# Design and Implementation of Pay for Performance

Michael Gibbs

I Z A

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

# Design and Implementation of
# Pay for Performance

**Michael Gibbs**
*University of Chicago Booth School of Business
and IZA*

# ABSTRACT

# Design and Implementation of Pay for Performance[*]

A large, mature and robust economic literature on pay for performance now exists, which provides a useful framework for thinking about pay for performance systems. I use the lessons of the literature to discuss how to design and implement pay for performance in practice.

Corresponding author:

Michael Gibbs
University of Chicago
Booth School of Business
5807 S. Woodlawn
Chicago, IL 60637
USA
E-mail: Michael.Gibbs@ChicagoBooth.edu

**I. INTRODUCTION**

Incentive compensation is an important tool for a firm trying to improve employee performance. If designed well, an incentive plan motivates the employee to work harder, smarter, and in better alignment with the firm's objectives. It can also improve recruitment and retention. If designed poorly, it can undermine these objectives. The design and implementation of pay for performance involves complex tradeoffs and has several interacting components. Fortunately, incentive compensation has been one of the most actively studied topics in economics in the last two decades, with related work in accounting and management. A large, mature and robust literature now provides a useful framework for designing and implementing pay for performance. Several excellent surveys of this research are available (e.g., Gibbons 1998; Murphy 1999; Prendergast 1999; Bushman & Smith 2001). The purpose of this chapter is not to provide yet another survey, but to use the lessons of the economic literature on incentives to discuss how to design and implement pay for performance in practice.

Compensation systems have several roles beyond creating incentives (e.g., Waggoner, Neely & Kennerly 1999; Ittner & Larcker 2002). This fact is most evident in performance evaluation. Firms may use an evaluation to monitor and reward effort. However, performance also depends on factors other than effort, such as the employee's abilities, training, information, and working relationship with colleagues. Firms may therefore use an evaluation for job assignments, promotions, to diagnose weaknesses in job design or personnel policies, and to provide feedback and coaching. I focus on incentives, but it is important to keep in mind that an incentive plan interacts with, and may undermine or reinforce, other personnel objectives.

The approach used in economics is called Agency Theory. Many situations involve agency relationships, such as a real-estate agent working on behalf of a home buyer, or a politician working on behalf of voters. Although the intuition of agency theory applies quite broadly, I focus on an employee working as an agent on behalf of the principal, his employer.

**Three Components of an Incentive Plan**

I will argue that a useful approach to designing an incentive plan is to consider three components, and two types of evaluation, in a particular sequence. I follow that sequence in this chapter; thus we walk through the logical steps of designing an incentive plan.

To illustrate the major points, I use a simple model of pay for performance that is similar to most of the literature.[1] The employee's effect on the employer's objective, "output" Q, depends on efforts $e_i$ in multiple tasks performed on the job, i = 1, …, n. The employee is risk and effort averse, with utility U = E[Pay] – ½R$\sigma_{Pay}^2$ – ½C$\sum_i e_i^2$. The second term adjusts the first term for risk aversion; combined they are the certainty equivalent value of risky income to the employee. R is the coefficient of absolute risk aversion. With our assumptions below, from the employee's perspective $\sigma_{Pay}^2$ does not depend on $e_i$, so the risk premium is independent of effort choice. C is the marginal disutility of effort. The employee chooses efforts $e_i$ to maximize U, subject to a participation constraint that U ≥ 0, with the employee's reservation utility (from leisure or alternative employment) normalized to zero.

The firm maximizes expected profit from the employee = E[Q – Pay] by using an incentive plan with three components, as in Figure 22.1. It chooses a *Performance Measure* P and ties that to pay. P must depend on efforts for any incentive to exist. Pay equals *Base Salary* S plus *Pay for Performance Relationship* I(P), Pay = S + I(P). The employee's optimal efforts on each task are found by solving:

$$max_{e_i}\ S + E[I(P)] - \tfrac{1}{2}R\sigma_{Pay}^2 - \tfrac{1}{2}C\sum_i e_i^2$$

$$\Rightarrow e_i^* = \frac{dI}{dP}E\left[\frac{\partial P}{\partial e_i}\right]/C,\ \forall i. \tag{1}$$

Equation (1) immediately provides useful insights about incentive-plan design. Base salary S does not appear – it is not a key driver of incentives because it does not vary with performance.[2] The role of base

---

[1] By "pay" and "rewards" throughout, I mean anything the employee values that the firm can give more or less of, as a function of performance. Monetary compensation is most important, but other examples include benefits, a nicer office, promotions, preferred job assignments, or access to training.

[2] This is a slight simplification, as base salary may be higher than in alternative jobs. In such a case, an employee who is terminated for poor performance suffers a loss. That differential is an additional way in which total compensation varies with perfor-

salary is to adjust overall compensation to labor market rates, so that the firm can maximize profits while attracting and retaining employees with appropriate skills. Thinking about base salary first when considering a pay plan is common, as it is a key issue in employment negotiations. However, it is the least important component of pay for incentives.

[INSERT FIGURE 22.1 HERE]

Equation (1) reveals the two components to emphasize in designing an incentive plan. The first is the *Incentive Intensity*, dI/dP, which is sometimes called the *shape* of the pay-performance relationship, because it is the slope in Figure 22.1. The second is how sensitive the performance evaluation is to effort, $\partial P/\partial e_i$. I will show that the effectiveness of the performance evaluation is the major determinant of the incentive intensity. Therefore, the first step in designing an incentive plan is performance evaluation, our next topic. Once the firm has chosen the performance evaluation method, and understands the properties of the evaluation, it can set how pay varies with performance.

**Two Methods of Performance Evaluation**

There are two general methods of performance evaluation: numeric and subjective. Examples of numeric performance measures include unit sales for a salesperson, stock price for a CEO, or a customer satisfaction index for a service representative. *Subjective Evaluation* uses supervisor discretion and judgment to gauge performance. An example is an annual merit rating (on a numeric scale or using verbal categories such as "Meets Expectations").

The distinction between numeric and subjective evaluation is not cut-and-dried, because subjectivity can arise in the evaluation P, pay for performance I(P), or both. Consider a salesperson whose performance is measured as unit sales *S*. However, suppose the bonus is not calculated by formula (such as *Bonus = b·S*, where *b* is the commission), but instead the supervisor uses her discretion to award a bonus. As a second example, consider an executive whose performance is quantified using two measures: profits

---

mance. Including such effects (sometimes called *efficiency wages*) is a simple intuitive extension of our discussion. Base salary might also imply income effects on the marginal utility of income and marginal disutility of effort.

and percentage growth in sales. His bonus is based solely on those two measures, but the weights on each are at the discretion of the supervisor. In both cases, the measures are numeric, but how they are used to calculate rewards is subjective. A third example arises when the reward's value cannot be quantified and so a formula cannot be stated. The value of a promotion is the discounted present value of uncertain increased compensation over an unknown future period.

With that caveat, let us first focus on the narrow question of whether a firm should use one or more numeric measures to quantify the employee's contributions, or instead use subjective evaluation. All else equal, a numeric evaluation is preferable. The employee knows exactly how his performance will be measured, and numeric measures usually can be tracked easily, so he will know where he stands at any point in time. By contrast, subjective evaluations raise issues of fairness, bias, and trust (see below). Furthermore, I will argue that firms often use subjective evaluation to complement numeric measures. For those reasons, the proper starting place is to quantify the employee's performance as well as possible, yielding one or more numeric measures, and analyze the properties of those measures to decide what weight should be placed on them for determining rewards. As a second step, the firm should then decide whether it might use subjective evaluation in addition to, or instead of, numeric performance measures.

## II. NUMERIC PERFORMANCE MEASURES: BASIC PROPERTIES

Because this section focuses on the term $E[\partial P/\partial e_i]$ in equation (1), assume incentive pay in its simplest form, a linear bonus with commission rate b: $I(P) = b \cdot P$; in fact, this form is extremely common in practice. I now show the incentive intensity b depends on several important properties of the performance measure.

Let output and the performance measure have the simple forms $Q = \sum_i q_i e_i + \mu$ and $P = \sum_i p_i e_i + \varepsilon$, where $q_i$ and $p_i$ are the marginal products of effort on Q and P, and $\mu$ and $\varepsilon$ denote the influence of random factors on each (uncontrollable by the employee or firm; both with mean = 0). Both sets of marginal products $q_i$ and $p_i$ may be random from the firm's perspective. However, I assume the employee knows the values of $p_i$ before choosing effort levels. Finally, a performance measure may be scaled

in many different units, with I(P) rescaling into dollars. As an example, a firm might use a customer satisfaction index as a measure, and pay a commission on the level of the index. To abstract from such scaling details, assume the vectors |q| and |p| are of equal length 1: $\sqrt{\sum q_i^2} = \sqrt{\sum p_i^2} = 1$.

