The Impact of Love and Money on Performance in the Federal Government: Implications for Pay for Performance

Laura Langbein

Dept. Public Administration and Policy

American University

Abstract:

I summarize what we can learn from personnel economics about the likely empirical veracity of the statements from OPM and ICMA asserting the efficacy of pay-for-performance for public employees. I then provide empirical evidence from the U. S. federal government showing that, controlling for indicators of human capital, agency location, and many other variables, that individual discretion on the job and intrinsic satisfaction with the job are more predictive of a (subjective) measure of individual performance than pay satisfaction and pay for performance (i.e., detailed lists of "objective" measures, or closed contracts). Higher salaries predict higher performance, but the relative impact of objective pay is less than the impact of intrinsic values and discretion.

DRAFT 091209

Prepared for the Public Management Research Conference, Columbus, Ohio, Oct. 1-3, 2009.

The Impact of Love and Money on Performance in the Federal Government:

Implications for Pay for Performance

Laura Langbein

Dept. Public Administration and Policy

American University

Background

According to a recent (Nov. 2007) item in the federal government press (Risher, 2007), pay for performance (PFP) unquestionably "works." In this article, titled "Pay for Performance: The Road to Success," the goal of PFP is to improve performance (in the public sector), presumably of both the employee and the organization. The article states that PFP works because it changes the basis for salary increases, so that the employee's performance affects the employee's pay increase. The article, published by the International City Managers Association (ICMA), makes several specific empirical assertions, but it does not provide evidence to verify them. Among the statements are the following:

PFP is effectively universal in the private sector.

The use of PFP is growing at all levels of government

The use of PFP can reduce turnover, especially among the most qualified.

PFP helps employees focus on results important to the organization.

PFP attracts better-qualified workers.

In this paper, I review the literature in economics on the likely impact of performance measurement and performance pay on productivity in the private sector. I then examine

the impact of the role of pay (i.e., money) on the likely behavior of public sector employees, testing the hypothesis: "The use of PFP can improve performance, especially with respect to results important to the organization."

There is a large literature in public administration and public management that raises questions about the rationale, effectiveness, equity, and efficiency of PFP in the public sector. (See for example Perry, 1986; Perry, 2007; Mintzberg, 1995; Moynihan, 2007; Oh and Lewis, 2009.) In this paper, I ask whether statements like those made by ICMA can be justified as theoretically obvious or empirically supported from the point of view of rational choice of employers and employees operating in a competitive labor market. I also examine the likely role of PFP in the public sector by considering the relation between liking to work for money (assumed by PFP), liking your work (in the context of the public sector, sometimes called public service motivation (PSM), the detail with which your performance is measured or rated (necessitated by PFP), and your performance relative to others in your work group and agency.

To do this, I rely on a review of literature from a disciplinary field that is believed to be particularly supportive of PFP: economics. The standard model of labor markets assumes is that people are indifferent about the purpose of their jobs, dislike effort, but will work hard if they are paid for doing so; that more pay elicits more and better performance; and that pay that equals marginal revenue product elicits efficient production, such that the value of the marginal product just equals than the marginal labor wage. (Stiglitz and Shapiro (1984) show that fficiency wages will exceed marginal product to prevent shirking and to prevent productive employees from moving costlessly to another employer.) Yet it turns out that, in most circumstances, this is not what

happens in the "real" world, and, from the perspective of social efficiency, is not even what "should" happen in all circumstances.

In fact, economists whose professional subfield is that of personnel economics have tested many of the ICMA assertions, and much of the evidence (in economics) does not support them. A few of these "tests" use the standard approach of comparing empirically generated observations, experimentally or non-experimentally manipulated, to a hypothesis. However, most of the tests use formal models that produce "interesting" and "non-obvious" statements, which, based on rational choice by profit and utility maximizing actors, are unanticipated prior to the model being solved (Clarke and Primo, 2007). The conclusion is that, under many circumstances particularly likely to characterize many public and complex private organizations, PFP may have unintended adverse consequences. But, beyond supporting stories, there is a dearth of systemic empirical evidence to support these conclusions, especially with data from the public sector, where the theories should also apply.

In this paper, I summarize what we can learn from personnel economics about the likely empirical veracity of the statements from OPM and ICMA listed above. I then provide empirical evidence using the 2005 Merit Systems Protection Board (MSPB) survey of the U. S. federal government employees. Unlike recent research by Oh and Lewis (2009), I do not rely on attitudes about the fairness or opinions about the effectiveness of performance assessment systems to generate evidence. Instead, I directly measure job and pay satisfaction, and characteristics of performance assessment systems. Evidence from the second survey shows that detailed lists of "objective" measures (closed contracts), regardless of how closely they are linked to pay, have a positive, but

small, effect on performance; linking the performance assessment to pay has no impact on performance. Actual pay levels also have a significant and small effect on performance, and so does satisfaction with pay. But liking the job (compared to liking the pay) and having job discretion also have a positive and greater effect on performance, compared to the monetary drivers.

The View from Personnel Economics

The costs of measuring performance

Lazear and Shaw (2007) note that pay for performance in the private sector is less common when monitoring costs (i.e., the costs of measuring output) are higher. As the costs of measuring output drop, the use of performance pay increases. When output is easy to measure, pay for performance improves both individual and organizational performance. For example, Lazear (2000) finds that, when the employer switched from hourly to piece-rate pay, workers who install new windshields on vehicles to replace damaged ones increased productivity. Because performance in this case is easy to measure, pay for performance reduced errors and increased speed. Similarly Baker and Hubbard (2003) find that, when technological innovations made it easier to monitor the performance of long-haul truckers, firms that needed to ship goods found it more attractive to hire truckers by contract rather than to employ their own truckers. When performance is easy to measure, closed contracts are easier to write and implement; consequently, contracting out becomes relatively more attractive. Even in the pharmaceutical industry, when performance is easy to measure, the activity is more likely to be outsourced, because it is easier to tie pay to performance. For example, Azoulay

(2004) finds that pharmaceutical firms out-source making the pills, but "invent" them in-house. Lazear and Shaw (2007) present evidence of the predominance and effectiveness of performance pay in other occupations (public and private) where performance is easy to measure, including Navy recruiters, jockeys, tree-planting firms, and salesmen of industrial products.

However, performance is not always easy to measure. In fact, when tasks are complex, performance pay may be worse than alternative forms of pay and contracts. For example, it appears that open (i.e., vague) contracts (with low powered incentives) are socially optimal when jobs are complex and relations between the principal(s) and agents(s) are repeated. Numerous theorists (Gibbons, 1998; Chen, 2002; Hartwig, 2004; Frey and Oberholzer-Gee, 1997; Kreps, 1997; Frey, 1993; Nagin et al. 2002) add that, when output is hard to measure and the production function is unknown, incomplete (unspecified) contracts between employees and managers may be more efficient than complete contracts.

The errors in measuring performance

Measuring performance is tricky; consequently, pay for performance is really paying for what is measured as performance. Widely recognized in public administration as goal displacement, one implication in all organizations is that you may hope for performance A but you pay for performance B (Gibbons, 1998). In the context of the private sector, one consequence may be that performance pay raises output but not profitability: performance pay induces effort away from other goals (Lazear and Shaw, 2007). In the context of the non-profit sector, rewarding organizations for measured

performance can detract from socially preferred outcomes (Bohte and Meier, 2000). In a meta-analysis of studies of the public sector in the UK, Prentice et al. (2007) report considerable evidence of "gaming," implying that employees use resources that do not increase productivity. The possibilities for goal displacement are not discussed in a report on the expected benefits of performance-based management in several federal agencies (Association of Government Accountants, 2009).

