are more easily observable among colleagues and costly conflicts of opinion are more likely to occur in production than in departments whose output is less easily observable. For research departments, the overall effect of incentive provision on effort is negative. Separate estimations by department type confirm negative total effects of the incentive instruments on overtime provision for employees in these departments. It is plausible to assume that for scientists, both the incentive mechanism and overtime as an effort measure are not fully appropriate: the effort of researchers may not be reflected in their overtime hours. Since individual performance is particularly difficult to evaluate in these occupations, where effort also may substantially precede any measurable output, the deviation from our hypothesized patterns of behavior is not surprising. Employees in administrative departments do not respond as strongly to incentives as production employees; however, their overall effort response is still positive (even when evaluated in separate estimations).[28]

As a second robustness check, we modified the definition of our incentive measures. Fundamentally, we are testing whether past experience of department policies affects future behavior. Employees are at times reassigned to new departments; therefore, we can specify the indicator of department policies in two ways. So far, we have measured our incentive indicators for the last year in the department in which an individual works today. One could argue instead that it is the *individual* experience with bonus incentives that matters as opposed to the history of the *department* that a new employee has just joined. We redefined our incentive measures to reflect the average of last year's rating and bonus experience of all employees currently working in a department, independent of where this experience was gathered. Even with redefined incentive indicators, the estimated coefficients do not change sign or significance. Therefore, the definition of our incentive measure does not seem to bias our results.[29]

There are many ways to measure dispersion; so far, we have looked only at the standard deviation of changes in ratings over time. To investigate the sensitivity of the results to this choice, we applied three alternative measures of the dispersion in rating changes: the variance, the average deviation from the median, and the difference between the 90th and 10th percentile of the department-specific distributions. Overall, the results are robust to these specification changes.[30]

In theory, individuals might move endogenously between departments in response to departmental performance pay policies. Though employees do move—in our data about 7% of the observations changed departments over four years—the human resources management of the company considered it extremely unlikely that employees would do so in response to departmental performance pay policies. We suggested above that three robustness tests

[28] Since there might be (time-varying) shocks affecting only certain groups of departments, we re-estimated the models of Tables 3a and 3b, controlling for the department type separately for both time periods. This did not affect the results.

[29] To save space, the results are not presented here but are available from the authors upon request.

[30] To save space, the results are not presented here but are available from the authors upon request.

Table 4b. Linear Regression without Observations of Individuals Who Changed Departments

	Overtime Hours		*Propensity to Provide More than 120 Hours of Overtime*	
	No change ever	*No change last 2 years*	*No change ever*	*No change last 2 years*
Indicators of Performance Pay Incentives in $t-1$:				
Std. Deviation of Rating Changes	1.944***	2.009***	0.0037***	0.0042***
	(0.370)	(0.334)	(0.0015)	(0.0013)
Bonus Payments	0.840***	0.814***	0.0016***	0.0016***
	(0.156)	(0.151)	(0.0007)	(0.0007)
Number of observations	5,704	6,300	5,704	6,300

***Statistically significant at the .01 level.

reveal that individual-specific unobservables do not affect our results. The first was to compare estimation results with and without controls for individual fixed effects. The second was to drop those individuals from the sample who were observed to move between departments both for the four years of our data as well as for just the two years remaining for the analysis. If our results were due to endogenous sorting of employees across departments, we would expect a clear drop in the coefficient estimates when the "sorted employees" are omitted from the sample. The estimation results for both dependent variables (as presented in Table 4b) are robust to this change in sample. The third test was to re-estimate separate models for individuals with short and long tenures in the firm. We found no significant differences in the response to incentives for individuals in the two groups. In sum, all three tests suggest that individuals' endogenous department changes are an unlikely explanation of our findings.

Alternatively, our results might be affected by endogenous attrition of low productivity departments over time. To determine the relevance of this type of mechanism, we repeated the estimations, this time considering only those departments that are observed in both years and in consequence were observable over the full four years of our data (see Table 4c). The results are unchanged.[31] In a final robustness check, we randomly assigned department numbers and characteristics to employees and, as expected, found no significant effects of incentives on behavior.

Conclusions

Our study evaluates the effectiveness of performance pay based on subjective supervisor evaluations as an incentive mechanism to generate employee effort. We apply panel data on about 4,000 heterogeneous employees in an international company and investigate how their effort—measured by paid and unpaid overtime hours—responds to two types of incentives: the flexibility of individual ratings over time in a given department and the frequency of surprise bonus payments in a department. We expect positive effort effects for both incentive instruments.

The results support the hypothesis that employee effort responds positively to surprise bonus payments. Similarly, employees provide more effort if their supervisors

[31] In additional estimations, we tested whether the results might be affected by the fact that some departments do not pay bonuses at all. Omitting employees from these departments from the sample, however, did not affect the results.