With these further assumptions, (1) becomes

$$e_i^* = bp_i/C$$

$$\Rightarrow \quad \frac{de_i^*}{db} = p_i/C. \tag{2}$$

The latter part of (2) describes how efforts respond to changes in incentive intensity. The firm chooses b to maximize expected profits E[Q − Pay], given (2) and the employee's participation constraint. This is equivalent to maximizing total surplus:

$$max_b \; E[\sum q_i e_i] - \tfrac{1}{2}Rb^2\sigma_\varepsilon^2 - E[\tfrac{1}{2}C\sum e_i^2] \; \Rightarrow \; E[\sum_i q_i p_i/C] - Rb\sigma_\varepsilon^2 - E[b\sum_i p_i^2/C] = 0$$

$$\Rightarrow \quad b^* = \frac{E\sum q_i p_i}{E\sum p_i^2 + RC\sigma_\varepsilon^2}. \tag{3}$$

Equation (3) describes the optimal incentive intensity as a function of properties of the performance measure. Several cases illustrate those properties.

**Uncontrollable Risk and Distortion**

Risk in the employee's work environment has two important effects on incentives. Consider a single-task job so that P = p·e + ε. By (2), $P = (bp^2/C) + \varepsilon$, so the variance of performance is $\sigma_P^2 = (b/C)\sigma_p^2 + \sigma_\varepsilon^2$. Because I assume the employee observes his marginal products p before choosing effort, the first term is not risk from the employee's perspective. It corresponds to what I call controllable risk below. Consider the effect of the second, $\sigma_\varepsilon^2$, which I call uncontrollable risk.

***Case 1:*** P = Q, $\sigma_\varepsilon^2 = 0$. The firm measures output without error, and uses it as the performance measure. The optimal incentive intensity is b* = 1. This is the benchmark case providing first-best effort, $e_i^* = p_i/C$. All cases below are second best because of imperfections in the performance measure.

***Case 2:*** $P = Q$ (so that $\mu = \varepsilon$), $\sigma_\varepsilon^2 > 0$. The firm measures output, but with error. The optimal incentive intensity is $b^* = \frac{1}{1+RC\sigma_\varepsilon^2} < 1$. This illustrates a key theme in the early theoretical literature on incentives: the tradeoff between uncontrollable risk and incentives (Holmstrom 1979; Banker & Datar 1989). Because $b^* < 1$, $e^*$ is lower than in Case 1, as is expected output. Our first property of a performance measure is the extent of ***Uncontrollable Risk*** or measurement error. The greater the measurement error $\sigma_\varepsilon^2$, the weaker the strength of the incentive. Measurement error imposes risk on the employee to the extent that pay varies with the measure. Stronger incentives punish more for bad luck, and reward more for good luck ($\varepsilon \lessgtr 0$). A firm may be able to reduce measurement error by spending resources to measure performance more accurately. Doing so would allow it to provide stronger incentives, though at additional measurement costs. To the extent that the measure has error, the firm must compensate the employee via a risk premium in the base salary. Therefore, the firm lowers $b^*$, even though doing so reduces incentives, trading off incentives with salary and measurement costs. Empirical evidence on whether the incentive intensity tends to be lower in more risky environments is mixed. I discuss why that mixed evidence makes sense below, after introducing the concept of controllable risk.

***Case 3:*** $P \neq Q$, $\sigma_\varepsilon^2 > 0$, $q_i$ and $p_i$ are not random. The optimal incentive intensity is

$$b^* = \frac{\sum q_i p_i}{\sum p_i^2 + RC\sigma_\varepsilon^2} = \frac{\cos(\theta)}{1+RC\sigma_\varepsilon^2}. \tag{4}$$

$\text{Cos}(\theta)$ is the cosine of the angle between the vectors $|p|$ and $|q|$, a measure of the extent to which the marginal products of effort on the performance measure match those on output (Datar, Kulp & Lambert 2001; Baker 2002; Feltham & Xie 1994). Our second important performance measure property is ***Distortion***. Note that $\cos(\theta) < 1$ if any $p_i \neq q_i$, so $b^*$ is lower than in Case 2. If P is distorted compared to Q, the employee has incentives to provide too much or too little of some types of effort. The more distorted the performance measure is, the lower the optimal incentive intensity.

To see the effect of distortion, suppose a worker has two tasks: producing high-quality parts, and producing as many as possible per hour. Output is $Q = q_1 e_1 + q_2 e_2$, where the first term represents quantity and the second quality. $q_i$ represents the relative weight of each task in creating firm value. Similarly, P =

$p_1e_1 + p_2e_2$, and $p_i$ represents the weight of each task in the performance measure. If P gives relatively more weight to task 1, compared to Q, the employee has too strong an incentive to perform task 1 compared to task 2. This is an important intuition about performance measurement and incentive-plan design. An incentive plan's goal is not merely to motivate an employee to work harder, but also to balance motivation across different tasks (Holmstrom & Milgrom 1991). A good measure rewards relatively more for more highly valued tasks, and relatively less for less valued tasks.[3] Many observed problems with incentive systems can be attributed to imbalanced multitask incentives (Kohn 1975).

Virtually all performance measures distort incentives, because they are only proxies for the firm's objective. For example, revenue ignores costs. Profit includes revenue and costs but is distorted because it is measured over one period. A manager evaluated based on this year's profits has an incentive to avoid maintenance or investments in R&D, which involve costs today to generate future benefits. The one common performance measure that arguably might *not* distort incentives is a publicly traded firm's stock price (rescaled by multiplying by the number of shares outstanding). One important benefit of a well-functioning stock market is that it provides a free performance measure for executives. A stock market is a prediction market providing an estimate of the discounted present value of future cash flows accruing to stockholders. To the extent that the Efficient Markets Hypothesis (Fama 1991) holds, stock prices weight present and future profits correctly and produce an undistorted performance measure. Of course, stock price also has a great deal of uncontrollable risk, for example, macroeconomic conditions, so it is also a volatile performance measure.

Distortion of a performance measure creates the first of three types of "gaming" of an incentive system described in this chapter. The second is manipulation, which arises in Case 4.

---

[3] I assume the marginal disutility of effort is the same for all tasks. If not, a measure should give more weight to tasks that are less costly to the worker, because the firm has to compensate for the disutility of effort.

**Controllable Risk**

***Case 4:*** $P \neq Q$; $\sigma_\varepsilon^2 > 0$; $q_i$ and $p_i$ are random; the employee observes $p_i$ before he chooses effort, but the firm does not. Here I add two important new issues: the employee's marginal products of effort may be stochastic, and the employee often has better information than the supervisor about how to focus efforts at any point in time, because he is doing the job. The optimal incentive intensity is as in (3), which is hard to interpret, but the single-task case provides useful intuition. If $Q = q \cdot e$ and $P = p \cdot e$, our scaling assumption implies $E(q) = E(p) = 1$, and

$$b^* = \frac{Eqp}{Ep^2 + RC\sigma_\varepsilon^2} = \frac{1 + \rho\sigma_q\sigma_p}{1 + \sigma_p^2 + RC\sigma_\varepsilon^2}, \tag{5}$$

where $\rho$ is the correlation between q and p, $\rho \leq 1$, and $b^* < 1$. As in Case 3, distortion affects incentives: the smaller is $\rho$, the smaller is $b^*$.

Equation (5) illustrates a new property of performance measures: ***Controllable Risk***, $\sigma_p^2$. Because the employee observes the marginal product of effort on the performance measure before choosing effort, he can use that information to work harder when p is larger, and to work less when p is smaller. That reaction is the efficient response to variation in marginal products of effort, as long as P is reasonably correlated with actual output Q. This effect can raise or lower the optimal incentive intensity, depending on $\rho$, $\sigma_q^2$, and $\sigma_p^2$. For example, if the firm can use output as the performance measure as in Case 2, $b^* = \frac{1 + \sigma_q^2}{1 + \sigma_q^2 + RC\sigma_\mu^2}$, which is larger than $b^*$ in Case 2. The difference is that now the employee has valuable information about varying marginal products of effort. The firm increases the incentive intensity to motivate the employee to use that information to allocate effort efficiently. More generally, define $\alpha = \frac{\sigma_q}{\sigma_p}$ as a measure of the relative precision of the employee's information about how to do his job, and rewrite $b^* = \frac{1 + \alpha\rho\sigma_p^2}{1 + \sigma_p^2 + RC\sigma_\varepsilon^2}$. The more precise the employee's information about how his effort affects the performance measure, the stronger the incentive should be, all else equal. The effect is, of course, reinforced by less distortion or larger $\rho$.

This is important new intuition about incentive-plan design. A good evaluation motivates the employee to use his information, expertise, experience of events as they unfold at work, and so forth to allocate effort and make decisions in ways that promote the firm's objective. When the employee knows things the firm does not about how to do the job, the firm cannot provide proper guidance. However, a well-chosen performance measure can, rewarding more for high-value tasks and less for low-value tasks. In fact, this insight helps explain why firms use incentives and performance evaluation at all. If no environmental uncertainty were present, the firm might simply tell the worker what to do and how to allocate time to different tasks. It would then *monitor* the worker for compliance. However, that approach is effective only if the supervisor knows more than the employee about how to allocate effort optimally. If the employee knows more, the firm needs to motivate him to use his information on the firm's behalf, by choosing a measure that captures the effects of the employee's controllable risk.