At the individual level, Langbein (2008) shows that basing faculty raises on student evaluations of teaching (SETs) results in inflated grades, which can distort external labor markets. Correspondingly, at the organizational level, the impact on grade inflation increases with greater reliance on SETs as a pay incentive (Rojstaczer, 2003).

There is a related literature (in public management and personnel economics) that investigates production in teams when performance is hard to measure pointing to the role of mid-level supervisors who can help to clarify goals or use persuasion (without using PFP). It appears that persuasion or clarifying goals may induce higher effort by agents at no cost to the principal (Brehm and Gates, 2008; Jung and Rainey, 2009; Van den Steen, 2009; Mader, Myers, and Kelman, 2009). (Persuasion and goal clarification may in fact be the same thing.) When principal's (or, in the public sector, principals') goals are unclear, mid-level managers can reduce agent's uncertainty, inducing more effort supporting the principals' goals (May and Winter, 2007; Langbein, 2000).

Heterogenous preferences: chocolate or vanilla

Paying for performance can also be costly because workers have heterogeneous preferences. Workers who value intrinsic internal motivation are likely to prefer salary-

type jobs (at lower pay than a worker who prefers the risk and potentially larger return from pay-for-output). Performance pay selects for (and retains) employees who like to be paid based on measurable, short-term output (Lazear, 2000; Lazear and Shaw, 2007). Firms use different types of pay systems to encourage workers to sort themselves to the firm that values them the most and that they also value the most. There is a signaling component not only in respect to the amount of the compensation package but also with respect to form of the package. Lazear and Shaw 2007 cite as an example that high quality scientists often are willing to work for less in order to work on "good" projects, and firms offer them a project/pay mix that attracts the type of employee they want. In the context of PSM, an implication is that, while money isn't everything, almost everything can be monetized. Because some workers are willing to accept less for a job they value intrinsically, markets actually monetize the non-pecuniary aspects of a job. This also means that PSM theories can be tested just like any other model based on rational choice. For example, Taylor and Taylor (2009) show that governments pay efficiency wages, but also that the elasticity of worker effort is greater with respect to increases in PSM than with respect to increases in pay. One implication is that intrinsically motivated employees can be paid less than employees whose main motivation for work is monetary. Another implication is that both money and motivation matter: the world is not populated exclusively by either knaves (who love money) or knights (who love their job). If all were knights, then incentive systems (of any sort) will not matter; but even knights need to be told what to do. That is, they need some sort of performance measure (but not pay for performance) (LeGrande, 2009; Mader, Myers and

Kelman, 2009). A world populated by both knights and knaves requires both pay and performance measurement, but not necessarily pay for (measured) performance.

Another implication is that, when skills are not firm-specific (e.g., computer programming), so that workers have other opportunities in competing firms, pay-for-performance can be used to retain workers (especially when the worker is young.) But in other cases, paying workers for performance can be more costly than other pay systems, including salaries based on inputs or the type of job.

Teams

Production in teams is increasingly common in both the public and private sector. In the private sector, it is most likely to be used when different skills are needed for the task, and when persons who each have all the different skills are rare. In this case team production can be cheaper and/or more productive than production by separate individuals (Mas and Moretti, 2009). But there is an important impossibility result: when production is team-based, outcome-based rewards, whether awards to the group or the individual, cannot perfectly align employee self-interest and organizational efficiency (Miller and Whitford, 2002; Holmstrom, 1982). Instead, a more optimal response is to rely on reputation, which means selecting and retaining employees who intrinsically value the work they are doing. This not only reduces costs of monitoring by the principal; it also means that the agent doesn't need to be paid as much. An implication is that efficiency wages may be lower in a team-based PSM market. (Monitoring will not disappear in this market. The principal still has an incentive to monitor, because the principal is the agent of multiple principals or bosses).

There is also theoretical and empirical evidence (both observational and experimental) that agents in teams do not free-ride on their teammates; instead, productive peers induce more productivity by each team member (Siegel, 2009; Mas and Moretti, 2009; Chen and Li, 2009).

When individuals work in teams to produce firm output, and when pay for performance is adopted, evidence suggests that performance incentives need to complement other aspects of human resource management (Lazear and Shaw, 2007). In fact, teamwork and group-based incentive pay appear to be complements and produce results better than teamwork and individually-based pay for performance (or either one alone). Teamwork is more likely to be present when individuals identify with their "team" or group (Chen and Li, 2009). When production is in teams and individuals are paid based on their own performance, individual workers not only try to make themselves look good but they may also try to make other workers, on their team or on other teams, look bad (Heap and Zizzo, 2009). Team production may not be optimal when high levels of trust within a group induce higher levels of distrust of outsiders. The result could be higher levels of non-cooperative behavior and lower net social benefits (Heap and Zizzo, 2009; Olson, 1982).

This is particularly likely when performance is ranked. Pay for individual performance in this situation results in less pay compression (i.e., greater variance) (Lazear, 1989). To reduce the incentive for sabotage, pay systems that reduce the variance in individual pay (i.e., more pay compression) may be optimal.

Respect and Relative position

The standard economic model implies that people value their income, and that they are rugged individualists who care about their income, and not their income relative to others. But people (and other animals) are social creatures with different preferences; some care more than others about non-monetary resources, including what others think about them. Among non-monetary attributes that (some) people value, employees often want respect, particularly from employers and those who have control over them. Ellingsen and Johannesson (2007) show that, when workers care about "respect" (or "esteem") (especially from an employer who values unobservable talent), workers may work harder under weak than clear material incentives. When A cares about B's beliefs about A's talent, then A will work hard to "signal" (unobservable) talent. If B appears to value A only for output and not for inherent talent, B will begin to closely monitor A to push for greater effort, but A will respond with less effort, since B has undermined A's attempt to signal her underlying talent. If employers B can pay employees A with "respect," then employees A should be willing to take these jobs for less monetary reward. (Employees will sort themselves accordingly.) Similarly, Benabou and Tirole (2003) develop a model showing that strong monetary incentives offered by principal B to employee A can have a hidden cost by undermining A's self-confidence. Such an incentive system may counterproductive if it signals to A that B does not value A's effort or output.

While absolute pay matters, relative income matters too. One empirical analysis (Torgler and Schmidt, 2007) showed that even when performance is easy to measure, but there is team production, the larger the income differences (within German soccer teams), the worse the performance of individual players. Other evidence (Fehr, Hoff, and

Kshetramade, 2008) shows that "spiteful preferences" reduce cooperative choices in joint production games that offer opportunities to cooperate, defect, and punish either cooperators or defectors. Spite, in this context, means that players will forego monetary gains (e.g. work less but also receive less pay) if it increases their payoff relative to that of another team member. Any gain that the spiteful individual A receives from cooperating with B is reduced if cooperation means that the difference between the payoff to A relative to the payoff to B is reduced. If the police get a bigger raise than the firemen, then the firemen feel less well off and may shirk.

Autonomy

In addition to respect, where the principal values each agent's competence, agent autonomy (i.e., agent's ability to control her job) also contributes to performance. Deci and Ryan (2000) cite literature (in psychology) showing that arrangements that fail to recognize the importance of agent autonomy and competence, and the alignment between agent preferences (intrinsic rewards) and the organization, are associated with poorer performance. The economics literature also recognizes the importance of respecting agent competence and the relatedness of agent preferences to organizational goals.

