Human Resource Management Review 24 (2014) 41–52

Contents lists available at ScienceDirect

Human Resource Management Review

journal homepage: www.elsevier.com/locate/humres

Pay for (individual) performance: Issues, claims, evidence and the role of sorting effects

Barry Gerhart *[,a], Meiyu Fang [b]

[a] University of Wisconsin-Madison, 4194 Grainger Hall, Madison, WI 53706, United States
[b] National Central University, Taiwan

ARTICLE INFO

Keywords:
Compensation
Motivation
Pay for performance
Performance
Incentives
Sorting

ABSTRACT

We address two broad questions: how much pay for individual performance (PFIP) is there and what are the positive and negative effects of PFIP? We consider specific claims, including that PFIP does not motivate (or even de-motivates), that it is ineffective in teams, and that it is ineffective in some national cultures. We demonstrate how incorporating sorting effects of PFIP into conceptual treatments of PFIP can change how one views the likely effectiveness of PFIP across contexts.

1. Introduction

Pay or employee compensation is a topic that generates considerable interest among scholars and practitioners. Over the years, they have asked questions like how important is money in motivation, to what degree do organizations pay for performance (especially individual performance), does paying for performance contribute to higher performance or does paying for performance (again, especially individual performance) sometimes have undesired (and often unanticipated) consequences? The present paper will focus on pay for individual performance (PFIP) and will offer a selective survey that builds on classic work, but primarily highlights recent conceptual and empirical developments.[1] In particular, the present paper will demonstrate how taking fuller account of sorting effects, the effect that PFIP has "on the composition (e.g., in terms of ability, personality) of an organization's workforce via attraction, selection, and attrition processes" (Gerhart & Rynes, 2003, p. 151), can change thinking regarding PFIP's effects in several arenas.

To begin, we classify pay for performance programs on three dimensions (Gerhart & Rynes, 2003): the emphasis on results-oriented (e.g., physical production output, sales, profits, total shareholder return) or behavior-oriented performance (e.g., supervisor or customer ratings of specific behaviors) measures, the emphasis on performance measured at the individual employee or aggregate (e.g., group, unit, organization) level, and the degree of incentive intensity. Virtually all private sector organizations in the United States (and in many other countries) use multiple pay for performance plans with the design (see the preceding three dimensions) varying as a function of factors such as job level, type of occupation, the way the work is organized, and the organization's strategy (Gerhart, 2000; Gomez-Mejia & Balkin, 1992; Gomez-Mejia, Berrone, & Franco-Santos, 2010; Milkovich, Newman, & Gerhart, 2013). For example, results-based measures are more likely to be used for sales occupations, executives, and other positions where objective performance measures are available. Incentive intensity is typically stronger in higher level jobs and/or in jobs where results-based performance measures are available. Merit pay (a focus on subjective judgments of performance behaviors) is found in some form for most jobs, though it is less likely to be used for lower level positions, especially when a labor union is present. However, even in this context, one is increasingly likely to find plant and/or firm level performance-based plans (e.g., gain sharing, profit sharing).

* Corresponding author. Tel.: +1 608 262 3895.
E-mail address: bgerhart@bus.wisc.edu (B. Gerhart).

[1] More complete reviews, especially of earlier work, as well as of work on other forms of pay for performance and on other dimensions of compensation (e.g., pay level, pay structure, benefits) are available elsewhere (Gerhart & Milkovich, 1992; Gerhart & Rynes, 2003; Gerhart et al., 2009; Gibbons, 1998; Lawler, 1971; Lazear, 1995; McKenzie & Lee, 1998; Prendergast, 1999; Rynes et al., 2005).

1053-4822/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.hrmr.2013.08.010

As noted, our primary focus here is on PFIP plans. Theories of compensation (e.g., reinforcement, expectancy, efficiency wage, agency) largely agree that incentives and reinforcement (central to any PFIP plan) are key drivers of important workplace behaviors such as employee performance and employee attraction/retention. Indeed, there is overwhelming (meta-analytic) evidence that incentives drive behavior and that the effect is substantial (Locke, Feren, McCaleb, Shaw, & Denny, 1980; Gupta & Shaw, 1998; Guzzo, Jette, & Katzell, 1985; Jenkins, Mitra, Gupta & Shaw, 1998; Judiesch, 1994; Stajkovic & Luthans, 1997). (Gerhart & Rynes, 2003 and Gerhart et al., 2009 provide reviews of the broader literature on pay for performance consequences.) In addition, PFIP decisions may be especially strategic. For example, Locke et al. reported a 30% improvement in productivity due to the introduction of incentive plans. Gerhart and Milkovich (1990) found that organizations differentiate more with regard to how they pay (e.g., how much they rely on PFIP), somewhat less on how much they pay, and that organizations placing more emphasis on PFIP performed better on average.