Controllable risk is a relatively new idea in the principal-agent literature (Baker 1992; Prendergast 2002a, b). The role of risk is not as simple as the early literature suggested (Case 2). If the employee cannot control environmental uncertainty affecting performance, the firm lowers the incentive intensity. However, to the extent that the employee can control environmental uncertainty, raising the incentive intensity to motivate good decision making may be optimal (to the extent that the measure is not too distorted). This distinction explains why the empirical literature attempting to test whether greater risk leads to weaker incentives yielded mixed results: it failed to recognize the distinction between controllable and uncontrollable risk. More recent empirical work (DeVaro & Kurtulus 2010; Barrenechea-Méndez, Ortin-Ángel & Rodes-Mayor Eduardo 2011) does find incentives tend to be stronger when the employee has more controllable risk, and weaker when he has more uncontrollable risk.

Controllability is a term the managerial accounting literature has used, but not rigorously defined, for many years. It does not necessarily mean an employee can control whether an event occurs. Rather, it means the employee has some ability to control a random event's impact on firm value. Consider a factory roof that collapses under accumulated snow. The storm is a random event, but the plant manager can partially control the effect of the storm on firm value. The manager can take precautions against foreseea-

ble risks, such as proper maintenance. He can react to events when they occur, such as sending employees to the roof with shovels. Finally, he may be able to act after the event to mitigate damage, such as finding a roofing company that can make immediate emergency repairs so that production can resume quickly. Along similar lines, employees can take preparations and react profitably to exploit unexpected positive events.

As the collapsing-roof example illustrates, whether variability in the performance measure is controllable or uncontrollable is a subtle issue. Economic theory can help explain the distinction, but ultimately a firm must use some judgment to decide the extent to which risks are controllable or uncontrollable. In fact, I will argue this is a primary reason for the use of subjective evaluation.

**Manipulability**

Earlier I stated that distortion in a measure causes the first of three types of gaming of an incentive system. I refer to the second type as manipulation, a special case of controllable risk. Return to (5) but assume $\rho < 0$. This can occur if an employee can take actions that improve the measure but at the same time reduce firm value.[4] Consider Prendergast's (2002a) example of a surgeon whose performance measure is mortality rates. The surgeon has an incentive to use his expertise and examination of patients – information he possesses that his employer does not – to avoid operating on patients with the most complex and risky medical problems. Those may be the patients the hospital most hopes to serve. This example illustrates an additional property of a performance measure, ***Manipulability***. Manipulation is a special case of controllable risk, but distinguishing it is useful to highlight that an employee may be able to use his information not to increase firm value, but strategically to game the incentive system to increase rewards.

The literature has not consistently used the terms "gaming," "distortion," and "manipulation." In an attempt to clarify, I distinguish three ways an employee can game an incentive system (distortion and

---

[4] If $\rho < 0$, $b^*$ might be negative, in which case the firm uses the measure to punish undesirable outcomes. However, it might still be that $b^* > 0$ when $\rho < 0$ (Baker 1992). Incentive pay motivates total effort as well as allocation across tasks. When $\rho < 0$, the measure gives the wrong signals about how to allocate effort, but might be used anyway to motivate greater average effort.

manipulation are the first two), and distinguish between distortion and manipulation. Courty and Marsch-ke (2008) and Sloop and van Praag (2011) use "distortion" to refer to both distortion and manipulation (they do recognize the theoretical distinction). I define manipulation as different from distortion because the distinction is useful in designing and implementing an incentive plan. The firm takes both into account in ex-ante incentive-plan design by reducing the incentive intensity, choosing additional measures, or using subjective evaluation. It may also address manipulation ex post through subjective evaluation or implicit rewards as described below.

### III. NUMERIC PERFORMANCE MEASURES: OTHER CONSIDERATIONS

**Degrading of Performance Measures**

An interesting implication of distortion and manipulability is that numeric performance measures may degrade over time (Courty & Marschke 2004, 2008). Suppose a firm finds that historically P has been a good proxy for Q (marginal products of effort on P are highly correlated with those on Q). It implements a bonus based on this measure. However, the performance measure is distorted, giving too much weight to Task 1 and too little to Task 2. The employee's behavior would not have been distorted if no incentive had been in place beforehand. Placing an incentive on the measure, though, will change the employee's behavior, resulting in greater emphasis on Task 1, which means the measure will be less effective than anticipated. Such an effect is inevitable, because virtually all performance measures distort incentives in some way.

Manipulation causes additional degradation of a measure in a more subtle way. Incentives spawn creativity. Once rewards are based on a measure, the employee has an incentive not just to increase output, but also to learn how to better manipulate the measure. For example, a more experienced CFO may be better able to "manage earnings" (a familiar example of manipulation in accounting) than an inexperienced CFO.

Finally, one should expect a measure's effectiveness to degrade because the environment changes. If Q evolves over time, a P chosen because it matches the initial $Q_0$ becomes more distorted when it does

not evolve alongside Q. For all of these reasons, a firm should expect a performance measure to be somewhat less useful in practice than initially expected and to degrade gradually over time. A good incentive system needs regular updating of performance measures.

**Scope of the Performance Measure**

I described four performance measure properties: controllable and uncontrollable risk, distortion, and manipulation. In choosing a measure, a firm may face a tradeoff between these measures. Some argue a tradeoff is likely between uncontrollable risk and distortion, so that measures with low distortion tend to have high uncontrollable risk, and measures with high distortion tend to have low uncontrollable risk (Datar, Kulp & Lambert 2001; Van Praag & Cools 1991; Baker 1992). The intuition is that if only two relevant properties are present (uncontrollable risk and distortion), and measurement cost is concave in $\rho$ and $\sigma_p^2$, a firm would never choose a measure if an alternative is no worse on either dimension, and better on at least one. This logic implies a negatively sloped "performance measurement frontier" in which the firm trades off uncontrollable risk and distortion.

We can generalize this idea with the concept of performance measurement, **_Scope_**: the extent to which the measure is more narrowly focused or broadly defined. A measure's scope may vary on several dimensions. For example, profit is broader than revenue or cost, because it incorporates both of those measures. In a sense, profit includes more tasks in the performance measure than the other two. The scope of a profit measure could also be expanded along the dimension of time, say, three years instead of one, or a discounted present value style measure such as economic value added (EVA). Similarly, a measure's scope can vary from narrow to broad on the group of employees that are included, from individual to group, business unit, or firm-wide performance.

Measurement scope illustrates nicely the argument that a tradeoff exists between a measure's uncontrollable risk and distortion. Consider the performance measure with the broadest scope, stock price. It is the most common measure for CEOs of publicly traded companies, but most firms also use other measures such as earnings in CEO incentive contracts (Murphy 1999). This measure includes the effects

of all possible actions the employee can take that affect firm value, and therefore has little or no distortion. At the same time, it has a large amount of uncontrollable risk because it includes the effects of all co-workers, plus many other random variables. Because of this high risk, most firms also use other measures for CEO incentives and rarely use stock price as incentives for employees below the executive level.[5] A firm can filter out much of the uncontrollable risk by choosing a measure that is narrower in scope. Accounting earnings is such an example. Revenue and cost are far more controllable than stock price. However, the narrower focus generates greater distortion, because some employee actions are left out of the measure.

As a firm varies a performance measure's scope, it is also likely to face similar tradeoffs with controllable risk and manipulation. The possibility that controllable and uncontrollable risks are positively correlated across measures seems likely. Random events tend to be partly controllable and partly uncontrollable. In the roof-collapse example, the storm was an act of nature, but to some extent, the manager could control its effects on firm value. Even macroeconomic conditions may be partially controllable by a firm, through actions such as currency or interest-rate hedging, or varying production over the business cycle. If uncontrollable and controllable risks are positively correlated, broader measures will tend to have more of both, and narrower measures less. We have a second tradeoff, between the two types of risk. Narrower measures are also likely to be more manipulable. Manipulating a single line item is much easier than manipulating the overall size of a budget. When a measure is affected by a smaller number of variables, the effects of each tend to be larger, so manipulating one is more likely to be effective. In addition, because narrow measures tend to have less uncontrollable risk, random noise is less likely to overwhelm attempted manipulation.

Putting these arguments together, we see that in choosing a performance measure, a useful concept is the measure's scope. Once the firm decides which aspects of performance it wishes to incorporate

---

[5] This statement holds except in employee stock or option plans. These plans are a puzzle from the perspective of this chapter, because for lower-level employees, the performance measure has extreme uncontrollable risk, and the share of increased stock value accruing to an employee due to his actions is essentially zero. See Oyer and Shaefer (2005).

in the evaluation, it may have latitude in how broadly or narrowly to implement the measure along several dimensions. Broader measures tend to have less distortion, include more controllable risk, and be less manipulable. However, they also tend to have more uncontrollable risk. To reduce uncontrollable risk, almost all measures narrow their focus at the cost of more distortion, less controllable risk, and more potential for manipulation.