Consider for example the theories and evidence about agent autonomy summarized in George A. Akerlof and Rachel E. Kranton (2008). In their model, workers with standard preferences gain utility from the difference between income and effort (so that, as in traditional economics, effort is a cost). The principal wants the workers to choose the optimal action, but cannot observe it directly. However, while the optimal action is not observable, less optimal choices by the agent are observable. To observe whether the employee makes an optimal choice, the principal must appoint a supervisor. However, appointing a supervisor affects the agent's preferences or identity: with no supervisor, the

agent gains utility from working with the group; if the principal appoints a supervisor, the agent gains no utility from the group, because the only relation that counts is the relation between the agent and the supervisor. While supervision lowers the cost of incentive pay to the principal because of monitoring, it also requires more incentive pay to elicit the optimal action. If there is no supervision, the workers enjoy working together. While they might not produce the optimal action, they may well produce an intermediate outcome at lower cost to the principal, so that the net utility to the principal of the intermediate action without the supervisor is higher than that from the optimal action with the supervisor. Thus, agent autonomy can be optimal, especially when agents have (endogenous) preferences about their work and their work identity.

Otto H. Swank and Bauke Visser (2006) add that the optimality of delegation (compared to monitoring or "paying attention") depends on the accuracy of the supervisor's assessment of the agent's ability. When the supervisor can accurately assess the agent's ability (or effort, which is assumed to be a complement), delegation is a costless signal to the agent that the agent is performing well; close supervision is not only a direct cost but it may also be an inaccurate signal.

The lack of a supervisor, together with the difficulty of (objectively) measuring performance, presents opportunities for agents to shirk. Fuchs (2007) shows that the repeated game between principal and agents results in agents working (and not shirking), if they are paid efficiency wages and if they do NOT receive bonuses or feedback on performance until they are fired. Mas and Moretti (2009) show that, when individual output is not observable but group output is observable, the ability of individuals to observe and monitor each other, and to exert peer pressure, reduces the incentive of individuals to shirk and makes them work harder than when there is less interaction.

Hidden work

There are other consequences of behavior in employment settings where workers have heterogeneous preferences and some aspects of performance are particularly hard to measure. It is common, but not universal, that inputs to performance (e.g., hours worked) are easier to measure than outputs or outcomes. But in many cases inputs cannot be readily measured or observed by others. Hours worked (or "effort") are often not observable by employers, but outcomes (or outputs) are observable. (University employment is typical of this sort of work.) Besley and Ghattak (2005) construct a model in which (some) agents intrinsically value their work (given a minimally acceptable wage); outcomes are observable, but effort is not. Mission can be either exogenous (set by the principal) or endogenous (affected by the agent). When the agent is more motivated than the principal, and the agent has few or no outside options, no incentive pay to induce effort is needed. This result depends on the amount of agreement between agent and principal: mission agreement requires less incentive pay. If mission choice is contractible (endogenous) between principal and agent, then the principal can use mission choice to motivate the agent. There is also sorting in the labor market, so that principal/agent missions will be aligned and may reduce the reliance on performance pay (and reduce reliance on differences in performance pay). This depends on unemployment in the for-profit sector and on the number of outside options in the mission-oriented sector (which includes portions, but not the entirety, of the public and non-profit sector). The implication is incentives will be less high powered in mission-oriented production with motivated agents.

Intrinsic preferences

Prendergast (2008) shows that “firms, in general, hire agents who do not share their interests—instead, agents are disproportionately motivated to carry out a subset of what the firm cares about.” Because tasks in most firms are specialized, firms hire agents who are biased to do just ONE of the tasks that need to be done; for example, government contractors hire engineers who love to design but don’t care about cost control. Thus, intrinsic motivation (viewed in isolation) is costly; but firms with tasks whose performance is hard to measure assemble a package (“team”) of biased agents. Each agent loves her (non-contractible) job. The firm thereby spreads the risks and reduces the costs of hiring biased agents.

Most theory (and what limited empirical evidence there is) supports the expectation that, in norm-driven organizations, extrinsic rewards crowd out intrinsic rewards. (See Hartwig, 2004; Frey, 1993, 1997; Frey and Oberholzer-Gee, 1997; and Kreps, 1997.) Using an experimental design, Nagin et al. (2002) find that employees who "trust" their employer and view the employer positively cheat less and are less responsive than others to increases in monitoring (i.e., direct controls). Even in the case of primary schooling, where student achievement is commonly measured by standardized tests, merit pay, at least in the random-assignment Tennessee STAR study, had mixed success in rewarding the teachers who increased student achievement (Dee and Keys, 2004).

Benabou (2003) also investigates conditions under which explicit (high powered) rewards crowd out performance, compared to low powered alternatives. These conditions include information asymmetry between agent and principal (about own ability, about each other, and about the nature of the task); and the nature of the reward system, which signals not only the agent’s payoff, but also signals what the agent infers

about how much the employer knows about the agent. Empowering the agent is also part of the low-power reward and signal system, and can increase implicit rewards, while offering help can sometimes be insulting. It is important to stress that these results depend critically on the nature of the task and the multidimensional nature of the information asymmetry between principal and agent.

Repeated games

The difficulty of performance measurement together with the added complication of team production also suggests that discretion and low-powered incentives may be optimal (Anja Schöttner, 2008). In this formulation, team production means that the relations among team members (among agents A) and between team member and the manager (the principal P) are repeated. There is only one imperfect performance measure of output (e.g. quantity) that is produced by several tasks whose contribution to the (imperfectly measured) output is not verifiable. There are more tasks than agents, and the tasks are not separable (e.g., quantity and quality). Schöttner shows that optimal job design favors broad task assignment (i.e., a wide span of control and discretion given to one agent), with an implicit contract. Reward is based on the performance of the firm, which is hard to measure. Because the relations are repeated, the implicit contract is self-enforcing, so that P has little incentive to renege, and A has little incentive to shirk. While this model is probably more applicable to management than production tasks, it turns out that, in the context of production tasks, sometimes it may be better to let A (re)design the production process rather than to assign a specific task.

Bureaucratic drift

Years ago, Downs (1967) referred to a class of bureaucrats that he called "zealots." Today, they would be regarded as employees who have strong normative preferences for doing a particular line of work, and may prefer to work for public sector agencies that do the type of work they like. The theories and evidence in contemporary personnel economics and political economy note both the benefits and the costs of "zealots": the principal can pay them less if she gives them discretion to do their work, but they may overproduce (Niskanen, 1975). For the zealot, over-production is the most likely form of shirking; it implies that the marginal social costs of bureaucratic activities (MSC) exceed the marginal social benefits (MSB). Examples include evidence of excessive regulation and excessive protection against risks from terrorists and natural catastrophes. Zealots may also resist modern financial innovations, and ignore other aspects of their job, such as attention to costs, that appear to detract from the mission; hence they appear resistant to cost-effective changes (Besley and Ghattak, 2005).

Zealots may also pursue goals that are inconsistent with those of principals, but they provide a source of expert advice for principals, who face problems of loss of control when agents know more than principals (Sean Gailmard and John W. Patty, 2007). In this situation, policy motivated agents gain by controlling their own job tasks (discretion), and principals gain because the agents receive less pay than if they worked elsewhere. Principals also gain by having experts rather than clerks to carry out complex tasks, but lose because the experts' advice may be somewhat biased.

Performance, Performance Measurement, Pay, and Motivation in the U.S. Federal Workforce: An Empirical Examination

Heterogeneous preferences among employees mean employees who are not particularly policy or service motivated and seek to work for pay will prefer employers who reward them accordingly, and employers will prefer that sort of employee. By contrast, employees who are service-oriented, and work not only for pay but also for intrinsic satisfaction from the job may be willing to accept less pay in exchange for a looser link between their pay and some external measure of their performance. Public sector employees are likely to be more service oriented, but money matters too. It is also particularly difficult to measure performance in the public sector, partly because there are multiple principals with conflicting goals (Langbein, 2000), but also because the goals (or outcomes) are collectively valued, happen in the future, and have uncertain production functions. In these cases, a close reading of the literature in economics suggests that, while performance responds to money, it responds more to intrinsic motivation, clear goals and supervisors who can clarify goals, but not pay for measured performance. It also suggests that work occurs in teams, and that workers in productive teams will not free-ride, but will be more productive than otherwise comparable workers in less productive teams.

