PERSONNEL PSYCHOLOGY: Performance Evaluation and Pay for Performance

Sara L. Rynes  
*Department of Management & Organizations, Tippie College of Business, University of Iowa, Iowa City, Iowa 52242; email: Sara-Rynes@uiowa.edu*

Barry Gerhart  
*Graduate School of Business, University of Wisconsin, Madison, Wisconsin 53706; email: bgerhart@bus.wisc.edu*

Laura Parks  
*Tippie College of Business, University of Iowa, Iowa City, Iowa 52242; email: Laura-Parks@uiowa.edu*

**Key Words**  
appraisal, pay, motivation, incentives, feedback

**Abstract**  
Although there is a voluminous psychological literature on performance evaluation (PE), surprisingly little of this research examines the consequences of linking pay to evaluated performance in work settings. Rather, PE research has been dominated by cognitive processing, measurement, and construct validity issues. At the same time, a large literature on pay-for-performance (PFP) linkages does exist, but most of it has been conducted in disciplines other than psychology. We think this pattern should change. To this end, we briefly trace the origins of the general separation of PE research from PFP research in psychology. From there, we review recent research on the relationship between PE and performance improvement, particularly with respect to multisource or 360-degree evaluation. We then turn to research on various PFP systems, such as merit pay and individual and group incentives. We conclude with suggestions as to how psychological research can make useful contributions to knowledge of PE, PFP, and performance improvement.

**CONTENTS**

- **INTRODUCTION** .......................................................... 572
- **ORIGINS OF THE SEPARATION OF PERFORMANCE**
  - Evaluation and Pay in Psychological Research .................. 573
  - Motivation Theories .................................................. 573
  - Empirical Evidence .................................................. 575
  - Kluger & DeNisi’s Feedback Intervention Theory ................. 578
  - Multisource (360-Degree) Performance Evaluation Research .......... 579

0066-4308/05/0203-0571$14.00  
571
INTRODUCTION

“The one issue that should be considered by all organization theories is the relationship between pay and performance” (Lawler 1971, p. 273).

The vast majority of organizations (at least in the United States) claim to use pay-for-performance (PFP) systems, and most U.S. workers say they want to be paid on the basis of performance (LeBlanc & Mulvey 1998, U.S. Bureau of National Affairs 1988). In addition, meta-analytic results show that increasing the connection between performance and pay can be very effective for improving performance. After conducting the first meta-analysis comparing alternative motivational interventions, Locke et al. (1980) concluded, “Money is the crucial incentive . . . no other incentive or motivational technique comes even close to money with respect to its instrumental value” (p. 379). Subsequent meta-analyses have tended to support this conclusion (e.g., Jenkins et al. 1998).

Given the importance of pay and performance to employers and employees as well as the potential for well-designed PFP systems to improve performance, one would think that research examining PFP would be plentiful in psychology. However, this has not been the case, particularly in recent years. Although there is a voluminous psychological literature on performance evaluation (PE), surprisingly little of this research examines the consequences of linking pay to performance in
work settings (Arvey & Murphy 1998, Smither et al. 2004). Conversely, although there is a large literature on the effects of PFP, most of it has been produced in disciplines other than psychology.

We would like to see this pattern change. We believe that psychology has much to contribute to some of the major practical issues concerning pay and performance in work settings. For example, academics in other disciplines often assume a great deal about the psychological mechanisms (perception, evaluation, goal choice) that explain employee reactions to pay plans. Greater attention to psychological theories and development of new research streams would put these assumptions to the test and also help identify reasons why plans do not always work as intended.

However, in order for psychology to contribute, it is important to understand the current state of findings from both of the two distinct literatures mentioned above: research on PE (particularly with respect to PE’s role in improving performance) and research on the performance effects of PFP. To this end, we briefly trace the origins of the separation of PE research from pay research in psychology.

ORIGINS OF THE SEPARATION OF PERFORMANCE EVALUATION AND PAY IN PSYCHOLOGICAL RESEARCH

Although organizations conduct performance evaluations for many reasons, the most basic one is to improve performance (Murphy & Cleveland 1995). Performance, in turn, is believed to be a joint function of both motivation and ability (Campbell & Pritchard 1976, Vroom 1964). Consistent with this notion, PE is believed to be capable of improving performance in two ways: through developmental feedback (directed primarily at improving ability to perform), and through administrative decisions that link evaluated performance to organizational rewards and punishments such as pay, promotion, or discharge (aimed primarily at enhancing motivation). In practice, however, psychological research on PE has focused far more heavily on the first route to performance improvement (i.e., feedback) than the latter (rewards). Indeed, the relative neglect of the administrative function of PE by research psychologists has become so notable in recent years that the most recent Annual Review of Psychology chapter on PE (Arvey & Murphy 1998) did not even mention research linking PE to pay or other rewards.

This relative neglect of the administrative functions of PE is unfortunate because performance is likely to improve most when employees both (a) receive information that will enable them to perform better and (b) have an incentive to act on that information. As such, we begin by examining the roots of the split between PE and PFP in psychological research, which appears to have begun with certain motivational theories that originated around the middle of the last century.

Motivation Theories

Typically, Annual Review chapters do not review much old research. However, it is important for a new generation of readers to understand the origins of the curious
split between PE and PFP in psychological research, a split that occurred despite the fact that the two practices are integrally related in most work organizations.

At least three historical theories of motivation have dampened psychologists’ interest in investigating the links between performance, PE, and pay: Maslow’s hierarchy of needs theory, Herzberg’s motivator-hygiene theory, and Deci & Ryan’s cognitive evaluation theory. A major implication of all three theories is that monetary rewards are not a major determinant of work motivation, except perhaps for employees at low income levels (Maslow 1943).

Maslow’s need hierarchy theory (1943) posited that human needs are arranged in a hierarchy of prepotency that is biological or instinctive in nature (Miner 1980). According to the theory, a need that is deprived acts as a primary motivator (i.e., “monopolizes consciousness”), while a need that is satisfied has less motivational impact.

At the bottom of Maslow’s hierarchy are physiological (food, sleep) and safety (e.g., housing) needs—precisely those needs posited to be most effectively satisfied by money. However, once these basic needs are satisfied, individuals are hypothesized to focus on “higher” needs, such as love, esteem (including achievement, independence, confidence), and self-actualization (“to become everything that one is capable of becoming”) (Maslow 1943, p. 382). Maslow hypothesized that these higher-order needs are more likely to be met through engagement in meaningful work than through monetary rewards. Although Maslow did not view need satisfaction as an all-or-nothing process (i.e., a person could simultaneously be motivated to varying degrees by needs at different levels), the expectation was that on average, all people progress through the hierarchy in roughly the same manner, and that as people move up the hierarchy, pay becomes less important.

The second theory, Herzberg’s motivation-hygiene theory, focused on identifying factors that contribute to either satisfaction, or dissatisfaction, at work (Herzberg et al. 1957). Herzberg viewed satisfaction and dissatisfaction not as opposite ends of the same continuum, but rather as two distinct constructs: “The factors involved in producing job satisfaction (and motivation) are separate and distinct from the factors that lead to job dissatisfaction” (p. 9). Like Maslow, Herzberg saw hygienic needs as being driven by people’s “animal nature” (p. 9). In contrast, motivational factors or factors associated with the work itself were posited to “relate to that unique human characteristic, the ability to achieve and, through achievement, to experience psychological growth” (p. 9). Most importantly for present purposes, Herzberg posited that money was more likely to be a “hygienic” factor—i.e., one capable of causing or reducing dissatisfaction—than a satisfying or motivating one. Thus, Herzberg believed that money played a role in creating or reducing dissatisfaction, but not in contributing to satisfaction or motivation.