**Combining Multiple Performance Measures**

We saw that weaknesses in performance measures imply lower optimal incentive intensities. A firm can also address these weaknesses by using additional performance measures or subjective evaluation. Here I consider the use of additional measures. These can be combined into a single measure, with weights the firm chooses (e.g., profit combines revenue and cost). Managers can also use them separately by awarding multiple bonuses for different measures. Firms use both methods (Murphy 1992; Gibbs, et al. 2009). If we assume linear pay for performance, the two methods are equivalent. Banker and Datar (1989) showed that for a wide class of joint density functions for uncontrollable risk, measures can be aggregated linearly (pay may then be a non-linear function of the aggregate measure).

Holmstrom (1979) initially developed the "Informativeness Principle": an additional measure firms should use in an incentive plan if it provides marginal information about employee effort. Research has discussed the question of relative weights, primarily in the context of uncontrollable risk and distortion. Adding measures can reduce distortion, because measures usually emphasize different aspects of the job. A common example is the Balanced Scorecard, in which a firm collects a variety of measures including qualitative factors such as customer satisfaction (Kaplan & Norton 1996). Balanced Scorecards were developed to incorporate dimensions of performance that traditional measures tend to ignore. A problem with combining measures to reduce distortion is that they often vary substantially in uncontrollable risk. Measures with more uncontrollable risk should receive lower incentive intensities, and measure with more controllable risk should receive high incentive intensities. This implies an incentive system combin-

ing performance measures with different levels of risk will have imbalanced multitask incentives, or weak overall incentives in order to preserve balanced motivation (Holmstrom & Milgrom 1991).

Firms may also combine performance measures to reduce uncontrollable risk, if measurement errors are negatively correlated. Consider $P_3 = b_1 P_1 + b_2 P_2$, where $b_i$ is the weight on measure $P_i$, which has measurement error $\varepsilon_i$. Measurement error of $P_3$ is $\sigma_3^2 = b_1^2 \sigma_1^2 + b_2^2 \sigma_2^2 + 2 b_1 b_2 \sigma_{12}$. If $\sigma_{12} < 0$, lowering overall risk by combining measures may be possible. One important example is relative performance evaluation (Lazear & Rosen 1981; Gibbons & Murphy 1990), in which supervisors evaluate an employee relative to colleagues or some other reference group. This evaluation may be effective if measurement error is common to employees in the group, which relative comparison filters out. Once more, however, tradeoffs can exist. Relative evaluation may also distort incentives if the employee can take actions to affect the group against which he is compared or if he can cooperate with or sabotage colleagues (Lazear 1989).

Virtually no analysis exists regarding how controllable risk affects the optimal combination of performance measures in an incentive system, reflecting the relatively new nature of that concept. Speculatively, a positive correlation between the extent of controllable risk and the use of multiple measures seems likely. See the discussion of how incentives are related to job design below.

## IV. SUBJECTIVE PERFORMANCE EVALUATION

An important method to address shortcomings in performance measures is subjective performance evaluation. Thinking of *any* job in which subjective evaluation or supervisor discretion does not play some role in incentives is difficult. Managers implicitly reward or punish employees through job assignments, office location, threat of termination, promotion, and so forth. Formal bonus plans must occasionally be reset (e.g., measures degrade over time, or targets must be reset each year), which involves discretion. Finally, many plans make explicit use of discretion: firms determine evaluations subjectively, or use measures as inputs into discretionary raises and bonuses. In this section, I consider the important role of discretion in incentive systems, focusing on the costs and benefits of subjective performance eval-

uation. The discussion of subjectivity is much shorter than the discussion above about numeric evaluation, but that difference does not reflect the relative (un)importance of subjectivity. This section is shorter because it builds easily on the prior discussion, and because research on this topic is much more sparse.

A substantial literature discusses potential problems arising from subjective evaluation (Murphy 1992; Prendergast & Topel 1993; Murphy & Cleveland 1995; Prendergast 1999). Because the evaluation is at the supervisor's discretion, supervisor preferences and incentives may play a role in the evaluation. A supervisor might engage in favoritism toward some employees, or manipulate the evaluation process to reduce compensation costs. Such behavior might expose the firm to legal liabilities from lawsuits alleging discrimination or wrongful termination. Subjective evaluation can also distort incentives. The employee might act as a "Yes Man," behaving in ways the supervisor prefers and that correlate imperfectly with firm value (Prendergast 1993). That employee may try to manipulate the evaluation by making negative evaluations personally costly for the supervisor (Milgrom 1988; Milgrom & Roberts 1988). As a response to such influence costs, the supervisor might be too lenient or reduce the variance of evaluations (leniency and centrality biases). Supervisors might also have hindsight bias, holding the employee responsible for factors that are known by the supervisor ex post, but that the employee did not know at the time he performed the work. Finally, subjective evaluations have their own form of uncontrollable risk for the employee: they are difficult to verify and enforce contractually, so they require relational contracting and adequate trust of the supervisor (Baker, Gibbons & Murphy 1994).

Because virtually all jobs use subjectivity, it must have some benefits. A smaller literature considers these (e.g., Murphy & Oyer 2003; Gibbs et al. 2004). Subjective evaluations are an alternative means by which a firm can address limitations of numeric measures. Subjective evaluation allows the supervisor to incorporate her own observations about employee performance that are not reflected in a numeric measure, and allows her to do so dynamically and to incorporate ex-post settling up.

The supervisor can reduce distortions by incorporating other dimensions of performance (quality, customer satisfaction) into the overall evaluation. In principle, such factors might be quantified as in a Balanced Scorecard, but as noted earlier, such an approach often suffers from significant measurement

16

error. A supervisor's judgment may be more accurate. If so, she can give such dimensions of the job relatively more weight in the overall evaluation, balancing multitask incentives with stronger overall incentive intensity.

Subjective evaluation may allow for a more subtle treatment than the simple tradeoffs of uncontrollable risk with controllable risk, distortion, and manipulation in selecting numeric measures. A supervisor can reduce uncontrollable risk by "backing out" effects of adverse uncontrollable events from the evaluation, while holding the employee responsible for controllable risks by punishing poor decisions and rewarding good ones. Consider again the collapsed-roof example. The supervisor can evaluate the extent to which the storm was predictable, whether the manager adequately planned and prepared, how he reacted as the event unfolded, and so on. The supervisor can weigh the evidence and decide the factors for which to hold the manager responsible, and which to ignore. No measure chosen ex ante will be able to incorporate all of these considerations and adjust to the unique circumstances of a given situation. Subjective evaluations seem almost inevitable in such situations.

Discretion in an incentive system provides a way to reduce manipulation, because it allows for ex-post settling up. If a measure indicates high performance but the supervisor has other information or personal observations suggesting the performance was due in part to manipulation, she can reduce the evaluation or punish the employee. The threat of such punishment can deter manipulation. One way firms might implement this is through a discretionary cap on the total level of a bonus.

Discretion in incentive systems has additional benefits beyond mitigating weaknesses in numeric measures. One such benefit is flexibility. Changing a performance measure or bonus formula in the middle of the year is possible, but at some renegotiation cost. Subjective evaluations and discretionary rewards are, by definition, changeable over time, thereby allowing the supervisor to adapt the incentive system to changing economic conditions, or adjust it if errors in design are discovered.

Another benefit of subjectivity is that the evaluation can promote other goals. Consider again the manager of the factory with the collapsed roof. When the supervisor evaluates performance, she can identify ways to improve job design, resources, information, or training. She can coach the manager on what

he should have done differently, what to learn from the experience, and how he should change his behavior on the job. A firm should keep such multiple roles in mind when using subjective evaluations. For example, many firms schedule evaluations for employee coaching at different times of the year than evaluations for determining compensation and incentives, in order to try to minimize conflicts between the two goals (Murphy & Cleveland 1995).

Only a small empirical literature exists on the use of subjective evaluation or discretion in incentive systems (e.g., Hayes & Schaefer 2000; Ittner, Larcker & Meyer 2003; Murphy & Oyer 2003; Gibbs et al. 2004, 2009), presumably because quantifying the concepts is difficult. The evidence is consistent with the uses of discretion described above. For example, firms employ subjective evaluations to reduce short-term incentives implied by numeric measures, and to motivate employees to cooperate more with other organizational units. They use implicit rewards such as threat of termination to punish detected manipulation. Gibbs et al. (2004) found supervisors were more likely to award subjective bonuses if achieving a bonus target was difficult and had significant consequences if not met, and if the manager's department had an operating loss for the year. These findings indicate the use of subjectivity to filter out uncontrollable risk. A department's loss may be due to poor performance, but it may also be due to bad luck. Subjectivity allows the firm to punish the former but reward the latter in ways that may be impossible with numeric measures.