I test these expectations using data from the 2005 survey of federal employees carried out by the Merit Systems Protection Board (MSPB). The survey was mailed to a randomly selected sample of 74,000 employees in 24 Federal agencies, and produced a response rate of 50%. The MSPB reports that the demographic characteristics of respondents did not differ significantly from the characteristics of the survey population

(van Rijn, 2005). The results show that, controlling for indicators of human capital, agency location, and many other variables, that individual discretion on the job and intrinsic satisfaction with the job are more predictive of a (subjective) measure of individual performance than pay satisfaction and pay for performance (i.e., detailed lists of "objective" measures, or closed contracts). Higher salaries predict higher performance, but the relative impact of objective pay is less than the impact of intrinsic values and discretion.

The Dependent Variable

Measuring the productivity of individual employees is difficult in any organization that produces complex goods, or even simple goods assembled in teams. It is particularly difficult in federal agencies because, unlike many state and local public sector agencies, they do not directly provide services (e.g., HHS, Dept. of Education; Dept. of Transportation) or because they directly provide services whose outcomes are hard to measure, or where the link between the service and the outcome are not known (e.g., Depts. of State, Justice, Defense, Homeland Security).

In this situation, a subjective measure of self-reported performance (e.g., high, medium, low) is probably the best practicable alternative. First, it is comparable across all tasks (from social security clerk in a local office to state department diplomat). Second, it should provide meaningful ordinal comparisons or ratings: because humans continually compare themselves to others, those who report that their own performance is high probably have higher levels of performance than those who report medium levels. Third, the behavior of the measure appears to have face validity. Even in the public

sector, there should be a correlation between objective pay and self-rated performance; poor performers are not promoted, on the average.

The (unadjusted) measure of performance in the survey is explicitly comparative. The survey asks respondents: "How would you rate your own performance in comparison to those in you immediate work unit?" The response options are: Above average (3), Average (2), or Below Average (1). In my study, I seek to compare high to low performers, not high relative performers. Hence I create a measure of performance that adjusts for how well the work unit performs. Fortunately, the survey also asks respondents to respond to the following statement: "My work unit produces high quality products and services," using a 5-point scale with an ordinal response continuum from strongly agree (5) to strongly disagree (1).[1]

The scale that I use to measure the respondent's own level of performance adds 1 point to the unit performance response if the respondent says that her performance is above the unit average. If the respondent reports that her performance is also the work unit average, the work unit response is the individual response. Finally, if the respondent reports that his performance is below the work unit average, I subtract one point from the work unit response. The result is a 7-point scale that ranges from 0 (meaning that the respondent's performance is below that of the performance of an underperforming work unit) to 6 (meaning that the respondent's performance is above that of the performance of a high-performing work unit). The observed scores range from 0 to 6, with a mean of 4.9 and a median of 5. Only 2% of the sample give themselves scores of 2 or below, while 70% give themselves scores of 5 or 6. Clearly this is a dependent variable with little

[1] This is an inversion of the original MSPB scale, which gave low numbers to positive outcomes.

variance ($s^2 = 0.9$). Nonetheless, the results below show that at least some of this variance is not random.

There is also support for the expectation that this measure of individual performance has a degree of face validity. First, the correlation between the log of actual pay and the adjusted measure of self-reported performance is positive, but low (r=.16). Second, respondents with higher levels of supervisory status also report higher levels of (adjusted) performance (r=.23). While these correlations are not "high", they are in the expected direction. Moreover, if I regress the adjusted measure of self-reported performance on both the log of actual pay and supervisory status, both variables significantly affect the performance measure; the beta weight for supervisory status is about twice that for salary (.19 versus .08).[2] Finally, a regression of self-reported performance on level of education reports a significant and positive coefficient for having a BA, having an MA, and having a PhD, relative to the reference group (high school or equivalent). Moreover, the coefficient for the PhD (.28) is twice that for the MA (.15); and the coefficient for the MA is larger than that for the BA (.11). These results comport with what one would expect if increasing education leads to higher real performance.[3]

Independent Variables

The main independent variables examine whether various aspects of performance measurement and pay for performance actually make employees perform better. One aspect of performance measurement is its specificity. When performance is hard to

[2] In every example, using the response for the work unit average (unadjusted for the individual relative to the work unit) gives a poorer fit. The implication is that the adjusted score is a better measure of individual performance than the unadjusted work-unit score. Moreover, in every example, using ordered logit gives a poorer fit than regression with robust standard errors.

[3] Table not shown. Ordinal logit produces the same results.

measure, using a detailed performance rating instrument may in fact be counterproductive if it leads to poorer performance, especially if it is a high-stake measure, closely linked to pay. On the other hand, in the context of a federal agency tasked with hard-to-measure, collectively produced goals, detail might add clarity to an otherwise vague and uncertain task, could improve performance, especially the connection between performance measure and pay is weak. Much of what Federal workers do is hard to measure: not only are the outputs and outcomes hard to measure, but many inputs are not observable, and, even if both inputs and outputs were observable, the production process is unknown. In this context, measuring performance (without basing pay on it) may improve (reported) performance.

I measure detail in the performance-rating instrument by summing the responses to two questions. The first item asks how many rating levels (from 2 to 6) are used in the respondent's performance appraisal system, and the second asks how objective the measures are, ranging from 1 to 5 (most objective). My expectation is that the number of rating levels is an indicator of how "closed" or "high-powered" the rating system is, and the "objectivity" of the system is an indicator of reduced uncertainty about the job (and possibly greater likelihood of goal displacement). I sum the responses to these to items to create an indicator of how clearly the respondent views his job (goal clarity). The resulting scale of how closed the rating system ranges from 2 to 10, with a mean of 6.1 and a median of 6.

Supervisors can act as go-betweens to help translate unclear or ambiguous policy directions to front-line employees, reducing agents' uncertainty about principals' preferences. In this way, they can contribute to the process of clarifying the nature of the

task; if effective, they help to identify what is to be measured or "counted" as performance. I measure the supervisors efforts by the response to the question: "My supervisor provides coaching, training opportunities, or other assistance to help me improve my skills and performance." The (observed) responses range from 1 (strongly disagree) to 5 (strongly agree).[4] The mean is 3.3, and the median is 3.

Measuring performance is not the same as basing pay on performance. I assess the degree to which employees report that their pay is linked to their performance using responses to two questions in the MSPB survey. The first question asks: "If your team performs well, how likely is it that you will receive a cash award or pay increase?" The responses range from 1 (very unlikely) to 5 (very likely). The second measure is the response to the question: "I understand how my pay relates to my job performance." Responses range from "Strongly agree" (5) to "Strongly disagree" (1).[5] The mean score for the first measure (likelihood of cash award) is 3.1 and the median is 3; the observed range is from 1 to 5, with a variance of 1.95. The observed range for the second measure is also 1 to 5, but the mean is 3.6 and the median is 4; it follows that the variance is smaller (slightly over 1.0). The two measures are positively but weakly correlated ($r = 0.28$).[6]

I measure the importance of intrinsic and extrinsic values by constructing two scales. The first scale (intrinsic values) is the sum of responses to 4 items that ask how important each one is in motivating the respondent to do a good job: agency mission; duty as a public employee; desire to help the work unit meet its goals; and personal pride

[4] This also is an inversion of the original MSPB scale, which gave low numbers to positive outcomes.
[5] These also are inversions of the original MSPB scale.
[6] I created a new variable by multiplying the likelihood of the link by the clarity with which the link is understood, and used it as a substitute for the two separate indicators in the regression results reported below. The results reported below remain unchanged by this modification.

or satisfaction with the work. The highest possible value is 20, and the lowest is 4. As a measure of scale reliability, the alpha = .70. As a measure of construct validity, indicative that the scale measures one underlying concept, the eigenvalue in a principal component analysis is 2.11 out of 4, which means that, together, the 4 items explain more than half of the total factor space. The scale mean = 18, s.d.= 1.8, with an observed range of 4-20. Most respondents like their work.