The third theory to challenge the role of money in motivation is Deci & Ryan’s (1985) cognitive evaluation theory (CET). Although the theory is somewhat more complicated than described here, in general, CET argues that placing strong
emphasis on monetary rewards is likely to decrease people’s intrinsic interest (i.e., interest in the work itself), thus dampening a potentially powerful alternative source of motivation. Deci & Ryan (1985) describe intrinsic motivation as “based in the innate, organismic needs for competence and self-determination” (p. 33), and argue that it occurs in its purest form when “a person does an activity in the absence of a reward contingency or control” (p. 35).

Deci and Ryan argue that when effort is exerted in exchange for pay, pay takes on a controlling aspect that threatens the individual’s need for self-determination. According to Ryan et al. (1983), if one must perform a task “in some particular way, at some particular time, or in some particular place...to receive the reward, the reward tends to be experienced as controlling” (p. 738). This perception of being “controlled” is assumed to be demotivational and to work against the potential incentive effects of extrinsic rewards (i.e., rewards not associated with the work itself, such as pay, candy, trinkets, or praise).

For the most part, Deci and his colleagues have tended to argue that the net effect of monetary rewards on intrinsic interest is generally negative. However, Ryan et al. (1983) propose that the overall effects of pay on intrinsic interest depend on the information embedded in pay outcomes. Specifically, they argue that intrinsic interest is likely to be impaired to the extent that information about rewards is seen to be controlling. However, to the extent that pay provides meaningful information regarding self-competence in contexts where people have discretion in choosing how to perform tasks, then monetary rewards can actually increase intrinsic interest. Thus, in statements of CET theory that are more complex, the net effect of rewards on intrinsic interest depends on the relative impact of monetary rewards on perceived control versus perceived self-competence.

Empirical Evidence

As we show below, extensive prior reviews of PFP research suggest that pay is far more motivational than assumed by Maslow, Herzberg, or Deci & Ryan (Gerhart & Rynes 2003). Thus, it is important to evaluate the cumulative evidence with respect to each of these theories.

With respect to Maslow’s hierarchy, reviews of the evidence suggest that pay is important for fulfilling higher-order needs, as well as lower-order ones. Lawler’s (1971) review of then-existing evidence concluded, “Money can buy food, security, social relations, and esteem, and to some extent, it can satisfy self-actualization needs” (p. 26), and subsequent research has continued to support this conclusion (e.g., Frank 1985).

Tests of Herzberg’s research also challenge the idea that pay cannot satisfy higher-order needs. Indeed, Herzberg’s 1987 summary of his own research (12 studies, 1685 subjects) showed that pay was mentioned as a motivator nearly as often as it was a dissatisfier. Pay was particularly likely to be mentioned as a motivator when it was seen as a form of recognition. According to Lawler, “the tendency for pay to be mentioned as a contributor to satisfaction as often
as it is mentioned as being unfairly low or dissatisfying has appeared in most of the studies that have attempted to replicate or test Herzberg’s theory” (1971, pp. 32–33). Moreover, Herzberg’s conclusions are even less well supported when methodologies other than Herzberg’s own (i.e., storytelling critical incidents) are used to assess the proposed dichotomy (Campbell & Pritchard 1976).

Evidence regarding CET’s presumption of a negative relationship between rewards and intrinsic interest is still a matter of debate, with support for the theory being mixed at best. For example, a meta-analysis by Eisenberger & Cameron (1996) examined 83 studies that compared a rewarded group (verbal or tangible rewards) with a no-reward control group on two traditional measures of intrinsic interest: amount of free time devoted to a task when the tangible or verbal reward is subsequently withdrawn (44 effect sizes) or self-reported attitude toward the task (e.g., interest, enjoyment, or satisfaction; 39 effect sizes). Across studies, the mean correlation between reward condition and free time was (only) \( -0.04 \), while the correlation between reward condition and task-related attitudes was actually positive (\( \rho = 0.14 \)) rather than negative (as would be predicted by CET). Therefore, the detrimental effects of rewards were generally not supported in their meta-analysis, even using the free time measure.

Subsequently, Deci et al. (1999) conducted a meta-analysis on the same issues addressed by Eisenberger & Cameron (1996). Although Deci et al. found more overall support for the detrimental effects of extrinsic rewards than did Eisenberger & Cameron, even they found “no effect for performance-contingent rewards” on the attitude measure of intrinsic interest (p. 644). In addition, Deci et al. found that extrinsic rewards were “more detrimental for children than for college students” (p. 656). This is a very important point because in the workplace, it is the motivation of adults (rather than of elementary school-aged children) that is at issue.

This last point raises a more general concern about the generalizability of CET-inspired research, which is that the vast majority of this research has been conducted in the laboratory under conditions that differ substantially from real work settings. For example, the monetary rewards in CET lab experiments generally fall in the sub-$10 range. In contrast, compensation for full-time work in real organizations ranges from the tens to hundreds of thousands of dollars (or more). Similarly, lab studies on motivation typically observe performance over trials of a few hours or less. In contrast, most people spend approximately two thousand hours at work each year. Finally, most subjects in CET research have been school-aged children rather than working adults. Given these substantial differences in background conditions, the effects of monetary rewards might play out considerably differently in the workplace than they do in the educational laboratory.

In a field examination of CET theory, Fang & Gerhart (2000) investigated the idea that Deci & Ryan (1985) may have focused too much on the potentially controlling aspects of rewards and not enough on the informational aspect. Consistent with this hypothesis, they found that employees covered by PFP plans reported higher intrinsic interest than employees not covered by such plans. Although causality cannot be demonstrated, this result is consistent with the fact that
people frequently report pay as a motivator when it is perceived as recognition for good performance (Lawler 1971). It is also consistent with the positive correlation between intrinsic and extrinsic motivation reported by Amabile and her colleagues (1996), as well as with meta-analytic evidence that is even more recent (Cameron et al. 2001).

In summary, although the ideas developed by Maslow and Herzberg have had considerable appeal to many people, the prevailing view in the academic literature is that the specific predictions of these theories are not supported by empirical evidence (Kanfer 1990). Although Deci and Ryan’s ideas have received stronger (albeit mixed) support, they remain largely untested in ongoing work settings, which differ in important ways from the laboratory settings in which they have been investigated.

Nevertheless, it would be a mistake to underestimate the influence these theories have had on both research and practice. For example, in research, we probably owe the dearth of psychological research on PFP primarily to these theories. This outcome remains a serious concern, as the influence of CET theory seems to be spreading toward human resources and industrial/organizational psychology research, rather than receding (e.g., Sheldon et al. 2003). In addition, these theories’ influence on practice also remains substantial. For example, in the widely read *Harvard Business Review*, Stanford professor Jeff Pfeffer (1998a, p. 112) called the idea that people work for money a “myth.” In addition, he asserted, “a substantial body of research has demonstrated, both in experimental and field settings, that large external rewards can actually undermine intrinsic motivation” (1998b, p. 216). Kohn (1993) has made similar claims.

Meyer, Kay, and French’s “Split Roles” in Performance Evaluation

In addition to the three motivation theories reviewed above, a final development that contributed to the separation of PE and pay research in psychology was Meyer et al.’s (1965) notion of “split roles” for PE. As mentioned earlier, PE has two distinct functions: to develop employees through such mechanisms as feedback and goal setting, and to evaluate employees for purposes of making administrative decisions (e.g., pay increases or promotions; Murphy & Cleveland 1995).

In a seminal article that had widespread impact on both researchers and practitioners, Meyer and colleagues (1965) argued that these two functions of performance appraisal should be kept completely separate from one another—that is, evaluations of performance and discussions of pay should be separated in time from discussions of how to improve (or “develop”) performance. This recommendation was based on the supposedly distracting and demoralizing aspects of criticism, as well as the authoritarian nature of evaluation: “Interviews designed primarily to improve a person’s performance should not at the same time weigh his or her salary or promotion in the balance... It seems foolish to have a manager serving the self-conflicting role as counselor (helping someone improve performance)
when, at the same time, he or she is presiding as a judge over the same employee’s salary action” (Meyer et al. 1989, p. 26).