Table 4c. Linear Regressions without Observations of Departments that Existed Only Temporarily

	Overtime Hours	*Propensity to Provide More than 120 Hours of Overtime*
Indicators of Performance Pay Incentives in *t*–1:		
Std. Deviation of Rating Changes	2.050***	0.0048***
	(0.298)	(0.0012)
Bonus Payments	0.847***	0.0018***
	(0.141)	(0.0007)

Notes: The tables present estimated coefficients and standard errors in parentheses. The standard errors are clustered at the individual level. The models use the same specification as in column 4 in Table 3. In Table 4a, each row (except for the last two) presents the interaction terms generated in a separate regression for each effort indicator. In the restricted samples used in Table 4c, the department fixed effects are not identified since the individuals moving between departments were omitted. Table 4a uses 7,335 person-year observations and 42 different departments. Table 4b uses all 42 departments but a reduced number of individual observations as indicated, and Table 4c is based on 6,635 person-year observations and 32 departments.
***Statistically significant at the .01 level.

re-evaluate their performance anew from year to year as opposed to leaving individual positions unchanged over time. Both correlation patterns are robust to controls for department and individual fixed effects. What explains the positive response of employee effort to the higher probability of bonus payments and higher rating flexibility over time? We model employee behavior as a function of the marginal benefits effort. With increasing rating flexibility and a higher expected probability of bonus payments, the marginal payoff of high effort goes up such that, *ceteris paribus*, we expect to see more of it. This is what the analysis yields. Thus, the "contestability of ratings" generates the expected behavioral response and favoritism appears to bear a direct cost in terms of reduced effort. The effectiveness of the two incentive instruments seems to be substantial and of comparable magnitude: an increase in incentive intensity by one standard deviation beyond the mean is correlated with an increase in the propensity to provide unpaid overtime by more than 20%.

We corroborate the positive incentive effects of surprise bonuses and performance-oriented pay in numerous robustness tests that investigate different estimators, subsamples, and definitions of the incentive indicators. The effectiveness of the incentive mechanisms varies with the extent to which employee output is observable.

This is the first study to provide evidence on the incentive effects of surprise bonuses as well as to demonstrate the importance of flexible individual performance evaluations and the implicit cost of favoritism. The results confirm the predictions of MacLeod (2003) and Levin (2003), since performance is reduced in the presence of favoritism and can decline as a consequence of conflict.

REFERENCES

Aggarwal, Rajesh H., and Andrew A. Samwick. 2003. "Performance Incentives Within Firms: The Effect of Managerial Responsibility." *The Journal of Finance*, Vol. 58, No. 4, pp. 1613–49.

Baker, George. 2000. "The Use of Performance Measures in Incentive Contracting." *American Economic Review,* Vol. 90, No. 2, pp. 415–20.

Bandiera, Oriana, Iwan Barankay, and Imran Rasul. 2005. "Social Preferences and the Response to Incentives: Evidence from Personnel Data." *Quarterly Journal of Economics*, Vol. 120, No. 3, pp. 917–62.

____, ____, and____. 2006. "The Evolution of Cooperative Norms: Evidence from a Natural Field Experi-

ment." *Advances in Economic Analysis & Policy*, Vol. 6, No. 2, pp. 1–26.

Bertrand, Marianne, and Antoinette Schoar. 2003. "Managing with Style: The Effect of Managers on Firm Policies." *Quarterly Journal of Economics, Vol.* 118, No. 4, pp. 1169–1208.

Booth, Alison L., and Jeff Frank. 1999. "Earnings, Productivity, and Performance-Related Pay." *Journal of Labor Economics, Vol.* 17, No. 3, pp. 447–63.

Cable, John, and Nicholas Wilson. 1989. "Profit-Sharing and Productivity: An Analysis of UK Engineering Firms." *Economic Journal*, Vol. 99, No. 236, pp. 366–75.

____, and ____. 1990. "Profit-Sharing and Productivity: Some Further Evidence." *Economic Journal*, Vol. 100, No. 401, pp. 550–55.

Drago, Robert. 1991. "Incentives, Pay, and Performance: a Study of Australian Employees." *Applied Economics*, Vol. 23, No. 9, pp. 1433–46.

Engellandt, Axel, and Regina T. Riphahn. 2004. "Temporary Contracts and Employee Effort." *Labour Economics*, Vol. 12, No. 3, pp. 281–99.

Fehr, Ernst, and Klaus M. Schmidt. 2007. "Adding a Stick to the Carrot? The Interaction of Bonuses and Fines." *American Economic Review*, Vol. 97, No. 2, pp. 177–81.

Gibbs, Michael, Kenneth A. Merchant, Wim A. Van der Stede, and Mark E. Vargus. 2004. "Determinants and Effects of Subjectivity in Incentives." *The Accounting Review*, Vol. 79, No. 2, pp. 409–36.