The second scale measures extrinsic value. That scale is the sum of responses to 4 similar items that ask about the importance of money and leisure: a cash award of $100, a cash award of $1000, desire for a good performance rating, and a time off reward of 8 hours. The highest possible value is 20, and the lowest is 4. With alpha = .75, the scale is reliable. The eigenvalue in a principal component analysis is 2.3 out of 4, meaning that the items explain nearly 60% of the factor space, suggesting that the 4 items are measuring the same underlying concept. The scale mean = 14, s.d. = 3.2, with an observed range of 4-20. Most respondents like their pay, but they like their job more than their pay.

I also expect that, for the average worker, having discretion improves performance. I measure job discretion by the sum of responses to two items. The first item is the degree of agreement (from strongly disagree to strongly agree, or 1-5 respectively) with the statement: "My opinions count at work." The second is the degree of agreement with the statement: "Creativity and innovation are rewarded." The possible range is from 2 to 10. The responses to these items are highly related: (r=.66). The mean is 6.75 (s.d.=2), and the observed range is from 2 to 10.[7]

[7] The scores for these items are also inversions of the original MSPB scale.

Finally, I expect that peer pressure and norms mean that workers in productive agencies will themselves be productive. I measure agency productivity as the sum of responses to two 5-point agree-disagree items, "My agency produces high quality products and services," and "My agency is successful in accomplishing its mission).[8] The observed range is from 2 to 10; the mean (and median) is 8 (s.d. = 1.6)

I expect the relation between performance measurement, pay for performance, motivation, discretion and performance to be affected by many other variables. Supervisory status clearly is likely to affect both factors. Relative to having no supervisory status (the reference group), I include an indicator variable for whether the respondent is a team leader (an informal supervisory role) (13%), a supervisor (30%), a manager (who oversees other supervisors) (19%), or an SES or equivalent (3%). There is also an indicator for whether the respondent works at headquarters (in DC) (25%), or in a regional or field office (the reference). I controlled for the years of federal service (mean=20.4, s.d.= 9.5), and for the type of the respondent's federal pay plan: GS is the reference; there are separate indicators for the wage system (4%), the SES 2.5%), and other types of pay plans (10%). There are also controls for: the respondent's salary (in thousands; mean = 87, s.d. – 41); whether the respondent is a dues-paying union member (11%), or in a position that is covered by a union agreement (15%); the reference is not being a union member, or not sure; for gender (66% male); for education, with separate indicators for associate degree (11%), college degree (38%), master's degree (22%), and PhD degree (9%), with high school or less as the reference. I also control for race/ethnicity with separate indicators for Asian (3%), Black/African American (12%), and white (75%), with Hispanic/Latino and other as the reference group. There is also a

[8] Ibid.

dummy for the respondent's agency. There are 24 agencies; the Dept. of Agriculture is the reference group. Table 1 reports the OLS regression results with robust standard errors.[9]

The Results

The regression examines the impact of standard economic instrumental aspects of job design (pay and pay satisfaction, pay for performance, detailed tasks, close supervision) with consumption aspects of job design (autonomy, intrinsic satisfaction with the job, behavior of others in your work unit), controlling for other characteristics of the respondent. Overall, the results show that money matters, but the consumptive aspects of the job matter more. Considering first the instrumental variables, an additional $10,000 in salary is associated with a higher self-reported performance score of 0.005 on the 0-6 scale; while this is significant statistically ($p < .013$) (driven by a large N=23,891), its substantive importance is underwhelming. The indicator of relative importance (beta) is also tiny, at .02[10]. The scale measuring satisfaction with pay is not significant at conventional levels, which are particularly appropriate with the N is so large ($p < .104$). Pay for performance is supposed to elicit higher levels of performance, but respondents who respondents who report that it is likely that better performance will be rewarded with higher pay do not report significantly higher performance ($p < .633$), and responds who report that the link between their performance and their pay is clear also do not report significantly higher performance ($p < .973$).

[9] With one exception, the results from an ordinal logistic regression with robust standard errors are identical to those reported in Table 1. In the ordinal logistic regression, race is significant and negative, while in the robust ordinary least squares regression race is not significant. I report the OLS results since the goodness of fit is twice that of the ordinal regression.

[10] The log of salary produced a similar result.

Measuring performance helps, independent of the link with pay: the more detailed the measure of performance, the higher the self-reported performance: for each additional point on the 2-10 measure of detail, performance increases by .02 on the 0-6 scale ($p < .000$). While this is statistically significant, the magnitude is not large, and nor is relative magnitude (beta = .04). Supervisors who clarify the nature of the job also raise self-reported performance. For each additional point on the 1-5 scale measuring the supervisor, performance increases by .05 points on the 0-6 scale ($p < .000$). Like the amount of detail in the measure of performance, this is statistically significant and positive, but the magnitude is not especially "large", and nor is relative magnitude (beta = .05).

Consider next the consumptive aspects of the job: liking the job, having discretion, and cooperating with others. Considered together, the sum of the beta weights for these three variables is .62, which means that a one standard deviation unit change in each of these variables accounts for a .62 standard deviation unit change in self-reported productivity. Recall that the mean of this 0-6 measure is 4.91, with a standard deviation of .93. Thus, the consumptive aspects of the job account for most of the variance. These three factors are not equally important.

Most important is the perceived productivity of the agency. A one point increase in the agency productivity scale (which ranges from 2 to 10) is associated with an increase of .2 in the average respondent's own performance ($p < .000$). In standardized units, the corresponding beta weight is .35. The result implies that respondents do not appear to free ride. It is likely that social pressure, or the ability to observe what others

are doing, accounts for individual behavior: when others work hard, those in the group do also.

Intrinsic values count also. A one point increase in the scale measuring intrinsic value of the job (ranging from 4-20) is associated with an increase of .1 in the average respondent's own performance ($p < .000$). In standardized units, the corresponding beta weight is .17. The beta weight for intrinsic values is many times that of the beta weight for extrinsic values (which was .01, and the variable was not significant). With respect to individual autonomy, a unit increase in the 2-10 discretion scale is associated with a .04 increase in the average respondent's own performance ($p < .000$). In standardized units, the corresponding beta weight is .09. Together with the result for agency productivity, the results are not consistent with a model of an intrinsically motivated employee who shirks (e.g., eats doughnuts) when given discretion.

Supervisors rate themselves as more productive (higher performance) than those who have no supervisory status. Correspondingly, education also appears to be associated with higher performance: compared to those with no post-secondary degree, those with a BA (but not an AA), MA, and PhD rate their performance as better.

It is important to note what variables are not significant; the large N makes the observation worth noting. Working in the field or headquarters has no relation to performance, and nor does the type of pay grade, or union membership, or gender. Compared to the reference group (Hispanics), race does not matter, with one disturbing exception: the average white respondent appears to rate his or her own performance higher.