We are aware of three field studies that attempted to test this assertion, and none obtained supportive results. Dorfman et al. (1986) found that discussion of pay and advancement during the PE session led to higher employee satisfaction with the PE process. Prince & Lawler (1986) found that discussion of salary during PE had either no impact or a slightly positive impact on PE processes (e.g., employee participation) and outcomes (e.g., satisfaction with the PE). Boswell & Boudreau (2002) found no difference in employee satisfaction with PE, satisfaction with supervisor, or awareness of developmental opportunities between development-only PEs and those where development was combined with evaluation for administrative and reward purposes. Moreover, they found that employees reported higher intentions to use developmental feedback when developmental feedback was combined with evaluation.

Another of Meyer et al.’s assertions was that criticism has a negative effect on goal achievement, while praise has little effect in either direction. Two bodies of empirical research bear on these assertions: a meta-analysis of (mostly laboratory) studies of feedback interventions (Kluger & DeNisi 1996), and recent field research pertaining to multisource feedback. We begin with the meta-analysis.

Kluger & DeNisi’s Feedback Intervention Theory

Kluger & DeNisi (1996) conducted an extensive historical review and meta-analysis (607 effect sizes, reflecting 23,663 observations) that reached some interesting conclusions about the effects of feedback interventions (FIs) on performance. Their findings suggested that, on average, FIs (i.e., providing information about performance) improved subsequent performance ($d = 0.41$). One of the effects they proposed is that following an FI, “effort is increased if the feedback sign is negative, and decreased or maintained if the sign is positive” (p. 263). In other words, people work harder when they find out that they are not performing up to expectation and relax when feedback suggests that performance is adequate, consistent with cybernetic or self-regulating theories of motivation (Kanfer 1990).

On the other hand, Kluger & DeNisi also found that despite the generally positive effect of feedback, in more than one third of the cases, performance decreased following feedback. In light of the nonuniform effects of feedback on performance, Kluger & DeNisi proposed and tested a feedback intervention theory (FIT) suggesting that different types of performance feedback differentially affect people’s locus of attention among three hierarchically organized levels of control: task learning, task motivation, and meta-task processes (including focus on the self). They hypothesized, and found, that the effectiveness of feedback decreases as it causes attention to move up the hierarchy—i.e., closer to the self and away from the task. For example, feedback specifically designed to be demoralizing was more likely than other kinds of feedback to have detrimental effects (p. 273). Thus, Meyer et al.’s contention that negative feedback can impair performance
appears to be true, at least when the feedback draws attention away from the task and toward the person him- or herself.

On the other hand, negative feedback does not always, or perhaps even “usually,” have that effect. For example, FIs that focused on the task or on task learning did not have such effects. Moreover, the overall test of differences for feedback direction (i.e., positive versus negative) was not significant. Thus, the common advice given to practicing managers to “criticize the work, but not the person” appears to be a sound one. However, Meyer et al.’s assertion that negative feedback usually causes a decline in performance does not appear to be supported. This is also the case in field studies of PE, as described below.

Multisource (360-Degree) Performance Evaluation Research

The vast majority of studies examined by Kluger & DeNisi were conducted in the laboratory and used student subjects. More recently, there have been a number of studies in which multisource or 360-degree feedback (from peers, subordinates, and customers as well as supervisors) has been given in ongoing work environments. Multisource feedback is seen as potentially more useful than supervisor-only evaluations in modern-day work environments, which increasingly incorporate team-based production, organizational structures that are more flat, and work organized around horizontal rather than vertical flows. In addition, because 360-degree feedback is gathered from multiple individuals, feedback reliability and validity may be substantially improved over the typical supervisor-only evaluation (Mount et al. 1998).

Results across these studies show that individuals generally improve their performance following 360-degree feedback, at least in terms of subsequent ratings by the same observers (e.g., Hazucha et al. 1993, Reilly et al. 1996, Smither et al. 1995; for an exception see Atwater et al. 2000). However, average effect sizes generally appear to be modest. For example, a recent unpublished meta-analysis by Smither and colleagues (2004) estimated that unweighted $d$s for feedback from direct reports, peers, and supervisors were 0.24, 0.12, and 0.14, respectively, with weighted $d$s being even smaller.

However, most studies also show wide variations in improvement across individuals. As such, it is important to examine personal or organizational factors that may moderate responses to multisource feedback. In light of Meyer et al.’s hypotheses, of particular interest here is the variable most likely to convey either positive or negative information to the recipient—i.e., the degree of congruence between self- and observer ratings. In general, studies have found that individuals who overrate their own performance (as compared to observer ratings) tend to improve most following feedback (Atwater 1995, Johnson & Ferstl 1999, Smither et al. 1995, Smither et al. 2004), even after taking into account initial performance and the possibility of regression to the mean. These findings are consistent with Kluger & DeNisi’s (1996) speculation that, in general, effort increases after receipt of negative feedback.
Thus, there is evidence that the degree of incongruence between self- and observer ratings impacts subsequent changes in performance. The more positive improvement for overraters (i.e., those who receive negative feedback) is once again consistent with cybernetic theories of motivation (Kanfer 1990) and Kluger & DeNisi’s (1996) FIT theory, which posits that only “feedback-standard gaps receive attention and foster change.”

In summary, laboratory research on feedback interventions and field research on split-role and multisource evaluations suggests that (a) there is no general decrement to satisfaction, motivation, or performance when evaluative aspects of PE are combined with developmental ones; (b) on average, FIs produce positive performance improvements, although these appear to be smaller in field than in laboratory settings; (c) there are substantial situational moderators of the feedback-subsequent performance relationship, with stronger performance gains tending to follow the receipt of negative feedback; but (d) this last finding may not hold in cases where the feedback causes the individual to focus on the self or other meta-processes, rather than on the task or task learning.

Unfortunately, 360-degree field research has not yet looked directly at this last question (i.e., focus of attention). However, field research has produced evidence that managers’ ratings improve more when they seek input from others (e.g., Hazucha et al. 1993), meet with subordinates to discuss feedback results (Walker & Smither 1999), or set improvement goals following feedback (Atwater et al. 2000). We also hypothesize that the consequences of substandard performance will act as a moderator. For example, in organizations where substandard performers are terminated or encouraged to leave, negative feedback would seem likely to have a stronger effect. We would expect this effect to show up either through improved performance (an incentive effect) or through turnover (e.g., quits or firing), either of which might be functional from the organization’s point of view in improving average employee performance.

**SUMMARY OF MOTIVATION THEORY AND PERFORMANCE EVALUATION RESEARCH**

As we have shown, a good deal of work in psychology has focused on testing hypotheses that negative evaluative feedback and extrinsic rewards (including money) are largely detrimental to motivation and performance. For the most part, empirical evidence on both points fails to support these hypotheses, particularly in work settings.

Before moving to research on PFP, it should be noted that some psychologists have taken strong positions supporting the importance of pay as a motivator. For example, Lawler (1971) reviewed empirical evidence on pay, finding that it was rated more highly than Herzberg et al. (1957) had reported years earlier. In addition, Lawler developed a process model of motivation suggesting that pay should in fact be quite an important motivator because of its instrumentality for obtaining so
many other outcomes. Similarly, Locke has produced both empirical (Locke et al. 1980) and conceptual work (e.g., Bartol & Locke 2000) suggesting that pay has considerable effectiveness as a motivator.

