Performance-related Pay in the Public Sector

A Review of Theory and Evidence

Zahid Hasnain

Nick Manning

Jan Henryk Pierskalla

The World Bank
East Asia and Pacific Region
Poverty Reduction & Economic Management Sector Department
Poverty Reduction and Economic Management Network
Public Sector Governance
April 2012

Abstract

The objective of this paper is to provide a review of the theoretical and, in particular, empirical literature on performance-related pay in the public sector spanning the fields of public administration, psychology, economics, education, and health with the aim of distilling useful lessons for policy-makers in developing countries. This study to our knowledge is the first that aims to disaggregate the available evidence by: (i) the quality of the empirical study; (ii) the different public sector contexts, in particular the different types of public sector jobs; and (iii) geographical context (developing country or OECD settings). The paper's main findings, based on a comprehensive review of 110 studies of public sector and relevant private sector jobs are as follows. First, we find that overall a majority (65 of 110) of studies find a positive effect of performance-related pay, with higher quality empirical studies (68 of the 110) generally more positive in their findings (46 of the 68). These show that explicit performance standards linked to some form of bonus pay can improve, at times dramatically, desired service outcomes. Second, however, these more rigorous studies are overwhelmingly for jobs where the outputs or

outcomes are more readily observable, such as teaching, health care, and revenue collection (66 of the 68). There is insufficient evidence, positive or negative, of the effect of performance-related pay in organizational contexts that are similar to that of the core civil service, characterized by task complexity and the difficulty of measuring outcomes, to reach a generalized conclusion concerning such reforms. Third, while some of these studies have shown that performance-related pay can work even in the most dysfunctional bureaucracies in developing countries, there are too few cases to draw firm conclusions. Fourth, several observational studies identify problems with unintended consequences and gaming of the incentive scheme, although it is unclear whether the gaming results in an overall decline in productivity compared to the counterfactual. Finally, few studies follow up performance-related pay effects over a long period of time, leaving the possibility that the positive findings may be due to Hawthorne Effects, and that gaming behavior may increase over time as employees become more familiar with the scheme and learn to manipulate it.

This paper is a product of the Poverty Reduction & Economic Management Sector Department, East Asia and Pacific Region; and the Public Sector Governance and Poverty Reduction and Economic Management Network. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at <http://econ.worldbank.org>. The author may be contacted at zhasnain@worldbank.org and nmanning@worldbank.org

The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

Performance-related Pay in the Public Sector: A Review of Theory and Evidence

Zahid Hasnain¹, Nick Manning², and Jan Henryk Pierskalla³

JEL Codes: H11, H83, I18, I28

Acknowledgements: This paper has benefitted greatly from comments by Mike Stevens, Willy McCourt, Mariano Lafuente, Gary Reid, and Svetlana Proskurovska.

¹ World Bank (zhasnain@worldbank.org)

² World Bank (nmanning@worldbank.org)

³ Duke University (jhp5@duke.edu)

Contents

1. Introduction	1
2. Theoretical debates	4
Expectancy and reinforcement theory	4
Incentive and principal-agent theory	5
Behavioral economics —intrinsic versus extrinsic motivation	7
3. Organizing the empirical evidence: “Craft” and “Coping” jobs.....	9
Methodological approaches.....	11
4. The empirical literature reviewed.....	13
Observational studies	13
Public sector coping jobs	13
Public sector craft jobs.....	14
<i>Tax administration, job placement</i>	14
<i>Teaching</i>	16
<i>Health care jobs</i>	17
Private sector: Craft or coping jobs	19
Experimental studies	21
Meta-studies.....	21
Laboratory experiments	22
Public sector craft jobs.....	23
<i>Tax administration, job placement</i>	23
<i>Teaching</i>	24
<i>Health sector</i>	25
Private sector: Craft or coping jobs	26
5. Assessing the evidence	26
6. Summary.....	28
Appendix A: List of empirical PRP studies reviewed	33
Appendix B: List of High Quality Studies of Craft and Coping Jobs	37
References	40

Figures

Figure 1: Aggregate findings on performance-related pay.....	27
Figure 2: Findings by internal and external validity.....	27
Figure 3: Findings by job type.....	28
Figure 4: Findings for craft and coping tasks by research quality and country context.....	29