A slightly different specification upholds the expectation that, with respect to performance, the consumptive aspects of the job may be more important than the instrumental aspects. Table 2 reports results from a model that examines the interaction between discretion and what employees value in their job. An economic theory that ignores heterogeneous preferences (i.e., employees place different values on the relative mix of pay and inherent value in a job) and assumes that effort is always costly would predict that discretion leads to shirking. Accounting for heterogeneous preferences implies that employees who value their job intrinsically (because it is "fun," or because they believe in its mission) and who also have discretion are likely to be more productive. By contrast, employees who work for pay could use discretion to be less productive, or to be no more productive than they would be otherwise. The implication is that the impact of discretion will depend on values. I create these interactive terms by multiplying the discretion scale by the intrinsic satisfaction scale, and then by multiplying the same discretion scale by the extrinsic satisfaction scale. The result is a discretion-intrinsic satisfaction scale that ranges from 8 to 200, with a mean of 123, and a discretion-extrinsic satisfaction scale with the same range and a mean of 96.

The results in Table 2 lend some support to the expectation of statistical interaction. The model fit (based on the F-test and R-squared) is only slightly poorer than that reported in Table 1. Working in a productive agency continues to promote individual productivity, as does having a clear measure of performance and a supervisor who helps clarify the measure. With respect to the interaction terms, as the level of discretion increases, employees who value their jobs intrinsically are more productive (by .04 for each 10 points on the interactive scale, $p < .000$). Further, as the level of

discretion increases, employees who value their jobs extrinsically are *less* productive (by .01 for each 10 points on the interactive scale, $p < .001$). As in Table 1, the clarity of the link between pay and performance continues to have no impact on self-reported performance, but, unlike the results in Table 1, the closeness of the link appears actually to reduce performance in this model estimate.

With some exceptions, the sign and significance of the control variables in Table 2 remain unchanged from those reported in Table 1. Self-reported performance continues to rise (generally) with the level of supervisory status, and with salary, but not with education. Only PhDs reported significantly higher performance, compared to statistically comparable employees with only a high school or equivalent education. In the model in Table 2, none of the race/ethnicity indicators is significant, but males report lower performance levels.[11]

Conclusion

PFP is supposed to improve employee (and agency) performance because it changes the basis for salary increases, so that the employee's performance affects the employee's pay increase. PFP is based on the assumption that employees respond to more pay by working "harder" and "better" for the employer who offers the extra pay. That employees have different preferences, are motivated to work not only because of money but also because of consumptive values (it is "fun;" it corresponds to values), and also by the social context (the "team") in not included in the unvarnished economic model. The evidence, however, suggests that, while money matters, consumptive values matter also. For example, with respect to U.S. federal government employees, previous

[11] (Agency) random effects estimates of the results in Tables 1 and 2 are essentially unchanged.

research reveals that pay is not necessarily important for retention; there is evidence that it has no significant impact on the probability of resigning and the probability of retiring (Langbein, 2009a). Second, liking the job is more important than more pay in retaining federal employees (op. cit.). PFP has numerous other problems, particularly when performance is hard to measure and the production functions (especially in the presence of team work) are unknown (Langbein, 2008a).

This study has similar findings. Controlling for numerous individual level characteristics, while this study shows that instrumental variables are associated with performance (salary, clear measure of performance, and supervisors who also help to clarify the nature of the job), linking pay to performance has no positive impact on performance, and nor does subjective satisfaction with pay. Further, the results show that consumptive aspects of the job are more important. Liking the job, having discretion, and working with others who are productive are all related to higher levels of individual performance. In relative terms, the beta weights for the consumptive aspects of the job sum to .62, while the beta weights for the (significant) instrumental aspects of the job sum to .11. (The comparable beta weights from Table 2 sum to .57 and .01 respectively.)

These findings imply that, particularly in the public sector, using money as a reward for performance may be less effective than creating a good job. Money matters, but so does the nature of the employee-job match. In this study, the correlates of a good job include discretion, a clear idea of what the employed is supposed to do, a cooperative workgroup, and a clear link between the employee's work and the agency's mission, but with an open performance assessment contract, allowing employees to determine how to

get the job done. These variables are more highly correlated with job performance than pay satisfaction, and support other evidence of the existence of public service motivation in the public sector. Even in the private sector, pay for performance is not as universal or optimal as many advocates for its use in the public sector seem to believe (Langbein, 2008a). Labor market sorting based on heterogeneous preferences means that the public sector attracts (or retains) employees with "public" motivations. Thus pay is likely to be even less effective in the public sector than it is in the private sector as a reward for good performance, especially when performance is hard to measure and is accomplished by teams. It probably should be even less widely used in the public than in the private sector.

The question remains: Why is PFP such a popular remedy for perceived poor performance in the public sector? Echoing Lewis (2008), I expect that it is most likely to be implemented in agencies where there is the greatest mismatch between the preferences of political principals and the professional norms of government employees. PFP is likely to attract employees who prefer to be rewarded by pay rather than by the job itself, and reduces the loss of control that political principals have over employee-agents. Patronage and political control rather than professional norms are likely to dominate PFP systems. In a time when public problems are particularly complex, the prospect of a return to patronage is indeed sobering.

Table 1: Regression of individual performance on instrumental and consumption variables, with statistical controls (MSPB, 2005)

my own performance	Coef.	Robust Std. Err.	t	P>|t|	Beta
my agency's performance	.207	.005	42.23	0.000	.351
discretion	.043	.004	11.43	0.000	.095
intrinsic satisfaction	.092	.004	25.90	0.000	.175
extrinsic satisfaction	.003	.002	1.63	0.104	.010
level of detail in performance measure	.018	.003	5.60	0.000	.036
closeness of pay-performance link	-.002	.004	-0.48	0.633	-.003
clarity of pay-performance link	-.000	.006	-0.03	0.973	-.000
coaching on performance measure	.046	.006	8.37	0.000	.057
team leader*	.119	.017	7.04	0.000	.043
supervisor*	.225	.014	15.55	0.000	.110
manager*	.230	.017	13.73	0.000	.096
executive*	.148	.057	2.59	0.010	.027
headquarters =1 (vs. field)	.015	.013	1.17	0.242	.007
years of federal service	.001	.0006	1.76	0.078	.010
WageSystem**	.012	.026	0.48	0.633	.003
SES**	.070	.058	1.20	0.229	.012
Other**	-.016	.019	-0.85	0.396	-.005
Salary (in 1000s)	.0005	.0002	2.49	0.013	.020
Unionyes***	-.011	.019	-0.57	0.571	-.004
Unioncover***	.024	.016	1.50	0.133	.009
Male	-.016	.011	-1.47	0.141	-.008
AA****	.004	.029	0.21	0.833	.001
BA****	.049	.015	3.25	0.001	.025
MA****	.050	.017	2.90	0.004	.022
PhD****	.150	.023	6.62	0.000	.046
Asian*****	.023	.031	0.73	0.464	.004
Black*****	-.028	.022	-1.25	0.213	-.010
White*****	.054	.018	2.98	0.003	.024
AirForce******	.045	.037	1.20	0.231	.007
Army	.046	.031	1.49	0.136	.009
Commerce	.094	.025	3.78	0.000	.029
Defense	.094	.024	3.99	0.000	.029
Education	.157	.043	3.62	0.000	.022

EPA	.155	.039	3.96	0.000	.023
Energy	.080	.046	1.72	0.085	.011
FDIC	.217	.042	5.22	0.000	.032
GSA	.052	.037	1.40	0.162	.009
HHS	.085	.028	3.08	0.002	.022
DHS	.145	.026	5.59	0.000	.043
HUD	.123	.050	2.45	0.014	.014
Interior	.069	.027	2.58	0.010	.019
Justice	.058	.026	2.25	0.025	.016
Labor	.029	.040	0.73	0.465	.004
NASA	.113	.038	2.99	0.003	.018
Navy	.073	.032	2.30	0.022	.015
OPM	.163	.041	4.00	0.000	.022
SSA	-.028	.040	-0.70	0.484	-.004
State	.094	.051	1.86	0.063	.011
Transportation	.107	.033	3.27	0.001	.021
Treasury	.163	.028	5.83	0.000	.040
Veterans Affairs	.057	.028	1.99	0.047	.014
cons	.621	.071	8.71	0.000	.