Still, the “taint” around evaluative appraisals and using money as a reward lingers, and continues to discourage the study of relationships between PE, PFP, and performance within psychology. As such, we turn now to other literatures—particularly management, economics, and finance—that have not shied away from examining relationships between performance-based rewards and subsequent performance. While the PE literature tends to focus on the developmental and ability side of the performance equation, the PFP literature tends to focus on the evaluative and motivational side.

PAY FOR PERFORMANCE: EVIDENCE AND PROCESSES

Despite concerns and arguments against the use of PFP, most organizations use it in some form. Moreover, this use appears to be growing, rather than declining (R. Heneman et al. 2000). In addition, our reading of the evidence on PFP is generally positive. To be sure, there are some very important caveats: pay is not the only important motivator in organizations, and PFP programs can yield serious, unintended negative results. Nevertheless, it can also deliver powerful improvements in performance. This combination of upside potential and downside risk is part of what makes PFP so interesting.

The clearest evidence of the impact of PFP comes from meta-analytic reviews of research conducted in actual work settings. In the first such study, Locke et al. (1980) found that the introduction of individual pay incentives increased productivity by an average of 30%. In contrast, Locke and colleagues found that job enrichment produced increases of 9%–17%, whereas enhanced employee participation yielded less than 1%. This meta-analysis is particularly compelling because it included only studies that used either control groups or before-and-after designs in real work settings, and measured performance via “hard” criteria (e.g., physical output) rather than supervisory ratings. Subsequent meta-analyses have yielded similar results regarding the substantial impact of individual pay incentive systems (e.g., Guzzo et al. 1985, Jenkins et al. 1998, Judiesch 1994).

What processes are responsible for these effects? In addition to the theories discussed earlier (e.g., Deci & Ryan 1985, Herzberg et al. 1957, Maslow 1943), a variety of psychological theories have been used to explain money’s role in motivation. These include equity and justice theories, drive theory, goal theory, self-efficacy theory, prospect theory, and expectancy theory (space limitations preclude reviews of these theories; see Bartol & Locke 2000 and Gerhart & Rynes 2003 for more information). In addition, economists and management scholars have turned to agency theory (Jensen & Meckling 1976) and tournament theories (Lazear & Rosen 1981).

Although the preceding theories propose a variety of ideas about pay and motivation, essentially PFP operates on motivation and performance through two
general processes (Gerhart & Rynes 2003). First, there is the potential for an “incentive effect”: the impact of PFP on current employees’ performance, holding the attributes of the workforce constant. Incentive effects have been the primary focus of most PFP theory and research.

Second, there is the potential for a “sorting effect,” which we define as the impact of PFP on the attributes of the workforce through differential attraction and retention processes (Gerhart & Milkovich 1992, Lazear 1986). For example, different types of PFP systems may cause different types of people to apply to and stay with an organization (i.e., to self-select), and these people may have different levels of ability or other attributes (e.g., need for achievement) that enhance effectiveness more in some organizations than in others. Organizations, too, may differentially select and retain employees depending on the nature of their culture or PFP strategies.

To the extent that studies reviewed in the preceding meta-analyses tracked the same individuals before and after the PFP intervention, the observed effects are pure incentive effects. However, to the degree that individuals making up the workforce changed in response to changes in pay systems, then some of the improvements in performance might be due to sorting effects. Lazear (1986), for example, found a 44% increase in productivity when a glass installation company switched from salaries to individual incentives. Of this increase, roughly 50% was due to existing workers increasing their productivity, while the other 50% was attributable to less productive workers quitting and being replaced by more productive workers over time.

In this vein, evidence suggests that PFP is more attractive to those higher in academic achievement (Trank et al. 2002), need for achievement (Bretz et al. 1989), and self-efficacy (Cable & Judge 1994). Research also shows that high performers are most likely to seek other employment if performance is not sufficiently recognized with financial rewards (Trevor et al. 1997). Conversely, low performers are more likely to stay with an employer when PFP relationships are weaker (e.g., Harrison et al. 1996).

CHOOSING A PERFORMANCE MEASURE

Although PFP can significantly improve performance, most of the above evidence has been obtained in contexts where individual contributions are separable and where performance can be measured objectively. However, in most jobs, individual contributions are difficult to identify and/or performance is more difficult to measure objectively. Consequently, subjective performance measures, particularly supervisory ratings, are more often used. In addition, organizations increasingly supplement individual-level measures with group, unit, or organization-level performance measures, especially to the degree that work is interdependent. Moreover, organizations are increasingly building risk into both individual (e.g., bonuses) and unit-based pay programs (e.g., gain sharing), despite the fact that employees are typically risk averse with respect to pay.
In deciding on performance measures that will be used to determine pay, there are at least three key choices: (a) How much emphasis can or should be placed on results-oriented performance measures (e.g., number of units produced) relative to behavior-based ones (e.g., supervisory evaluations of effort or quality)? (b) How strong (or intensive) should incentives be and how will risk aversion influence their effectiveness? (c) How much emphasis should be placed on individual contributions relative to collective contributions? Although we discuss each choice individually, in practice, many organizations use multiple performance measures to balance multiple (and sometimes conflicting) objectives.

Behavior-Based (Subjective) versus Results-Based (Objective) Measures

Turning to the first question, subjective behavior-oriented measures (such as traditional PE ratings) offer a number of potential advantages relative to results-based measures (Gerhart 2000). First, they can be used for any type of job. Second, they permit the rater to factor in variables that are not under the employee’s control but nevertheless influence performance. Third, they permit a focus on whether results are achieved using acceptable means and behaviors. Fourth, they generally carry less risk of measurement deficiency, or the possibility that employees will focus only on explicitly measured tasks or results at the expense of broader prosocial behaviors, organizational citizenship behaviors, or contextual performance (see, e.g., Arvey & Murphy 1998, Wright et al. 1993).

Despite these potential advantages, the subjectivity of behavior-oriented measures limits their ability to differentiate employees (Murphy & Cleveland 1995). In addition, meta-analytic evidence finds a mean interrater reliability of only 0.52 for performance ratings (Viswesvaran et al. 1996), making it difficult for organizations to justify differentiating employees based on such error-laden performance measures (360-degree appraisals may be helpful in this regard). The subjectivity in these PE measures has led the PE literature to focus on identifying cognitive biases in evaluation (in hopes of reducing them) and examining how various features of PE measures (and PE-related processes) influence employees’ perceptions of fairness (in hopes of improving PE’s legitimacy and effectiveness; e.g., Folger & Konovsky 1989).

Even if subjectivity in PE could be sufficiently controlled and performance reliably and credibly differentiated, managers may be reluctant to do so because of concerns about adverse consequences for workgroup cohesion, prosocial behaviors, and management-employee relations (H. Heneman & Judge 2000, Longenecker et al. 1987). Indeed, a long line of research suggests that managers tend to become considerably more lenient in PE when they know that ratings will be used for administrative rather than purely developmental purposes (e.g., Jawahar & Williams 1997). Perhaps for these reasons, research suggests that pay is rarely seen by employees as strongly differentiated across employees when it is tied to subjective measures of performance (HayGroup 1994, 2003).
Results-oriented measures of performance, such as productivity, sales volume, shareholder return, and profitability, would seem to provide an antidote to the subjectivity and unreliability of performance ratings. Lawler (1971), for example, concluded, “[I]n general, objective measures enjoy higher credibility; that is, employees will often grant the validity of an objective measure . . . when they will not accept a superior’s rating” (p. 166).

Nevertheless, such measures are not available for most jobs, at least at the individual level. Moreover, agency theory emphasizes that results-based plans (such as gain- or profit-sharing) increase risk bearing among employees (Gibbons 1998). Because most employees derive the bulk of their income from employment, they cannot diversify their employment-related earnings risk, making them more risk-averse than, say, are investors.