Groves, Theodore, Yongmiao Hong, John McMillan, and Barry Naughton. 1994. "Autonomy and Incentives in Chinese State Enterprises." *Quarterly Journal of Economics*, Vol. 109, No. 1, pp. 183–209.

Haley, M. Ryan. 2003. "The Response of Worker Effort to Piece Rates. Evidence from the Midwest Logging Industry." *Journal of Human Resources*, Vol. 38, No. 4, pp. 881–90.

Holmstrom, Bengt, and Paul Milgrom. 1987. "Aggregation and Linearity in the Provision of Intertemporal incentives." *Econometrica*, Vol. 55, No. 2, pp. 303–28.

____, and ____. 1994. "The Firm as an Incentive System." *American Economic Review*, Vol. 84, No. 4, pp. 972–91.

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

Kahn, Lawrence, and Peter D. Sherer. 1990. "Contingent Pay and Managerial Performance." *Industrial and Labor Relations Review*, Vol. 43, No. 3, pp. 107-S–20-S.

Landers, Renée M., James B. Rebitzer, and Lowell J. Taylor. 1996. "Rat Race Redux: Adverse Selection in the Determination of Work Hours in Law Firms." *American Economic Review*, Vol. 86, No. 3, pp. 329–48.

Lazear, Edward. 2000. "Performance Pay and Productivity." *American Economic Review, Vol.* 90, No. 5, pp. 1346–61.

Lee, Michael, B. and Yinsog Rhee. 1996. "Bonuses, Unions, and Labor Productivity in South Korea." *Journal of Labor Research*, Vol. 17, No. 2, pp. 219–38.

Levin, Jonathan. 2003. "Relational Incentive Contracts." *American Economic Review, Vol.* 93, No. 3, pp. 837–57.

MacLeod, Bentley W. 2003. "Optimal Contracting with Subjective Evaluation." *American Economic Review*, Vol. 93, No. 1, pp. 216–40.

____, and Daniel Parent 1999. "Job Characteristics and Form of Compensation." *Research in Labor Economics*, Vol. 18, pp. 177–242.

Milgrom, Paul and John Roberts. 2004. "Bargaining Costs, Influence Costs and the Organization of Economic Activity." In *Controversies and Challenges in the New Institutional Economics*, Elgar Reference Collection. International Library of the New Institutional Economics, Vol. 7, pp. 271–312. Cheltenham, U.K. and Northampton, MA: Elgar.

Morton, Peter. 1998. "Annual Bonuses, Employment and Productivity Change in Taiwan's Manufacturing Sector." *Journal of Economic Behavior and Organization*, Vol. 37, No. 4, pp. 481–507.

Murphy, Kevin J. 1999. "Executive Compensation." In Orley Ashenfelter and David Card, eds., *Handbook of Labor Economics*, Vol. 3, pp. 2485–2563. Amsterdam: Elsevier.

____, and Paul Oyer. 2001. "Discretion in Executive Incentive Contracts: Theory and Evidence." Unpublished paper, Graduate School of Business Stanford University.

Parent, Daniel. 1999. "Methods of Pay and Earnings: A Longitudinal Analysis." *Industrial and Labor Relations Review*, Vol. 53, No. 1, pp. 71–86.

Pekkarinen, Tuomas, and Chris Riddell. 2008. "Performance Pay and Earnings Evidence from Personnel Records." *Industrial and Labor Relations Review*, Vol. 61, No. 3, pp. 297–319.

Prendergast, Canice. 1999. "The Provision of Incentives in Firms." *Journal of Economic Literature*, Vol. 37, No. 1, pp. 7–63.

____, and Robert Topel. 1993. "Discretion and Bias in Performance Evaluation." *European Economic Review*, Vol. 37, No. 2–3, pp. 355–65.

____, and ____. 1996. "Favoritism in Organizations." *Journal of Political Economy*, Vol. 104, No. 5, pp. 958–78.

Schwab, Donald P., and Craig A. Olson. 1990. "Merit Pay Practices: Implications for Pay–Performance Relationships." *Industrial and Labor Relations Review*, Vol. 43, No. 3, pp. 237S–55S.

Shearer, Bruce. 2004. "Piece Rates, Fixed Wags and Incentives: Evidence from a Field Experiment." *Review of Economic Studies*, Vol. 71, No. 2, pp. 513–34.

Sousa-Poza, Alfonso and Alexandre Ziegler. 2003. "Asymmetric Information about Workers' Productivity as a Cause for Inefficient Long Worker Hours." *Labour Economics*, Vol. 10, No. 6, pp. 727–47.

Taylor, Paul J, and Jon L. Pierce. 1999. "Effects of Introducing a Performance Management System on Employees' Subsequent Attitudes and Effort." *Public Personnel Management*, Vol. 28, No. 3, pp. 423–52.