(*Reference supervisor status = Non-supervisor)
(**Reference type of pay plan = GS or similar to GS)
(***Reference union status = none)
(****Reference education = High school, GED)
(*****Reference race/ethnicity = Hispanic/other)
(******Reference agency = Agriculture)

F(51, 23,839) = 175.08
Prob > F = 0.0000
R-squared = 0.33
Number of obs = 23,891

Table 2: Regression of individual performance on instrumental and consumption variables, with statistical controls (interaction between discretion and type of job statisfaction) (MSPB, 2005)

my own performance	Coef.	Robust Std. Err.	t	P>|t|	Beta
my agency's performance	.217	.005	44.31	0.000	.369
discretion*intrinsic satisfaction	.004	.0002	18.64	0.000	.202
discretion*extrinsic satisfaction	-.00007	.0002	-3.21	0.001	-.029
level of detail in performance measure	.017	.003	5.22	0.000	.033
closeness of pay-performance link	-.009	.004	-2.14	0.032	-.014
clarity of pay-performance link	.006	.006	0.95	0.343	.006
coaching on performance measure	.035	.006	6.29	0.000	.044
team leader*	.128	.017	7.56	0.000	.046
supervisor*	.239	.015	16.35	0.000	.117
manager*	.245	.017	14.50	0.000	.102
executive*	.170	.057	2.97	0.003	.031
headquarters =1 (vs. field)	.012	.013	0.95	0.343	.006
years of federal service	.0006	.0006	0.95	0.340	.006
WageSystem**	.009	.026	0.34	0.737	.002
SES**	.069	.058	1.19	0.233	.012
Other**	-.022	.019	-1.14	0.252	-.007
Salary (in 1000s)	.0004	.0002	2.32	0.020	.019
Unionyes***	.002	.020	0.10	0.924	.001
Unioncover***	.021	.016	1.29	0.196	.008
Male	-.039	.011	-3.51	0.000	-.020
AA****	.003	.020	0.13	0.896	.001
BA****	.027	.015	1.78	0.075	.014
MA****	.028	.018	1.60	0.110	.012
PhD****	.127	.023	5.52	0.000	.039
Asian*****	.017	.032	0.52	0.600	.003
Black*****	-.030	.022	-1.33	0.183	-.010
White*****	.029	.018	1.57	0.116	.013
AirForce******	.048	.038	1.27	0.206	.007
Army	.042	.031	1.36	0.175	.008
Commerce	.077	.025	3.06	0.002	.024

Defense	.086	.024	3.58	0.000	.027
Education	.157	.044	3.59	0.000	.022
EPA	.149	.040	3.76	0.000	.022
Energy	.093	.047	1.99	0.047	.013
FDIC	.199	.042	4.72	0.000	.029
GSA	.054	.037	1.46	0.144	.009
HHS	.088	.028	3.13	0.002	.022
DHS	.158	.026	6.03	0.000	.047
HUD	.139	.050	2.75	0.006	.016
Interior	.063	.027	2.31	0.021	.017
Justice	.057	.026	2.19	0.028	.016
Labor	.034	.040	0.84	0.403	.005
NASA	.092	.038	2.45	0.014	.015
Navy	.065	.032	2.03	0.042	.013
OPM	.163	.041	3.92	0.000	.022
SSA	-.035	.041	-0.87	0.385	-.005
State	.068	.051	1.32	0.188	.008
Transportation	.092	.033	2.77	0.006	.018
Treasury	.138	.028	4.89	0.000	.034
Veterans Affairs	.071	.029	2.45	0.0 14	.017
cons	.621	.071	8.71	0.000	.

(*Reference supervisor status = Non-supervisor)
(**Reference type of pay plan = GS or similar to GS)
(***Reference union status = none)
(****Reference education = High school, GED)
(*****Reference race/ethnicity = Hispanic/other)
(******Reference agency = Agriculture)

F(50, 23840) = 164.76
Prob > F = 0.0000
R-squared = 0.32
Number of obs = 23,891

References

George A. Akerlof and Rachel E. Kranton (2008). "Identity, Supervision, and Work Groups," American Economic Review, May 2008 98 (2); 212-217.

Association of Government Accountants (2009). *Performance Based Management.* Alexandria, Va: AGA CPAG Research Series, Report No. 20 (March).

Pierre Azoulay (2004). "Capturing Knowledge within and across Firm Boundaries: Evidence from Clinical Development" American Economic Review, Vol. 94, No. 5, (Dec.), pp. 1591-1612.

George P. Baker and Thomas N. Hubbard (2003). "Make versus Buy in Trucking: Asset Ownership, Job Design, and Information", American Economic Review, Vol. 93, No. 3, (June), pp. 551-572).

Roland Benabou (2003). "Instrinsic and Extrinsic Motivation" Review of Economic Studies Volume 70 Issue 3, July, Pages 489 – 520.

Roland Bénabou and Jean Tirole (2003). "Intrinsic and Extrinsic Motivation" The Review of Economic Studies, Vol. 70, No. 3, (Jul., 2003), pp. 489-520

Timothy Besley and Maitreesh Ghattak (2005). "Competition and Incentives with Motivated Agents" American Economic Review June 2005 95(3): 616-636

John Bohte and Kenneth J. Meier (2000) "Goal Displacement: Assessing the Motivation for Organizational Cheating," Public Administration Review 60 (March/April 2000), 173-182).

John Brehm and Scott Gates (2008). Teaching, Tasks and Trust: Functions of the Public Executive. New York: Russell Sage Foundation.

Jongmin Chen (2000). “Promises, Trust and Contracts.” Journal of Law, Economics and Organization 16(1) April: 209-232.

Chen, Yan and Sherry Xin Li (2009). “Group Identity and Social Preferences”. *American Economic Review* 99 (1) March: 431-457.

Kevin A. Clarke and David M. Primo (2007). “Modernizing Political Science: A Model-Based Approach” Perspectives on Politics December 2007 | Vol. 5/No. 4: 741-753)

Crewson, P.E. (1997). “Public-Service Motivation: Building Empirical Evidence of Incidence and Effect. Journal of Public Administration Research and Theory, 7, 499-518.

Edward L. Deci and Richard M. Ryan (2000). “The "What" and "Why" of Goal Pursuits: Human Needs and the Self-Determination of Behavior” Psychological Inquiry, Vol. 11, No. 4. (2000), pp. 227-268.

Thomas S. Dee and Benjamin J. Keys (2004). "Does Merit Pay Reward Good Teachers: Evidence from a Randomized Experiment." Journal of Public Policy Analysis and Management 23 (3): 471-488.

Anthony Downs (1967). Inside Bureaucracy - 1967 - Little, Brown.

Tore Ellingsen and Magnus Johannesson “Paying Respect” Journal of Econ Perspectives 21 4 Fall 2007 133-149

Ernest Fehr, Karla Hoff, Mayuresh Kshetramade (2008). “Spite and Development” AER May 2008, 98 (2): 494-499

Bruno S. Frey (1993). “Does Monitoring Increase Work Effort? The Rivalry with Trust and Loyalty,” Economic Inquiry 31 (Oct.): 663-670.

Bruno S. Frey (1997). Not Just for the Money: An Economic Theory of Personal Motivation. Cheltenham, U.K.: Edward Elgar.

Bruno S. Frey and Felix Oberholzer-Gee (1997). "The Cost of Price Incentives: An Empirical Analysis of Motivation Crowding-out" American Economic Review 87(4): 746-755.