The consequences of risk bearing may go unnoticed when performance (e.g., profitability) is good and a PFP plan is paying out, but can quickly become a major employee relations issue when results decline and payoffs go down. If poor organizational results are perceived as being strongly influenced by factors that are beyond employees’ control (e.g., poor decisions by top executives), feelings of injustice can be strong. Indeed, many such plans are abandoned due to employee pressure (e.g., Petty et al. 1992).

Finally, even though results-based measures are objective and possibly more reliable, they may also be more deficient as indicators of the full domain of expected performance. Lawler warned nearly 35 years ago, “[I]t is quite difficult to establish criteria that are both measurable quantitatively and inclusive of all the important job behaviors,” and “if an employee is not evaluated in terms of an activity, he will not be motivated to perform it” (Lawler 1971, p. 171). Unfortunately, efforts to make measures more inclusive (e.g., unit profitability rather than individual piece rates) generally reduce employees’ expectation that they can influence the performance measure, thus reducing effort.

Incentive Intensity

Theories and empirical research generally concur that stronger PFP links increase motivation and performance. However, there are at least two offsetting disadvantages to strong PFP relationships: They may (a) exacerbate problems having to do with risk aversion and (b) be more prone to deficiency in performance measures (e.g., paying for quantity of production without adequate attention to quality).

Turning to the first issue, agency theory (Jensen & Meckling 1976) hypothesizes that employees are risk-averse and will prefer “insurance” to avoid downside earnings risk. In addition, employees or executives may be tempted to manipulate results-based systems by artificially inflating results measures (e.g., revenues or profits), resulting in short-term incentive payouts but long-term harm to organizations. Gaming may also arise if management believes the standards for payout are too easy, or if employees believe the standards are too difficult (Whyte 1955). Conflict over standards is a leading cause of plan failure. Thus, using strong incentives might be described as a high-risk, high-return strategy (Gerhart et al. 1996).
Milgrom & Roberts (1992) suggest that the optimal intensity of incentives depends on four factors: the extent to which better results can be created by additional effort, the precision with which the desired activities can be assessed, employees’ risk tolerance, and their responsiveness to incentives. The trick, of course, is getting the right balance.

**Individual versus Group (or Collective) Performance**

Criticisms have been leveled at organizations for focusing too much on individual performance and rewards. For example, Pfeffer (1998b) critiqued individual merit and incentive plans as being ineffective, inciting grievances, and reducing product quality. Similarly, Deming (1986) argued that management’s “excessive” focus on individual performance often obscures apparent differences in individual performance that “arise almost entirely from the system that (people) work in, rather than the people themselves” (p. 110). Deming and Pfeffer also argue that focusing on individual performance discourages teamwork: “Everyone propels himself forward, or tries to, for his own good. . . . The organization is the loser” (Deming 1986, p. 110).

While the potential pitfalls of individually based PFP are important, the literature is quite clear that group-based plans also have their own drawbacks. One is that most employees (at least in the United States) prefer that their pay be based on individual rather than group performance (Cable & Judge 1994, LeBlanc & Mulvey 1998). Another is that this preference is strongest among the most productive and achievement-oriented employees (see, e.g., Bretz et al. 1989, Lazear 1986, Trank et al. 2002, Trevor et al. 1997). This means that group-based pay may also have unfavorable sorting effects, causing the highest performers to choose alternative opportunities where individual results will be rewarded more heavily. Yet another drawback has to do with weakened incentive effects: “Unless the number of individuals in a group is quite small, or unless there is coercion or some other special device to make individuals act in their common interest, *rational self-interested individuals will not act to achieve their common or group interests*” (Olson 1965, pp. 1–2, emphasis in original).

This last phenomenon has been widely studied (Kidwell & Bennett 1993), and goes by many names (e.g., the common-resource problem, public-goods problem, free-rider problem, or social-loafing problem). The general idea is that when people share the obligation to provide a resource (e.g., effort), it will be undersupplied because the residual returns (e.g., profit-sharing payouts) to the effort often are shared relatively equally, rather than distributed in proportion to contributions. Empirical reviews suggest that the free-rider problem is sufficiently important (e.g., Albanese & Van Fleet 1985, Cooper et al. 1992, Shepperd 1993) that researchers have devoted considerable attention to how to mitigate free-rider effects.

One potential solution is to give differentiated rewards to group members based on their individual contributions (this solution has been used successfully for many years by Lincoln Electric). As mentioned above, differentiating rewards based on performance can yield benefits via both incentive and sorting effects (Bishop 1984).
In the same vein, differentiating pay on the basis of individual performance, even within group systems, may reduce the tendency of high-performing employees to leave organizations that switch to group-based pay systems (e.g., Weiss 1987).

On the other hand, to the extent that pay dispersion is not accepted as being based on true performance differences or other acceptable equity considerations, differentials may have negative effects on productivity. In a pair of studies that examined the effects of within-group pay dispersion on group productivity, both Pfeffer & Langton (1993) and Bloom (1999) concluded that higher pay dispersion harms group productivity. In Pfeffer & Langton’s study of faculty departmental research productivity, the authors obtained a negative regression coefficient for the independent variable representing salary dispersion. Similarly, in a study of the relationship between pay dispersion and team performance in major league baseball, Bloom (1999) reported that teams with lower pay dispersion had better performance (operationalized in a number of different ways).

However, we believe that both studies probably drew overly pessimistic conclusions about pay dispersion because of certain features of their analytic models. For example, in Pfeffer & Langton’s study, the departmental productivity equation controlled for the correlation between pay and productivity in each department. As such, the negative coefficient for pay dispersion essentially reflected the effect of pay dispersion that was unexplained by differences in performance (see Gerhart & Rynes 2003, pp. 180–182, for a more thorough explanation). Similarly, in specifying his model of baseball team performance, Bloom (1999) controlled for both team talent and team pay in estimating the coefficient for pay dispersion. However, team talent and team pay arguably should not be treated as control variables in this case because to do so parcels out the positive effects of pay dispersion, i.e., the attraction and retention of star players who are paid a great deal, thus resulting in better team performance, higher team pay, and greater dispersion. By controlling for both team pay and team ability, these advantages of performance-based pay dispersion are omitted from Bloom’s observed dispersion coefficient. In short, we suspect that both studies underestimated the positive effects of pay dispersion, at least in the contexts they studied.

In summary, both individual- and group-based pay plans have potential limitations. Individual-based plans may generate too little cooperation when work is highly interdependent and may be seen as unfair when system factors rather than individual effort and ability determine performance. In contrast, group-based plans can weaken incentive effects via free-rider problems, which generally increase with group size. Group-based plans can also result in detrimental sorting effects if high achievers go elsewhere to have their individual contributions recognized and rewarded. In our discussion below, plans that use aggregate performance measures (gain sharing, profit sharing, stock plans) would seem most likely to suffer from weakened incentive effects and unfavorable sorting with respect to individual performance. On the other hand, aggregate plans may have positive incentive effects if they promote cooperation, attract people with cooperative values, and avoid overly narrow individual goals.
**EMPIRICAL EVIDENCE ON VARIOUS PAY-FOR-PERFORMANCE PROGRAMS**

Table 1 classifies various PFP programs using two of the dimensions identified above: (a) results-based versus behavior-based performance and (b) group versus individual performance. Although not explicitly included in Table 1, the third factor (incentive intensity) also is captured to an important degree in that it tends to be stronger for results-based plans. Although we describe each program separately, it should be kept in mind that people are often paid using a combination of programs.

### Individual Incentives

The meta-analytic evidence reviewed earlier demonstrates a strong, positive average effect of incentives on employee productivity (e.g., Jenkins et al. 1998, Judiesch 1994, Locke et al. 1980).