## V. PAY-PERFORMANCE RELATIONSHIP

The second step in designing an incentive plan is to set the pay-for-performance relationship I(P). In (1), incentives are driven by dI/dP, how pay varies with performance. Little systematic evidence exists on how firms set pay-performance shapes. Gibbs et al. (2009) summarize primary bonus contracts for auto dealership department managers. Two percent of them involve lump-sum bonuses, 6% have a floor, 2% have a cap, and 98% are simple linear functions. Interestingly, if an employee receives a second or third bonus, those are much more likely to have complex shapes such as lump sums or caps. Murphy (2001) finds executive incentive contracts make greater use of pay caps (80%) and lump sum rewards at perfor-

mance thresholds (20%). Both find all contracts are linear or piecewise linear. In general, pay-for-performance shapes tend to be quite simple, perhaps because they are easier for employees to understand, which improves trust in the incentive system. Banker and Datar (1989) and Holmstrom and Milgrom (1991) provide theoretical justifications for linearity.

I discuss properties of two common pay-performance shapes. Figure 22.2a illustrates an incentive plan in which the employee receives a lump-sum bonus if performance exceeds a *threshold* T, so $dI/dP = 0$ to the left and right of T, and $dI/dP = \infty$ when the employee's performance puts him right at the margin between winning and losing the reward. Figure 22.2b illustrates a plan in which the employee earns a linear bonus for performance beyond a threshold ("floor"), but pay is capped for performance above a second threshold.

[INSERT FIGURES 22.2a & 22.2b HERE – ON SAME PAGE IF THEY FIT]

Consider Figure 22.2a. A couple of problems may arise. First, if the firm sets the threshold either too high or too low, the employee may have little incentive (Roy (1952) famously called such behavior "goldbricking"). Second, if the employee is near the threshold T, he has a strong incentive to engage in the third method of gaming an incentive system – shifting output between periods in order to increase compensation (Healy 1985; Courty & Marschke 2004). Managers and salespeople with this type of incentive plan often "sandbag" their performance, delaying sales until next year if they have already met the threshold, or trying to push customers to accelerate purchasing if they have not yet met the threshold (Oyer 1998). Both of these types of behaviors can have negative effects on firm value. In addition, if the performance measure is manipulable, the temptation to manipulate is strong when performance is close to T. Similar effects can arise if the incentive intensity (slope) changes less dramatically at a certain level of performance. The general lesson is that smoother pay-performance shapes, particularly simple linear relationships, are less likely to cause problems.

Promotion-based incentives have a similar pay-performance shape, because the reward is all or nothing if the employee does or does not meet the promotion criteria T. That form of incentive can have surprising effects on performance evaluation. If a supervisor provides accurate feedback, employees have

strong incentives if they are near T, but may slack off if they are below or above it. This effect would be most severe for employees performing below T, because reducing effort would push them further from T, whereas the opposite would occur for those above T. For these reasons, supervisors may give *less informative* evaluations, and particularly avoid giving negative feedback (Hansen 2010). These observations may help explain the observation that performance-rating distributions almost always exhibit leniency and centrality bias, especially when promotion stakes are high.

Given the negative effects of thresholds, why are they used? One reason is that they may allow the firm to use a stronger incentive intensity while keeping expected compensation at competitive rates or avoiding risks of high turnover in years when uncontrollable risk yields low performance (Oyer 2000; Murphy 2001). An additional benefit may be that paying a base salary for low performance reduces the employee's risk aversion – employees are most averse to large downside risks – which may motivate them to take more chances on the job. A subtle agency problem that has received almost no attention is conflicts between the firm and employee over risk taking. Firms are generally less risk averse than employees, so employees may be more cautious than is efficient. CEOs, for example, might pursue diversification strategies in order to reduce uncertainty in their compensation, even though shareholders prefer focused strategies. By limiting downside compensation risk as in Figure 22.2b, a firm may be better able to motivate employees to experiment and take risks on the job. Conversely, if a firm wishes to motivate employees to be conservative (e.g., risk management in a bank), it might use a pay-performance shape that limits upside compensation potential and reduces pay when performance is low.

Some firms use caps on incentives as illustrated in Figure 22.2b. To understand the tradeoffs in using a cap, remember two things about a performance measure. First, the measure is a function of the employee's effort, but also his ability and human capital. Second, a performance measure may be affected by manipulation and by measurement error / uncontrollable risk. A problem with a cap is that it is more likely to affect employees with high levels of skills and talent; therefore, it might adversely affect employee selection. A benefit of a cap, however, is that in many cases, high levels of a performance measure are more likely to reflect measurement error or manipulation than high levels of talent or effort. If an em-

ployee's performance is unusually good, he may simply have gotten lucky – or he may have cheated in some way. If odds are low that unusually high performance is due to effort and talent, a cap may be useful to deter manipulation and avoid rewarding good luck. Finally, a firm might also use a subjective cap, reserving the right to limit a bonus payout if the firm deems high performance to be due to luck or manipulation. However, although such a practice makes sense in theory, it runs the risk of a ratchet effect (see next section).

## VI. IMPLEMENTATION

In this section, I briefly discuss implementation of pay for performance. The first issue arises in implementing nearly all incentive plans. The last three involve ways in which pay for performance interacts with other personnel policies and objectives.

### The Ratchet Effect

Every incentive plan needs to be changed from time to time. Measures and targets degrade. Errors in design need correcting. Strategy and structure change, requiring focus on different activities. Changing an existing plan risks a *ratchet effect*. This effect occurs when an employee's expectations about future changes in the incentive system reduce incentives today, and it can occur for several reasons.

Suppose a firm sets commission rates for salespeople by determining expected sales and then dividing expected sales into the going rate for workers with similar skills. The firm uses current sales as one factor in setting expected sales for next year. This practice implies a ratchet effect because if the employee increases sales this year, he receives a lower commission next year. Similarly, growth-based targets or performance measures would seem to cause ratchet effects automatically. An additional cause may be a desire to lower labor costs. In our example, if the firm decides employees earn "too much," it lowers the commission. However, an employee may have high income because he has high performance due to strong talent or effort. The desire to reduce labor costs may end up punishing both.

Ratchet effects have received surprisingly little empirical study beyond Roy's (1952) famous description (Levenson et al. 2012). Some theoretical analysis exists (Lazear 1986; Gibbons 1987). Given the

many parameters of an incentive system that may need to be changed, ratchet effects are likely a common concern in implementing incentive systems. The famous Lincoln Electric case (Milgrom & Roberts 1995) provides insights into how a firm might reduce ratchet effects. The company has an explicit policy of not changing piece rates unless a change occurs in methods of production. Piece rates are set by a special department with great expertise at setting accurate rates (they set approximately 10,000 piece rates per year, according to a company executive I know). Employees are allowed to challenge piece rates, and their complaints are taken seriously. Expected compensation is pegged to wage data from the Bureau of Labor Statistics, eliminating discretion over one determinant of piece rates. Top managers have strong pay for performance, which reduces their temptation to lower compensation for employees in ways that reduce firm performance.

Perhaps most importantly, the firm makes effective use of relational contracting. Founder James Lincoln had a passionate belief in individual effort and reward – he even wrote pamphlets about the topic. His strong beliefs imprinted on the company's culture, which emphasizes hard work and high rewards based on individual output. This culture is preserved partly by a policy of promotion from within, so that most executives started in entry-level jobs. The firm has not varied incentive policies significantly in over 75 years. This history gives the firm a strong reputation with employees, who are thereby unlikely to worry high performance will be punished by ratcheting of the incentive system.

**Matching Incentives and Job Design**

Pay for performance does not operate in a vacuum, but interacts with other personnel policies and goals. One important example is job design. Holmstrom & Milgrom (1991 note that one way to deal with the problem of imbalanced multitask incentives arising from tasks with large differences in uncontrollable risk is to split the job, bundling easy-to-measure tasks in one job and hard-to-measure tasks in another. Prendergast (2002a) argues that delegation and multitasking are likely to have important effects on optimal performance evaluation. The degree of controllable risk in a job is, to some extent, within the firm's choosing. Consider two different work environments. In the first, work is largely the same from one day

to another, so there is little controllable risk. In the second, unexpected situations arise, and the optimal allocation of effort changes daily, so there is high controllable risk. In the first case, the firm has a good idea of what the employee should do, and delegates few decisions. It tells the employee what to do and monitors compliance, so little or no pay for performance is needed. In the second case, the worker possesses information about how to best perform the job, as in Case 5. To exploit that information, the firm delegates decisions, and uses pay for performance to motivate good decision making. Thus jobs with greater delegation are likely to have stronger pay-for-performance links.

Evaluation methods should also vary with the degree of delegation. I argued above that broader measures or use of multiple measures may be required if controllable risk is important. In such situations, a firm may also emphasize outputs instead of inputs in evaluation. Input-based measurement makes sense when the firm has a good idea about how the employee should do the job. Output-based measurement is better when the firm does not know how to guide employee efforts. Finally, subjective evaluation is likely to be more important for jobs with more delegation and controllable risk. It allows the firm to better incorporate controllable risk in the evaluation while filtering out controllable risk. It allows for ex-post consideration of new information in evaluating the employee, whereas the firm must choose a numeric measure in advance. Finally, jobs with high controllable risk and delegation are more likely in changing environments, and subjective evaluation provides greater flexibility.