William Fuchs (2007). "Contracting with Repeated Moral hazard and Private Evaluations" American Economic Review 97 (4), Sept 2007: 1432-1448

Sean Gailmard and John W. Patty (2007). "Slackers and Zealots: Civil Service, Policy Discretion, and Bureaucratic Expertise", AJPS 51 4 Oct. 2007 873-889.

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

Richard Hartwig (2004). "A Tiny Ring of Power: The Department Chair and Golden Rule Management." Journal of Public Affairs Education 10 (1): 31-42.

Bengt Holmstrom (1982). "Moral Hazard in Teams" Bell Journal of Economics 13(2) Autumn: 324-340.

Heap, Shaun P. Hargreaves and Daniel John Zizzo (2009). "The Value of Groups." *American Economic Review* 99(1) March: 295-323.

Estelle James (1983). "How Nonprofits Grow: A Model" Journal of Policy Analysis and Management, Vol. 2, No. 3 (Spring, 1983), pp. 350-365.

Chan Su Jung and Hal G. Rainey (2009). "Organizations Goal Characteristics and Public Duty Motivation in U.S. Federal Agencies." Presented at International Public Service Motivation Research Conference, Bloomington, Ind., June 7-9, 2009.

Kahneman, Daniel and Amos Tversky**.** 1979. "Prospect Theory: An Analysis of Decision under Risk." *Econometrica.* 47:2, pp. 263–91.

Krause, G.A., Lewis, D. C., and Douglas, J.W. (2006). "Political Appointments, Civil Service Systems, and Bureaucratic Competence: Organizational Balancing and Executive Branch Revenue Forecasts in the American States." American Journal of Political Science 50 (3), July: 770-787.

D. M. Kreps (1997). "Intrinsic Motivation and Extrinsic Incentives." American Economic Review 87(2): 359-364

Laura Langbein (2000). "Ownership, Empowerment and Productivity: Some Empirical Evidence on The Causes and Consequences of Employee Discretion," Journal of Policy Analysis and Management, 19 (3), 2000: 427-449.

Laura Langbein (2008) "Management by Results: Student Evaluation of Faculty Teaching and the Mis-measurement of Performance." Economics of Education Review, 27 (4), August 2008: 417-428

Laura Langbein (2008a). "Economics, Public Service Motivation and Pay for Performance: Complements or Substitutes?" Paper presented at the Midwest Political Science Association Annual Meeting, Chicago.

Laura Langbein (2009). "Controlling Federal Agencies: The Impact of External Controls on Worker Discretion and Productivity." International Public Management Journal 12 (1) 2009: 82-115.

Laura Langbein (2009a). "The Impact of Love and Money on Quitting in the Federal Government: Implications for Pay for Performance." Presented at the International Public Service Management Conference, Bloomington Indiana, June 5-7.

Edward P. Lazear (1989). “Pay Equality and Industrial Politics,” Journal of Political Economy 97 (June): 561-580).

Edward P. Lazear (2000). “The Power of Incentives,” American Economic Review (Papers and Proceedings) 90 (2), May: 410-414.

Edward P. Lazear and Katherine L. Shaw (2007) “Personnel Economics: The Economist’s View of Human Resources” J. of Econ Perspectives 21 (4), Fall: 91-114.

Julian Le Grand (2009). “Knights and Knaves Return: Motivation and the Delivery of Public Services.” Dept. of Social Policy, London School of Economics. Unpublished ms. May.

Lewis, D. E. (2008). The Politics of Presidential Appointments: Political Control and Bureaucratic Performance. Princeton, N.J: Princeton University Press.

Dave Mader, Jeff Myers, and Steven Kelman (2009). What it Takes to Change Government. McLean, Va.: Booz Allen Hamilton.

Mas, Alexandre and Enrico Moretti (2009). “Peers at Work.” *American Economic Review* 99 (1), March: 112-145.

Peter J. May and Soren C. Winter (2007). “Politicians, Managers, and Street-Level Bureaucrats: Influences on Policy Implementation.” J. of Public Administration Theory and Research 19: 453-476.

Gary Miller and Andrew Whitford (2002). "Trust and Incentives in Principal-Agent Negotiations: The 'Insurance-Incentive' Tradeoff." Journal of Theoretical Politics 14 (April): 231-267.

Miller, G.J. & Whitford, A.B. (2007). "The Principal's Moral Hazard: Constraints on the Use of Incentives in Hierarchy". Journal of Public Administration and Research Theory, 17, 213-233.

Mintzberg, H. (1996). Managing Government, Governing Management. *Harvard Business Review, 3,* 75-83

Moynihan, Donald P (2007). "The Normative Model in Decline? Public Service Motivation in the Age of Governance" Working Paper Series La Follette School Working Paper No. 2007-021 http://www.lafollette.wisc.edu/publications/workingpapers

D. Nagin, J. Rebitzer, S. Sanders, and L. Taylor (2002). "Monitoring, Motivation and Management: The Determinants of Opportunistic Behavior in a Field Experiment." NBER Working Paper 8811 (March).

William A. Niskanen (1975). "Bureaucrats and Politicians" Journal of Law and Economics, Vol. 18, No. 3, (Dec.), pp. 617-643.

Oh, Seong Soo and Gregory B. Lewis (2009). "Can Performance Appraisal Systems Inspire Intrinsically Motivated Employees?" *Review of Public Personnel Administration.* Published online Feb. 13 as doi:10.1177/0734371X09331616.

Olson, Mancur (1982). The Rise and Decline of Nations. New Haven: Yale University Press.

Perry, J. (1986). Merit Pay in the Public Sector: The Case for a Failure of Theory. *Review of Public Personnel Administration, 7,* 57-69.

Perry, J. (2007). Democracy and the New Public Service. *American Review of Public Administration, 37,* 3-16.

Canice Prendergast (2008) "Intrinsic Motivations and Incentives" American Economic Review: Papers & Proceedings 2008, 98:2, 201–205.

Graham Prentice, Simon Burgess, and Carol Propper (2007). "Performance pay in the public sector: A review of the issues and evidence." Office of Manpower Economics, UK (November)

Howard Risher (2007). ICMA IQ Report, Vol. 39, No. 6. ICMA, Washington, DC, Nov.

Rojstaczer, S. (2003). Grade inflation at American colleges and universities. http://www.gradeinflation.com/)

Anja Schöttner (2008). "Relational Contracts, Multitasking and Job Design," J. of Law and Econ 24 (1), May 2008: 138-162.

Carl Shapiro and Joseph E. Stiglitz (1984). "Equilibrium Unemployment as a Worker Discipline Device." American Economic Review 74 (3): 433-444.

David A. Siegel (2009). "Social Networks and Collective Action." American Journal of Political Science 53 (1): 122-138.

Otto H. Swank and Bauke Visser (2006). "Motivating Through Delegating Tasks or Giving Attention." J. of Law, Economics and Organization 23(3): 731-743.

Jeannette Tayor and Ranald Taylor (2009). "Do Governments Pay Efficiency Wages? Evidence fro a Selection of Countries." Paper presented at International Public Service Motivation Research Conference, Bloomington, Ind. June 7-9.

Benno Torgler, Sascha Schmidt (2007). "What shapes player performance in soccer? Empirical findings from a panel analysis" Applied Economics, Volume 39, Number 18, October 2007, pp. 2355-2369

Van den Steen, Eric (2009). "Authority versus Persuasion." American Economic Review: Papers and Proceedings 99(2): 448-453.

van Rijn, P. (2005). "Participating Agencies, Methods of Survey Distribution, and Response Rates." (retrieved from MSPB, 10/12/05).