### Merit Pay

Although merit pay continues to be the most widely used PFP program (especially among salaried employees), there is surprisingly little evidence about the performance implications of adopting, or not adopting, merit pay programs. R. Heneman (1992, pp. 247–252) listed 10 studies that purported to address the consequences of merit-based PFP. Five of the reported relationships were positive (using statistical significance as the criterion), whereas the rest of the relationships were nonsignificant. No statistically significant negative relationships between merit pay and performance were reported. These studies, however, are limited by the scarcity of control groups, longitudinal assessment, objective measures of performance, and unit measures of performance.

Perhaps the most well-known and best effort to incorporate these design features is a longitudinal four-year study of a governmental agency by Pearce et al. (1985), which is widely cited as suggesting that merit pay programs are not effective. Indeed, the authors themselves conclude, “the positive effects of the implementation of merit pay...were not supported by the data” (p. 271). Although Pearce...
et al. represents a laudable attempt to tackle a difficult issue, we believe that it is inappropriate to draw such a negative conclusion about merit pay from their study.

First, unit performance climbed during the entire 48-month period of their study. As such, the failure to obtain statistically significant (before-and-after) effects may be due in part to the small sample size (n = 20). In addition, the study ended before two-thirds of the units had actually distributed any merit increases based on performance (the program had been implemented and initial training begun, but no raises administered in the majority of their cases). In their words, “Eight of our 12 tests assess the effect of training and the start of the program on organizational performance, an emphasis somewhat different from testing changes in organizational performance after merit pay rewards were distributed” (p. 271). Another concern is that incentive intensity was weak, with high-performing units receiving merit increases that were not much different from those received by lower-performing units (Gerhart & Milkovich 1992). Other difficulties with the study, including the likelihood of political interference with program implementation in some units, are cited in the original study and reviewed in Gerhart & Rynes 2003 (pp. 187–188).

Although further research is greatly needed, the evidence that does exist on merit pay is more positive than negative (R. Heneman 1992). Given the typically weak link between performance and pay as well as problems with getting supervisors to provide credible measures of performance for administrative purposes (see, e.g., Jawahar & Williams 1997), perhaps any positive results at all should be regarded as impressive.

**Profit Sharing**

Profit-sharing plans pay out based on meeting a profitability target (e.g., return on assets or net income). As such, they are “risky” compensation schemes. Incentive effects of profit sharing, although hypothesized to be positive, may nevertheless be limited by the large number of employees involved (thus lowering expectancies of being able to influence results through individual effort), the riskiness of returns, and the fact that many profit-sharing plans defer the income generated until retirement.

Weitzman & Kruse (1990) reported generally positive employee attitudes toward profit sharing, although this was “tempered … by the risk of fluctuating income” (p. 123). They estimated that the mean effect of profit sharing on productivity was 7.4% (median = 4.4%). Kruse (1993) reported similar, but somewhat smaller effects, as did a meta-analysis by Doucouliagos (1995).

**Stock Plans**

Employee stock ownership and options have increased dramatically since 1990. (Note: A stock option is the right to purchase a fixed number of shares of stock during a fixed time period at a fixed price, regardless of the actual stock price.) According to the National Council on Employee Ownership website, in 2001
employees owned or had options to own stock worth “about $800 billion, or about 8% of all the stock in the U.S.” This figure compares with only 1%–2% a decade earlier.

Early research provided positive evidence on the effectiveness of stock plans, at least among the top five executives (for which public data are readily available; see Masson 1971 and Brickley et al. 1985). Gerhart & Milkovich (1990) moved beyond studying only the top five managers and examined the relationship between the percentage of top- and mid-level managers eligible for stock options in a firm and the firm’s return on assets. Their results suggested that companies having 20% of managers eligible for stock options had predicted returns on assets of 5.5%, as compared with predicted returns of 6.8% (or roughly 25% higher) for companies having 80% of managers eligible.

Despite these generally positive results, the potential drawbacks of stock-based pay have received increasing attention in recent years. Investors have questioned the number of options issued to employees (including executives), as well as the historically favorable accounting treatment of options. Indeed, the Financial Accounting Standards Board (2004) recently proposed a landmark change in this area (requiring companies to expense options, which typically reduces net income) that is causing companies like Microsoft to eliminate or drastically curtail their use of stock options (Krim & Spinner 2003). In addition, companies that historically relied heavily on stock options to attract and retain talent have had trouble competing when stock prices declined. Finally, a number of prominent examples of executives’ manipulation of stock prices for personal gain have further thrown the desirability of stock options as performance motivators into question.

Gain Sharing

Gain sharing links pay to results-based performance at a collective (usually facility) level. Although incentive effects should be lower relative to individual-level plans (due to lowered expectancy perceptions; Schwab 1973), they should be stronger than for corporate-level programs such as stock- and profit-based plans. Milkovich & Wigdor (1991, p. 86) suggested that group incentive plans such as gain sharing “may provide a way to accommodate the complexity and interdependence of jobs, the need for work group cooperation, and the existence of work group performance norms, but still offer the motivational potential of clear goals, clear PFP links, and relatively large pay increases.”

Indeed, the empirical evidence on gain sharing appears to be quite favorable (e.g., Welbourne & Gomez-Mejia 1995). In one interesting study, Petty et al. (1992) compared one division of an electric utility company that implemented a gain-sharing plan with another division that did not. The gain-sharing division performed better on 11 of 12 objective performance measures, providing an estimated savings between $875,000 and $2 million. In addition, employee perceptions of teamwork and other factors designed to be influenced by the plan were also positively affected.
In another well-designed multiyear longitudinal study, Wagner et al. (1988) found that even without a strong worker participation element, a foundry improved its productivity by more than 100% after implementing gain sharing, and experienced statistically significant decreases in labor costs and grievances. Moreover, the authors also reported greater employee concern for cooperative behaviors as well as coworker “policing” of quantity and quality to assure equitable contributions.

Moderating Variables in Group Pay-For-Performance Plans

One widely replicated finding is that group size is a moderator of group plan effectiveness, consistent with the idea that expectancy perceptions weaken as each individual sees less effect of his or her own efforts on total group output. For example, Kruse (1993) reported that the impact of profit sharing on annual productivity growth was roughly twice as great in companies having fewer than 775 employees than in larger companies. Similarly, Kaufman (1992) reported that doubling the number of employees covered by a gain-sharing plan from around 200 to 400 was associated with a reduction in the average productivity gain of nearly 50%.

Another variable that has long been hypothesized to moderate the effectiveness of group plans, as well as their specific design components, is the extent to which tasks are truly interdependent. For example, Shaw et al. (2002) found that in an industry where tasks are mostly independent (long-distance trucking), firm accident rates and time spent out of service were lowest (i.e., performance was highest) when there were high individual pay incentives and high pay dispersion across drivers. In contrast, in a second study in the concrete pipe industry, they found that safety performance was lowest when high pay differentiation was coupled with high task interdependence (measured via the existence of self-managed teams).

Summary

Every pay program has its advantages and disadvantages. Programs differ in their sorting and incentive effects, their incentive intensity and risk, their use of behaviors versus results, and their emphasis on individual versus group measures of performance. Because of the limitations of any single pay program, organizations often elect to use a portfolio of programs, which may provide a means of reducing the risks of particular pay strategies while garnering most of their benefits (Gerhart et al. 1996). For example, using only an individual incentive program could result in unacceptably high levels of competitive behavior and a focus on overly narrow objectives. On the other hand, relying exclusively on gain sharing could result in the under-rewarding of high individual performers, thus risking their attraction, motivation, and retention. However, offering a mix of these different programs offers the possibility that the advantages of each can be captured, while minimizing the disadvantages.
FUTURE RESEARCH

Moving to the Field

At several places in this review, we have discussed the issue of questionable generalizability from laboratory research on PE and pay (particularly when performed on children) to the reactions and behaviors of employees in the workplace. Among the issues we noted are the smaller amounts of pay available in the lab, the shorter time frames involved, and the fact that adults generally expect (and often need) to be paid for working, but not necessarily for solving puzzles in a laboratory.