Finally, Prendergast (2002a) argues the degree of multitasking in a job is also likely to be positively related to the use of pay for performance. The earlier literature on multitasking suggested the opposite might be true, because of the difficulty of balancing incentives across tasks with different uncontrollable risk. His counterargument is that jobs with more tasks are more costly to monitor, which might tip the balance toward incentives. An additional argument is suggested by Gibbs, Levenson & Zoghi (2010), who also provide evidence of a strong positive correlation between delegation and multitasking (but do not have data on incentives). A more complex job with multiple tasks may create greater worker knowledge that can drive continuous improvement. Continuous improvement in turn may require delega-

tion so that the worker can use his information to test and implement new methods. Delegation would then be complementary with incentive pay.

Because this area of study is new, the empirical literature is small but growing (e.g., Brown 1990; MacLeod & Parent 1998; Ortega 2009; DeVaro & Kurtulus 2010; Barrenechea-Méndez, Ortin-Ángel & Rodes-Mayor 2011) and is generally consistent with the arguments made here.

**Motivating Creativity**

A special aspect of how incentives match job design is motivation of creativity. Psychologists argue that pay for performance undermines creativity (Kohn 1975). To an economist, such a claim seems odd, given the enormous innovation the profit motive generates in markets. This topic is too broad to cover fully, but I provide brief remarks. When used properly, incentives do not necessarily undermine creativity.

A distorted performance measure, however, can easily undermine creativity. Rewarding a professor for the quantity of published articles distorts incentives toward publication in low-quality journals. In extreme cases, the professor might even manipulate the system by setting up his own journal. Most examples of incentives undermining creativity stem from distorted numeric measures. Quantifying creative activities is often hard; therefore they are likely to receive relatively low weight in numeric measures. Where creativity is important, a firm should consider broader, output-based measures and subjective evaluation, much as discussed in the prior sub-section.

The pay-for-performance relation also matters. Creativity requires experimentation with new methods, which implies mistakes. Reducing the employee's downside risk from mistakes will counteract employee risk aversion, leading to more willingness to experiment. A firm might reduce downside risk for the employee in several ways, including job-security provisions, a relatively high base salary for performance below some target level of performance, or discretionary bonuses to "back out" the effects of mistakes on formal bonuses. All incentives amount to ways to make the pay-performance shape similar to

that of a call option. This may be one reason why stock options are a common form of compensation in small startup firms, where innovation is particularly important.

**Pay for Performance and Employee Selection**

Performance depends on effort, but also on innate talent and accumulated human capital. Generally speaking, stronger pay for performance is likely to improve employee selection, because it provides stronger rewards to those whose ability and skills better match the job. Lazear (2000) studied the effects of a new pay-for-performance plan on the productivity of windshield installers. Productivity rose dramatically, and he estimated that approximately half of this increase was a result of selection (the other half to motivation). More research is required to gauge whether this strong selection effect of pay for performance occurs more broadly. Nevertheless, it does suggest a firm should consider potential selection benefits when designing a pay-for-performance plan.

Employee selection is affected not just by incentive intensity, but also by performance evaluation. Bouwens and van Lent (2006) find that selection improves when a firm uses performance measures that are less distorted and have less uncontrollable risk. This finding makes sense, as stronger incentives increase the links between the employee's performance, pay, and firm objectives. Helliwell and Huang (2010) find employees place a large value (compensating differential) on the extent to which they trust management. Together these findings suggest that how subjective evaluation and discretion in incentive systems are implemented is likely to have important effects on recruitment, turnover, and labor costs, in addition to motivation.

## VII. FUTURE RESEARCH

In this section, I sketch some areas for future research briefly, due to limited space. Some questions build on prior research in straightforward ways. Others involve collecting new data. Finally, many interesting questions involve thinking about incentives as a *system*. Most research focuses on a specific piece of an incentive plan, such as a single performance measure or bonus. However, we have seen that an incentive plan is complex, involving several instruments, with formal and informal components. Com-

plementarities or tradeoffs exist between most of these pieces. Incentives also arise implicitly, through promotions, career concerns, and so forth, all of which will interact with formal incentives. Finally, compensation and evaluation interact with other personal policies and objectives, notably job design and employee selection. A great deal of interesting research is yet to be done on how the various pieces of an incentive system fit together, and with other parts of organizational design.

**Worker Preferences**

Researchers rarely give much thought to the utility assumptions behind agency models, but doing so might be interesting. Many models are of the form employed here, with a risk-aversion adjustment for income variance. However, controllable risk implies effort variance. Even if an employee observes his marginal products before choosing effort, effort is still uncertain at the stage where the incentive contract is established, and so in principle should affect base salary. I am unaware of any evidence on whether effort variance affects compensation. For employees, being effort risk averse creates a cost of controllable risk.

Also of interest is whether diminishing marginal utility of income affects effort. If employees enjoy high payoffs on exercise, this issue may be relevant for companies that use employee stock options. It may be especially relevant for CEOs and executives with extremely high remuneration. The growing use of such compensation instruments, and of income levels for highly skilled workers in advanced economies, suggests this issue might be worth study.

**Performance Evaluation**

A great deal of research has been devoted to evaluation (especially numeric measures), but many important questions are not fully understood. How prevalent is manipulation (as distinguished from distortion)? How can a firm evaluate whether an employee is more likely to use controllable risk to increase firm value, or to manipulate the measure? How do firms address manipulation in practice? Finally, is manipulation by the firm in reporting performance also a concern?

Another interesting question is whether improved performance evaluation alone, without explicit incentives, might improve employee performance. Employees may have intrinsic motivation and also some extrinsic motivation from implicit incentives such as promotions. Performance evaluation can provide the employee with training and with feedback about the effects of his actions and decisions. Training and feedback could drive continuous improvement and human capital acquisition, even if no explicit incentives were applied by the firm.

Limited empirical research exists about how firms address flaws in performance measures, beyond reducing the incentive intensity. Presumably firms also use multiple measures, including subjectivity and implicit incentives, and may alter the shape of the pay-performance relationship. Linking such practices to properties of performance measures would be quite interesting.

The performance measurement frontier – tradeoffs faced when choosing performance measures – is an interesting question, though studying it may be difficult because inferring much about measures that firms do not choose is difficult. A good starting place for research would be to focus on a specific type of job where evaluation is simple, such as CEOs or sales. Studying measures that vary in scope along several dimensions may also be a useful approach.

I emphasized that subjectivity and discretion are almost inevitable and necessary in the design and implementation of incentives. Observation suggests they are in fact ubiquitous. Despite this fact, most research focuses on formulaic bonuses and numeric measures, likely because such compensation schemes are more amenable to formal modeling and data collection. However, widespread use of subjectivity suggests economists devote more effort to understanding this topic in theory and practice. Many issues are not fully understood. How prevalent are subjective evaluations? In what types of jobs? How are they conducted? What is the relative importance for incentives of subjectivity, discretion, and implicit rewards compared to explicit, numeric approaches?

An interesting issue discretion raises is incentives for the supervisor. Instead of a simple principal-agent problem, incentives for the evaluator matter (Murphy 1992). Several papers cited in this chapter consider this issue theoretically, but I am unaware of empirical research on how supervisor incentives af-

fect incentives for subordinates. Investigating the effects of supervisor incentives on the subordinate's evaluation method, distribution of performance ratings, incentive intensity, and effectiveness of relational contracting between the firm and the employee would be interesting.

**Group Incentives**

Group incentives appear prevalent (including gain sharing, profit sharing, options, and stock ownership plans), yet explaining them with standard agency theory is hard. Recall from (1) that $e_i^* = \frac{dI}{dP} E\left[\frac{\partial P}{\partial e_i}\right]$. Incentives require that efforts have tangible effects on the performance measure and that income varies reasonably with the measure. Group-based pay seems to yield low values for each. Performance is based on the actions of all employees, so the evaluation has a great deal of uncontrollable risk and a low signal-to-noise ratio. The tie of income to the evaluation is also usually small. For example, imagine an employee participates in profit sharing in a company with 5,000 employees, and by his own efforts increases profits by $1 million. On average, his share of his contribution is $200, or b = 0.0002. Putting these together, we see that it is implausible that such schemes might actually generate much incentive. A few studies have analyzed profit-sharing plans, but systematic evidence on how productivity varies with incentives in such systems would be revealing. Firms may adopt these policies for other purposes (e.g., implicit risk sharing, or to reduce variability in cash flows through a business cycle).

**Methodology**

Two recent advances in economics seem interesting for the study of pay for performance: behavioral economics and field experiments. Incentives involve individual worker behavior, supervisor judgment, and relational contracting. These topics may all benefit from study by behavioral economists. In addition, economists are making greater use of field experiments. Firms sometimes experiment with changes in pay-for-performance plans, and studying such examples to gain a more precise understanding of the effects of incentive-plan design would be of great interest. Lazear's Safelite study (2000) is an early example of such.