So, given the evidence that PFIP has positive effects, what is there to write about? Over the years, challenges to the conceptual underpinnings of PFIP have been raised, as have important and real-world challenges in successfully implementing PFIP. For example, work by Herzberg directly questioned whether money was an important motivator in the workplace, relegating money to the "hygiene" category of motivation-hygiene theory. Deci (1975) and Deci & Ryan (1985) went a step further than simply arguing that money did not motivate. They argued that money could demotivate. Specifically, their model makes the case that under common circumstances, PFIP undermines intrinsic motivation and interest to perform a task or activity. This hypothesized undermining effect of PFIP, which is based on research conducted in non-work settings (often using children as subjects in an education setting), was subsequently picked up by others (e.g., Kohn, 1993; Pfeffer, 1998) and extrapolated to workplace settings. Most recently, this undermining effect of PFIP has been popularized in a bestselling book, *Drive* (2009), by Daniel Pink (Pink was also the keynote speaker in 2011 at the World at Work, formerly known as the American Compensation Association, annual conference, thus having the maximum opportunity to communicate his message to compensation professionals).

Other concerns have to with PFIP's fit with different situational contexts. As one example, does PFIP fit the "new economy" where the use of teams is thought to be greater, with greater importance placed on collaboration and cooperation (as opposed to narrower individual performance goals)? As another example, even if PFIP is seen as continuing to be a central part of managing people in a country like the United States, does PFIP fit other countries that differ on dimensions such as national culture? For example, China is now the world's second largest economy (after the United States) and has a very different national culture profile (e.g., less individualistic; Hofstede, 1980, 2001; House, Hanges, Javidan, Dorfman, & Gupta, 2004) than the United States. Does that mean PFIP will not work well in China?

Two final concerns stand in direct opposition to one another. One concern is that although organizations say they use PFIP, when one looks at the actual pay-performance relationship, it sometimes does not appear to be strong enough to be meaningful. In other words, the incentive intensity of PFIP in this case is too weak to be of any consequence. A second, opposite concern, is that PFIP can be implemented with such strong incentive intensity (often using results-based performance measures) that the risk of serious unintended negative consequences becomes greater (e.g., top executives manipulating profits and/or taking too little or alternatively, excessive risks, Martin, Gomez-Mejia, & Wiseman, 2013; Sanders & Hambrick, 2007; employees focusing on quantity rather than quality or customer satisfaction and/or gaming the incentive system, rate cutting and goldbricking, Gerhart, 2001; Roy, 1952; Whyte, 1955; employees paying attention only to performance objectives explicitly covered in the incentive plan, Lawler, 1971; Milgrom & Roberts, 1992; Wright, George, Farnsworth, & McMahan, 1993). In this situation, agency theory (Berle & Means, 1932; Fama & Jensen, 1983; Holmstrom, 1979; Prendergast, 1999), with its focus on agency costs (adverse selection and moral hazard) that arise from information asymmetry and goal incongruence is a standard tool for understanding what can go wrong and also helps make apparent that when PFIP is designed to have strong incentive intensity, PFIP can be described as a high risk, high return strategy (Gerhart, 2004; Gerhart, Trevor, & Graham, 1996). Agency theory is also useful because of its recognition that while higher incentive intensity increases the magnitude of the incentive effect, it also increases the potential for agency costs and thus, unintended negative consequences. Higher incentive intensity also increases the amount of risk to be borne by employees, which can easily translate into employee relations problems (especially when the incentive plan does not pay out).

Another concern is that when one digs into the evidence on PFIP in work settings, there is less there than one might have expected. For example, the meta-analytic evidence cited earlier, which shows that PFIP has a strong positive impact on performance pertains to a narrow set of contexts. Specifically, the evidence comes primarily from studies that use results-based (objective) measures of performance and from settings where work is simple, easy to measure, and not interdependent.[2] Gerhart and Milkovich (1992) noted that these types of PFIP plans "are not applicable to many jobs" (p. 523) and that there are often "administrative problems" with such plans (p. 523), which include the cost of setting and maintaining setting standards, both in financial terms and in employee relations terms (e.g., due to gaming, rate cutting, and so forth). As such, there may be a selection bias when we study only the plans that survive, at least to the degree that there are many failed plans that do not make it into our sample (Gerhart et al., 1996).

Merit pay plans, with their broad applicability (i.e., to any job where a performance rating can be obtained) and behavior-oriented focus (which allows one to consider whether results are obtained in a way judged to be in keeping with the organization's values and goals) would seem to provide the solution to the potential limitations of results-based PFIP plans. However, merit pay plans have their own challenges, both in terms of execution and in terms of how to design research to study their effects (Milkovich & Wigdor, 1991). Unfortunately, there is little rigorous research on the direct impact of merit pay plans because of challenges in doing such research (Heneman & Werner, 2000; Rynes, Gerhart, & Parks, 2005). On the execution side, as noted, there is often doubt as to whether merit pay

[2] Consider, for example, the meta-analysis of financial incentives conducted by Jenkins, Mitra, Gupta, and Shaw (1998). Of the 47 studies covering 3124 employees they were able to find eight studies that were conducted in a field setting, covering 470 employees. The performance measures in the eight studies were: number of trees planted (2), number of animals trapped (2), supervisory/behavioral ratings (2), exam completion time (1), number of items tested in a manufacturing setting (1). Exam completion time involved transforming a number by adding digits to it and looking it up in a catalog.