We are aware that in many areas of personnel psychology, laboratory results generalize well to the field (Locke 1986). However, there is already some evidence that this may not be the case with respect to the relationship between intrinsic and extrinsic motivation (e.g., Fang & Gerhart 2000), and we suspect that this may be the case for certain aspects of performance feedback as well (see below). As such, although there is still a role for laboratory research in PE and compensation, at present we see a far greater need for research that is conducted under conditions that are more realistic (see also Arvey & Murphy 1998, Rousseau & Fried 2001 with respect to the growing recognition of the importance of context in industrial/organizational research).

For example, in addition to the issues discussed earlier with respect to intrinsic and extrinsic motivation, it is possible that some aspects of Kluger & DeNisi’s FIT theory will play out differently in real work settings. For example, we suspect that in many cases, employees who receive negative 360-degree feedback have their attention diverted—at least initially—away from managerial “tasks” to their own sense of self-worth, just as laboratory subjects do. According to FIT theory, such employees would be expected to be at risk of lower subsequent performance. However, on average, field research suggests that most individuals who receive negative 360-degree feedback subsequently improve as a result.

It may be that in real work settings, both the longer time frame and the crucial importance of succeeding at the task are likely to increase both effort and task focus, at least once the initial reactions of distress, anger, or self-denigration have worn off (see, e.g., Smither et al. 2003). In contrast, both the size of the incentives and the time available for emotional adjustment and learning usually are dramatically curtailed in the lab. This is just one more example of how findings from laboratory and field research might diverge due to substantially different contextual conditions in PE and pay research. As such, we applaud the field-oriented direction taken by 360-degree evaluation research and encourage more such research in future.

Intervening Processes

One of the most frequent calls in the burgeoning “strategic human resource management” literature has been for research that will help illuminate the “black
box” between various HR practices (e.g., profit sharing or 360-degree evaluation) and organizational outcomes (e.g., profits or growth; Becker & Gerhart 1996). Although economists and strategic management researchers have produced dozens of studies correlating various HR practices with organization-level outcomes, most such studies leave the reader guessing as to the causal processes involved and, consequently, how to enhance practical effectiveness.

To tease out causal processes as well as to provide guidance about how best to implement various PE and PFP programs, researchers need to begin measuring mediating psychological variables such as employee attitudes, beliefs, and behaviors. To date, few studies (in any discipline) have conducted cross-organizational research that simultaneously measures pay or evaluation policies, employee reactions or behaviors, and unit/organization performance. We believe psychologists could make particularly valuable contributions in this area.

One notable exception that addresses these issues is Klein’s (1987) study of 2804 employees covered by 37 employee stock ownership plans (ESOPs). Klein hypothesized that ESOP performance would be mediated by three factors: (a) mere “pride of ownership” from belonging to an ESOP, (b) the amount of the employer’s financial contribution to the ESOP, and (c) the extent to which the ESOP was a core part of the management philosophy and communications program. She found support for the last two factors, suggesting that managerial commitment to group plans is indeed associated with their degree of success.

Studies of intervening processes, particularly longitudinal ones, might also be used to reveal much more about the difficulties of successfully implementing and maintaining new PE or PFP practices. We know, for example, that many gain-sharing and individual incentive programs have been discontinued due to implementation difficulties. For example, in the Petty et al. (1992) study reported above, despite better performance on 11 objective performance measures and a variety of perceptual measures, the gain-sharing plan was discontinued because of disagreements between management and the union about how to divide the gains among employees if the plan were to spread to other units.

In another study, Pritchard et al. (1988) reported that a combined goal-setting, feedback, and incentive intervention was dismantled—despite productivity improvements of up to 75%—after the arrival of a new manager who was philosophically opposed to the use of incentives. There was also resistance from people who believed that employees “should not get something for doing what they are already supposed to do” (p. 354), as well as from some supervisors who felt that the incentive would undermine “their power and prerogatives to reward individuals and units informally” (p. 354). One thing both these studies illustrate is the serious difficulty of aligning perceived conflicts of interest between individuals’ (both employees’ and managers’) goals and objectives and those of the broader organization. In any event, these two studies are exemplary in terms of presenting important information on both causal processes and practical implementation challenges.
Sorting and Individual Differences

Another area in which psychologists could make valuable contributions would be to build on the progress that has been made in integrating individual difference variables into pay research (e.g., Cable & Judge 1994). In particular, we encourage additional work on compensation systems that investigates the distinction between incentive and sorting effects, the latter of which appear to be very important based on the limited work available (e.g., Lazear 1999, Trevor et al. 1997).

In particular, the sorting effects of group incentive plans need further attention, particularly since some studies suggest that there may be higher turnover of high performers under such plans (e.g., Park et al. 1994, Weiss 1987). However, it is also possible that contextual variables such as strong communications, inspirational leadership, a highly participative culture, or limited alternative employment opportunities might ameliorate or even reverse these potentially negative sorting effects (e.g., Hamilton et al. 2003, Reichheld 1996).

In addition, we encourage increased attention to individual differences in how people perceive risk (Gomez-Mejia et al. 2000). Despite the fact that risk is a central factor in many current forms of pay (e.g., incentives, gain sharing, profit sharing, and stock options), little attention has been given to its measurement. Recent work by Hall & Murphy (2000) and Wiseman et al. (2000) provides a glimpse of just how important risk can be in affecting employee perceptions of value. The rapid growth of “risky” pay, in combination with the dramatic fluctuations in stock value and profits in recent years, suggests that this may be an individual difference variable of considerable importance. More generally, the increased risk of job loss at all employment levels (Reingold 2004) may also make risk an important topic in future investigations of PE, although it has not been so to date. Psychologists would seem well suited to take on this challenge.

360-Degree or Multisource Performance Evaluation

Some of the potential advantages of using multiple raters in 360-degree appraisal include higher reliability, credibility, and perhaps lower deficiency (e.g., traditional appraisals capture supervisors’ perspectives but not those of peers, customers, or subordinates). With better PE measures, improved perceptions of fairness may make it possible to tie pay more closely to performance, which in turn should translate into employees seeing a stronger link between behaviors and rewards. At present, however, too few companies use multisource feedback for administrative purposes (Smither et al. 2004) to provide reliable tests of these hypotheses.

Given the generally modest average effect sizes but high variability within and across studies (Smither et al. 2004), future 360-degree research would do well to focus on determining which individual or program characteristics are most likely to improve the returns from multisource PE. Research to this point suggests that a variety of actions (e.g., discussing results with subordinates or coaches or setting goals) may be associated with improvements in performance following
multisource feedback. However, the fact that some employees take such actions while others do not suggests that individual differences in underlying traits such as conscientiousness, need for achievement, or mastery orientation may be important precursors of observed responses to multisource feedback.

With respect to situational differences, it seems that most implementations of 360-degree are quite weak in terms of providing motivation for improvement (Smither et al. 2004). By deliberately separating multisource feedback from organizational rewards, 360-degree implementation has nearly always taken the same route suggested by Meyer et al. (1965, 1989) in their “split roles” perspective on traditional appraisal. (For the logic behind the development-only approach to 360-degree appraisal, see London & Smither 1995 or Waldman et al. 1998.) Although most 360-degree researchers have speculated that linking multisource evaluations to money or other administrative consequences is unwise, we believe this is an important matter for future research. In saying this, we are taking into account our findings that the use of PFP can be considerably more effective than it is portrayed in much of the psychological literature, and that research has failed to validate Meyer et al.’s claims about the dangers of integrating feedback and reward in more traditional appraisal settings (e.g., Boswell & Boudreau 2002, Dorfman et al. 1986).