## VIII. CONCLUSIONS

I briefly conclude by summarizing the implications of research on incentives for designing and implementing pay for performance in practice. Incentive compensation is one of the most successful areas in economic research. An enormous body of research exists, from highly abstract and technical to empirical and descriptive. A healthy interplay between theory and empirics results in an advanced understanding of key issues and tradeoffs, and provides a structured way of thinking through the design and implementation of incentives.

A useful way to think about incentives is to use the structure of this chapter. Firms can design an incentive plan in three stages. First choose the performance evaluation method, including one or more numeric performance measures and subjective evaluation. This issue is by far the most important and complex to address in designing incentives. Most problems derive from the method of evaluation. When considering measures, a firm should account for uncontrollable and controllable risk, distortion, manipulability, and scope. The evaluation should, fundamentally, be matched to the job design. Second, establish the pay-for-performance relationship. This process involves setting weights or incentive intensities for bonus plans, choosing any performance targets, and deciding whether the incentive plan should have floors or caps. Third, set base salary to adjust for risk, effort, and prevailing labor market rates.

Importantly, an incentive plan is a *system* of interrelated parts. Understanding the individual pieces, such as the properties of a specific performance measure, is easy. Understanding how the pieces relate to each other is also necessary. One performance measure may help reduce distortion or uncontrollable risk in one dimension and raise them in another. Subjective evaluation will probably be necessary to detect and deter manipulation, encourage greater risk taking, or encourage the employee to take more initiative in using his knowledge to promote firm objectives. A measure may have a great deal of uncontrollable risk, implying the firm may want to put in a floor on incentive compensation to reduce risk. The measure might be manipulable, so that an explicit or discretionary cap may be useful.

Finally, an incentive plan has both formal and informal elements. The formal elements are the natural starting point, but thinking carefully about the informal elements is important. Subjective evaluation plays a critical role in improving the evaluation by filtering out certain risks while including others, reducing distortions, and deterring manipulation. This requires effort by the firm to avoid favoritism and other biases, training and incentives for supervisors so that evaluations are effective, and development of reasonable trust of the firm by employees. Effective relationship contracting can also reduce concerns about ratchet effects. For these reasons, careful implementation is important for an incentive plan to realize maximum potential.

**REFERENCES**

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

Baker, George P. 2002. "Distortion and Risk in Optimal Incentive Contracts." *Journal of Human Resources* 37(4):729-751.

Baker, George P., Robert Gibbons, and Kevin J. Murphy. 1994. "Subjective Performance Measures in Optimal Incentive Contracts." *Quarterly Journal of Economics* 109:1125-1156.

Banker, Rajiv, and Srikant Datar. 1989. "Sensitivity, Precision, and Linear Aggregation of Signals for Performance Evaluation." *Journal of Accounting Research* 27:21-39.

Barrenechea-Méndez, Marco A., Pedro Ortin-Ángel, and Eduardo Rodes-Mayor. 2011. "Uncertainty, Job Complexity and Incentives: An Empirical Analysis of Spanish Industrial Firms." Working paper, Universitat Autònoma de Barcelona.

Bouwens, Jan, and Laurence van Lent. 2006. "Performance Measure Properties and the Effect of Incentive Contracts." *Journal of Management Accounting Research* 18:55-75.

Brown, Charles. 1990. "Firms Choice of Method of Pay." *Industrial and Labor Relations Review* 43(3):165-182.

Bushman, Robert M., and Abbie J. Smith. 2001. "Financial Accounting Information and Corporate Governance." *Journal of Accounting and Economics* 32:237-333.

Courty, Pascal, and Gerald Marschke. 2004. "An Empirical Investigation of Gaming Responses to Explicit Performance Incentives." *Journal of Labor Economics* 22(1):23-56.

Courty, Pascal, and Gerald Marschke. 2008. "A General Test for Distortions in Performance Measures." *Review of Economics and Statistics* 90(3):428-441.

Datar, Srikant, Susan Cohen Kulp, and Richard Lambert. 2001. "Balancing Performance Measures." *Journal of Accounting Research* 39(1):75-92.

DeVaro, Jed, and Fidan Ana Kurtulus. 2010. "An Empirical Analysis of Risk, Incentives and the Delegation of Worker Authority." *Industrial and Labor Relations Review* 63(4):637-657.

Fama, Eugene. 1991. "Efficient Capital Markets: II." *Journal of Finance* 46(5):1575-1617.

Feltham, Gerald, and Jim Xie. 1994. "Performance Measure Congruity and Diversity in Multi-Task Principal-Agent Relations." *The Accounting Review* 69:429-453.

Gibbons, Robert 1987. "Piece-Rate Incentive Schemes." *Journal of Labor Economics* 5:413-429.

Gibbons, Robert. 1998. "Incentives in Organizations." *Journal of Economic Perspectives* 12(4):115-132.

Gibbons, Robert, and Kevin J. Murphy. 1990. "Relative Performance Evaluation for Chief Executive Officers." *Industrial and Labor Relations Review* 43:30-52.

Gibbs, Michael, Alec Levenson, and Cindy Zoghi. 2010. "Why Are Jobs Designed the Way They Are?" *Research in Labor Economics* 30:107-154.

Gibbs, Michael, Kenneth Merchant, Mark Vargus, and Wim Van der Stede. 2004. "Determinants and Effects of Subjectivity in Incentives." *The Accounting Review* 79(2):409-436.

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

Hansen, Stephen. 2010. "The Benefits of Limited Feedback in Organizations." Working paper, Universitat Pompeu Fabra.

Hayes, Rachel M., and Scott Schaefer. 2000. "Implicit Contracts and the Explanatory Power of Top Executive Compensation." *Rand Journal of Economics* 31:273-293.

Healy, Paul. 1985. "The Effect of Bonus Schemes on Accounting Decisions." *Journal of Accounting and Economics* 7:85-107.

Helliwell, John F., and Haifang Huang. 2010. "How's the Job? Well-Being and Social Capital in the Workplace." *Industrial and Labor Relations Review* 63(2):205-227.

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

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

Holmstrom, Bengt, and Paul Milgrom. 1994. "The Firm as an Incentive System." *American Economic Review* 84(4):972-991.

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

Kaplan, Robert S., and D.P. Norton. 1996. *The Balanced Scorecard: Translating Strategy into Action.* Boston, MA: Harvard Business School Press.

Kohn, Alfie. 1975. "On the Folly of Rewarding A, While Hoping For B." *Academy of Management Journal* 18:769-783.

Lazear, Edward P. 1986. "Salaries and Piece Rates." *Journal of Business* 59(3):405-431.

Lazear, Edward P. 1989. "Pay Equality and Industrial Politics." *Journal of Political Economy* 97(3):561-580.

Lazear, Edward P. 2000. "Performance Pay and Productivity." *American Economic Review* 90(5):1346-1361.

Lazear, Edward P., and Sherwin Rosen. 2001. "Rank Order Tournaments as Optimal Labor Contracts." *Journal of Political Economy* 89(5):841-864.

Levenson, Alec, Cindy Zoghi, Michael Gibbs, and George Benson. 2012. "Optimizing Incentive Plan Design: A Case Study." Working paper, USC.

MacLeod, W. Bentley, and Daniel Parent. 1999. "Job Characteristics and the Form of Compensation." *Research in Labor Economics* 18:177-242.

Milgrom, Paul. 1988. "Employment Contracts, Influence Activities, and Efficient Organization Design." *Journal of Political Economy* 96:42-60.

Milgrom, Paul, and John Roberts. 1995. "Complementarities and Fit: Strategy Structure, and Organizational Change in Manufacturing." *Journal of Accounting and Economics* 19:179-208.

Murphy, Kevin J. 1992. "Performance Measurement and Appraisal: Motivating Managers to Identify and Reward Performance." In *Performance Measurement, Evaluation, and Incentives*, edited by William J. Bruns. Boston: Harvard Business School Press.

Murphy, Kevin J. 1999. "Executive Compensation." In *Handbook of Labor Economics*, edited by Orley Ashenfelter and David Card, ed.1 v.3:2485-2563. New York: North Holland.

Murphy, Kevin J. 2000. "Performance Standards in Incentive Contracts." *Journal of Accounting and Economics* 30: 245-278.

Murphy, Kevin J., and Paul Oyer. 2003. "Discretion in Executive Incentive Contracts: Theory and Evidence." Working paper, USC.

Murphy, Kevin R., and J.N. Cleveland. 1995. *Understanding Performance Appraisal: Social, Organizational, and Goal-Based Perspectives.* Thousand Oaks, CA: Sage Publications.

Ortega, Jaime. 2009. "Employee Discretion and Performance Pay." *The Accounting Review* 84(2):589-612.

Oyer, Paul. 1998. "The Effect of Sales Incentives on Business Seasonality," *Quarterly Journal of Economics* 113(1):149-186.

Oyer, Paul. 2000. "A Theory of Sales Quotas with Limited Liability and Rent Sharing." *Journal of Labor Economics* 18(3):405-426.

Oyer, Paul, and Scott Schaefer. 2005. "Why Do Some Firms Give Stock Options to All Employees? An Empirical Examination of Alternative Theories." *Journal of Financial Economics* 76:99-133.

Prendergast, Canice. 1993. "A Theory of 'Yes Men.'" *American Economic Review* 83:757-770.

Prendergast, Canice, and Robert Topel. 1993. "Discretion and Bias in Performance Evaluation." *European Economic Review* 37:355-365.

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

Prendergast, Canice. 2002a. "The Tenuous Trade-off between Risk and Incentives." *Journal of Political Economy* 110(5):1071-1102.

Prendergast, Canice. 2002b. "Uncertainty and Incentives." *Journal of Labor Economics* 20(2 pt. 2):S115-S137.

Roy, Donald. 1952. "Quota Restriction and Goldbricking in a Machine Shop," *American Journal of Sociology* 57(5):427-442.

Sloof, Randolph, and Mirjam van Praag. 2011. "Testing for Distortions in Performance Measures: An Application to Residual Income Based Measures Like Economic Value Added." Working paper, University of Amsterdam.

Van Praag, Mirjam, and Kees Cools. 2001. "Performance Measure Selection: Aligning the Principal's Objective and the Agent's Effort." Working paper, University of Amsterdam.

Waggoner, Daniel B., Andy D. Neely, and Mike P. Kennerly. 1999. "The Forces That Shape Organisational Performance Measurement Systems: An Interdisciplinary Review." *International Journal of Production Economics* 60-61:53-60.
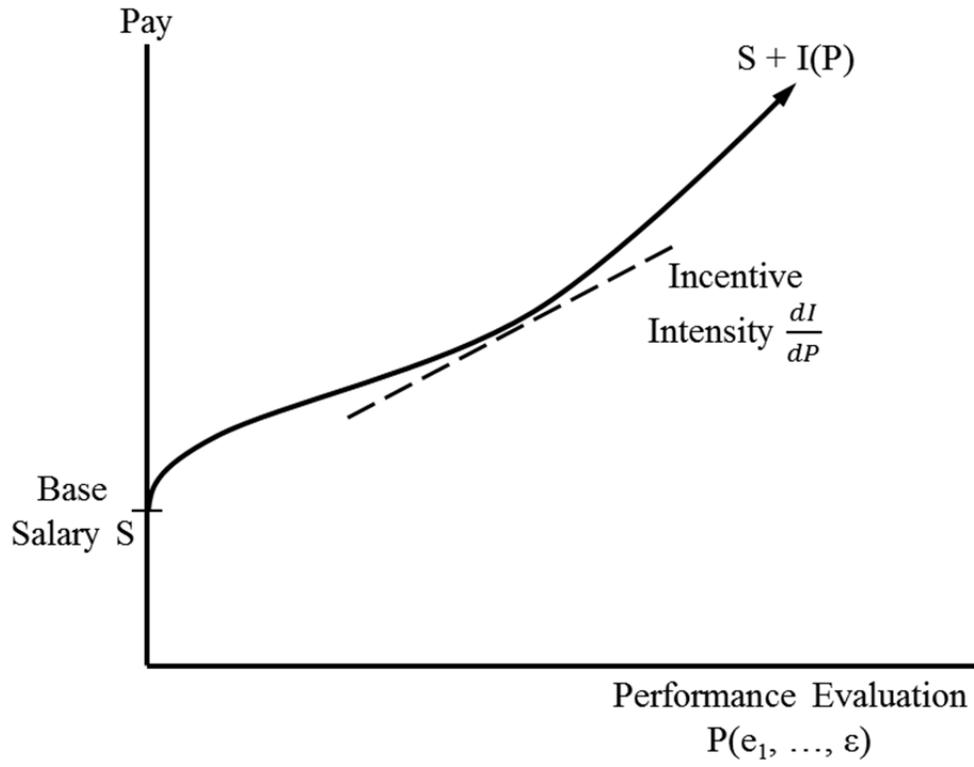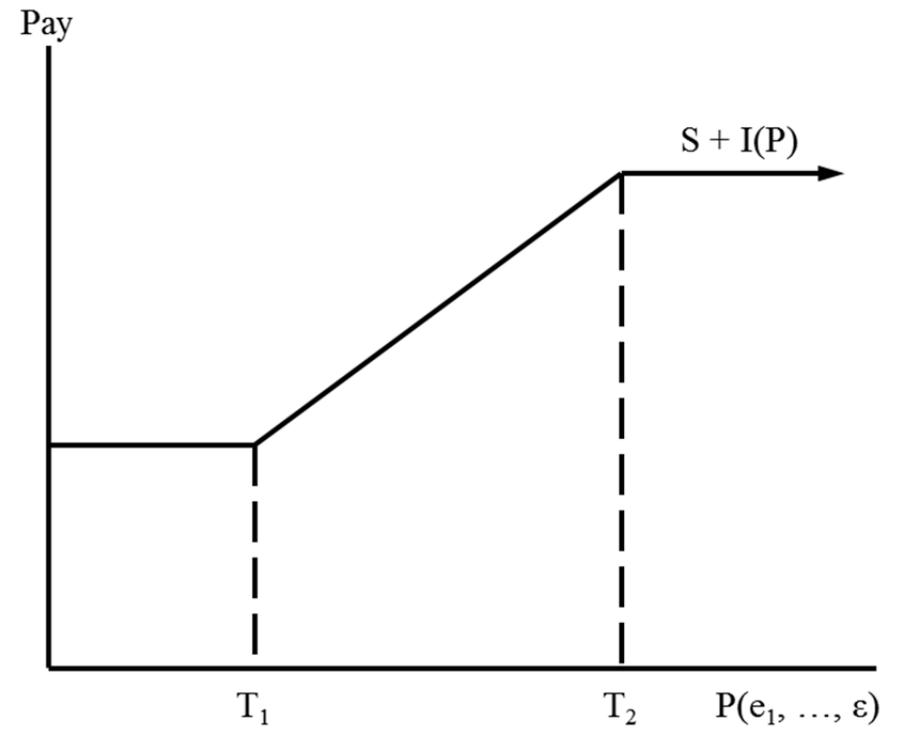
**Figure 22.1. Pay for Performance Plan**
A pay plan has 3 components: Performance Evaluation
P; Pay for Performance I(P), and Base Salary S.
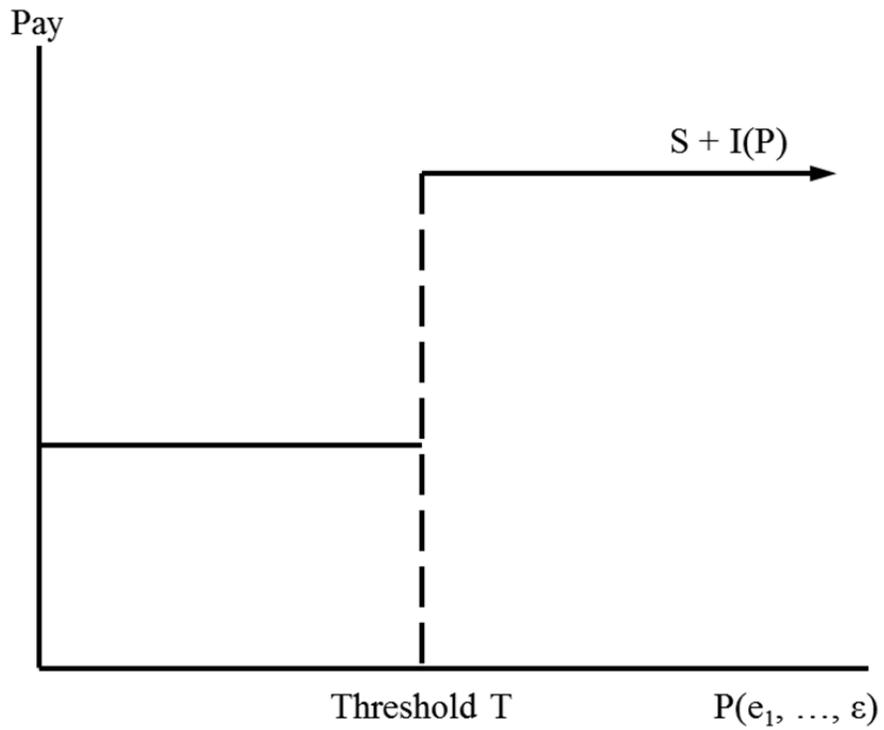
**Figure 22.2a. Lump Sum Bonus**
Lump sum bonus for performance above threshold T.
Promotion-based incentives have the same shape.



**Figure 22.2b. Bonus with Floor & Cap**
Base salary is paid for performance below $T_1$; a bonus
is paid for performance above $T_1$, up to $T_2$.