Merit Pay

We mentioned earlier how little well-designed research exists on the most common PE–pay combination of all: merit pay systems combined with subjective supervisory appraisals. Additional research is sorely needed of the quality exemplified by Klein (1987), Petty et al. (1992), or Pritchard et al. (1988) with respect to ESOPs and incentive systems—i.e., studies that include control groups, intervening process variables, and/or careful longitudinal analyses.

One important priority for future research is to conduct more studies that examine the link between the operationalizations of merit pay programs and unit- or organization-level performance. The optimal sites for such research will probably be in multiple units of the same organization (the strategy used by Pearce et al. 1985). Evidence should be gathered concerning the relationships between variability in both PE ratings and merit increases, the size of performance rating–merit increase correlations, and organizational performance. Ideally, these would be gathered in combination with measures of employee satisfaction, perceived fairness, and employee turnover that might help to tease apart the motivational and sorting processes involved in producing the observed outcomes.

In addition, future research should also incorporate the indirect (but potentially very important) effects of merit ratings on pay via their nontrivial influence on promotion, cumulative (year-after-year) earnings, and quality of employees attracted and retained (Gerhart & Milkovich 1992). Such modifications are almost certain to demonstrate stronger PFP linkages (and, consequently, larger impacts on performance) than those reported to date in the merit pay literature.
CONCLUDING REMARKS

In most organizations (at least in the United States), PE is used both to provide developmental feedback and to motivate employees via linkages between PE and rewards. However, psychological research has focused primarily on the former role to the exclusion of the latter, whereas other disciplines (e.g., economics and finance) have tended to do the opposite.

This is unfortunate, because in the real world of organizations, PE and PFP are two of the most powerful tools in an organization’s motivational arsenal (Welch 2001). Yet, in some branches of psychological research, PFP is still assumed a relatively ineffective motivator or even an impediment to motivation (Deci et al. 1999, Kohn 1993). However, as our chapter indicates, money is a very powerful motivator indeed. In fact, it is so powerful that one of the main challenges for managers is to make sure that their compensation systems are not motivating the wrong kinds of behaviors.

Psychologists potentially have much to contribute to the practical problems faced by managers who want to know the most effective and motivational ways of evaluating, developing, and rewarding people. However, psychology’s contribution will remain less than it otherwise might be if some of the evidence revealed in this chapter—e.g., that monetary rewards and intrinsic motivation generally do not work in opposition to one another among working adults and that, on average, performance improves after negative feedback—remains unknown by many psychologists who are interested in industrial and organizational issues. As such, we appreciate the wisdom of the editors in seeing the need for a chapter that addresses both PE and PFP simultaneously. We hope that it results in new and vibrant psychological research.

ACKNOWLEDGMENTS

We wish to thank Angelo DeNisi, Susan Fiske, Robert Heneman, and James Smither for helpful comments on an earlier draft of this manuscript.

The Annual Review of Psychology is online at http://psych.annualreviews.org

LITERATURE CITED

Atwater LE, Waldman DA, Atwater D, Cartier P. 2000. An upward feedback field experiment: supervisors’ cynicism, follow-up and


Krim J, Spinner J. 2003. Microsoft employees...


## CONTENTS

Frontispiece—Richard F. Thompson  xvi

### PREFATORY

In Search of Memory Traces, Richard F. Thompson  1

### DECISION MAKING

Indeterminacy in Brain and Behavior, Paul W. Glimcher  25

### BRAIN IMAGING/COGNITIVE NEUROSCIENCE

Models of Brain Function in Neuroimaging, Karl J. Friston  57

### MUSIC PERCEPTION

Brain Organization for Music Processing, Isabelle Peretz and Robert J. Zatorre  89

### SOMESTHETIC AND VESTIBULAR SENSES

Vestibular, Proprioceptive, and Haptic Contributions to Spatial Orientation, James R. Lackner and Paul DiZio  115

### CONCEPTS AND CATEGORIES

Human Category Learning, F. Gregory Ashby and W. Todd Maddox  149

### ANIMAL LEARNING AND BEHAVIOR: CLASSICAL

Pavlovian Conditioning: A Functional Perspective, Michael Domjan  179

### NEUROSCIENCE OF LEARNING

The Neuroscience of Mammalian Associative Learning, Michael S. Fanselow and Andrew M. Poulos  207

### HUMAN DEVELOPMENT: EMOTIONAL, SOCIAL, AND PERSONALITY

Behavioral Inhibition: Linking Biology and Behavior Within a Developmental Framework, Nathan A. Fox, Heather A. Henderson, Peter J. Marshall, Kate E. Nichols, and Melissa A. Ghera  235

### BIOLOGICAL AND GENETIC PROCESSES IN DEVELOPMENT

Human Development: Biological and Genetic Processes, Irving I. Gottesman and Daniel R. Hanson  263
CONTENTS

SPECIAL TOPICS IN PSYCHOPATHOLOGY
The Psychology and Neurobiology of Suicidal Behavior,
Thomas E. Joiner Jr., Jessica S. Brown, and LaRicka R. Wingate 287

DISORDERS OF CHILDHOOD
Autism in Infancy and Early Childhood, Fred Volkmar,
Kasia Chawarska, and Ami Klin 315

CHILD/FAMILY THERAPY
Youth Psychotherapy Outcome Research: A Review and Critique
of the Evidence Base, John R. Weisz, Amanda Jensen Doss,
and Kristin M. Hawley 337

ALTRUISM AND AGGRESSION
Prosocial Behavior: Multilevel Perspectives, Louis A. Penner,
John F. Dovidio, Jane A. Piliavin, and David A. Schroeder 365

INTERGROUP RELATIONS, STIGMA, STEREOTYPING,
PREJUDICE, DISCRIMINATION
The Social Psychology of Stigma, Brenda Major
and Laurie T. O'Brien 393

PERSONALITY PROCESSES
Personality Architecture: Within-Person Structures and Processes,
Daniel Cervone 423

PERSONALITY DEVELOPMENT: STABILITY AND CHANGE
Personality Development: Stability and Change, Avshalom Caspi,
Brent W. Roberts, and Rebecca L. Shiner 453

WORK MOTIVATION
Work Motivation Theory and Research at the Dawn of the Twenty-First
Century, Gary P. Latham and Craig C. Pinder 485

GROUPS AND TEAMS
Teams in Organizations: From Input-Process-Output Models to IMOI
Models, Daniel R. Ilgen, John R. Hollenbeck, Michael Johnson,
and Dustin Jundt 517

LEADERSHIP
Presidential Leadership, George R. Goethals 545

PERSONNEL EVALUATION AND COMPENSATION
Personnel Psychology: Performance Evaluation and Pay for Performance,
Sara L. Rynes, Barry Gerhart, and Laura Parks 571
CONTENTS

**PSYCHOPHYSIOLOGICAL DISORDERS AND PSYCHOLOGICAL EFFECTS ON MEDICAL DISORDERS**

Psychological Approaches to Understanding and Treating Disease-Related Pain, *Francis J. Keefe, Amy P. Abernethy, and Lisa C. Campbell* 601

**TIMELY TOPIC**

Psychological Evidence at the Dawn of the Law’s Scientific Age, *David L. Faigman and John Monahan* 631

**INDEXES**

Subject Index 661
Cumulative Index of Contributing Authors, Volumes 46–56 695
Cumulative Index of Chapter Titles, Volumes 46–56 700

**ERRATA**

An online log of corrections to *Annual Review of Psychology* chapters may be found at http://psych.annualreviews.org/errata.shtml