Tables

Table 1: James Q. Wilson’s classification of job type	11
Table 2: Studies by country environment, methodology, and job type.....	13
Table 3: Findings of high quality craft and coping studies by sector and country context.....	29

Boxes

Box 1: The elements of pay flexibility.....	3
---	---

1. Introduction

Performance-related pay (PRP) has been introduced in many countries as a possible tool for improving the productivity and accountability of the public sector. Over the past fifteen years, a majority of OECD countries have implemented PRP in the central administration (core civil service), in specialized entities such as revenue administration, and for key service delivery staff such as teachers and medical personnel. Middle income countries, and to some extent low income countries, perhaps drawing on the OECD example, have also experimented with PRP in an attempt to inject performance-orientation in otherwise dysfunctional bureaucracies. A vast theoretical and empirical literature has analyzed various dimensions of PRP, and there is now a small but growing body of robust evidence on the impact of PRP that is shedding new light on what is achievable and under what specific conditions.

The objective of this paper is to provide a review of the theoretical and in particular empirical literature on PRP in the public sector spanning the fields of public administration, psychology, economics, education, and health with the aim of distilling useful lessons for policy-makers in developing countries. This is by no means the first comprehensive review of this literature; but it is, to our knowledge, the first that aims to disaggregate the available evidence by (i) the quality of the empirical study; (ii) the different public sector contexts, in particular the different types of public sector jobs; and (iii) geographical context (developing country or OECD settings). The intention in so doing is to ensure that the findings from the empirical literature are appropriately nuanced.

PRP is a compensation arrangement in which the final salary of an employee is a function of some form of measured “performance”, where how performance is measured, who measures it, and how it is linked to salary can all vary considerably and are key aspects of the design of the scheme. Performance can be based on qualitative assessments or quantitative measures of inputs (effort allocation, attendance, voluntary contributions at the workplace and skills acquisition), outputs (completion of pre-agreed tasks or number of clients/cases served), or outcomes (student test scores, official client evaluations, service utilization rates or revenue creation). Salary can either be wholly a function of performance, for example as piece-rate pay in a manufacturing setting or commission-based salary in a sales environment, or a combination of base pay and one-off bonuses or merit increases of base pay. Bonuses and merit increases can be awarded on an individual, small team or larger departmental basis. Evaluations can be implemented by direct supervisors, human resource specialists, peer panels or outside agencies. Once performance has been measured, it has to be evaluated against a performance standard. This standard can be based on individually pre-agreed goals, absolute performance against minimum or scaled standards, relative improvements against past performance, rank-order of performance in a tournament evaluation or relative performance measured against co-workers, other teams of co-workers, other schools or agencies nationally or regionally.

PRP in the form of bonuses or merit increases to basic pay has been used more frequently in the OECD in recent years. According to one estimate, approximately two-thirds of OECD countries have introduced PRP in some form or the other (OECD 2005). The United Kingdom, Switzerland and the Czech Republic apply PRP more extensively than countries such as New Zealand, Austria and the Netherlands. In Finland, for example, the proportion of basic salary that PRP can represent can amount to over 40% of the total. In the US, while PRP is of limited use in the core public administration, it is at the center of efforts to improve teacher accountability over the past decade in the context of the No Child Left Behind Act. The literature suggests that there are similar movements underway in middle income countries and, perhaps more sporadically, in lower income countries where PRP is more often referred to in the health and education sectors than in the core administration.

Drawing on contract theory and the problems of moral hazard, a significant chunk of the academic literature has examined the impact of PRP on increasing effort and reducing shirking. In situations where effort is unobservable, fixed pay contracts provide little ability for employers to influence employee effort. This is especially likely to be the case in traditional civil service jobs characterized by uniform pay for jobs in similar grades, pay increases based largely on seniority, and negligible probability of termination. Contracts that tie observable outputs, which are correlated with unobservable effort, to desired pay incentives can mitigate such problems. Some of the literature has attempted to move beyond effort and to introduce the concept of “engagement” as a measure of an employee’s emotional and intellectual commitment to their employing organization and its success, encompassing both commitment (“I like working here”) and organizational citizenship (“I am prepared to go the extra mile”), with the question then being what impact can PRP have on staff engagement. Contract theory also suggests that PRP can help address the problem of adverse selection by encouraging high ability individuals who will do better under a performance pay scheme to join the agency and similarly discouraging low ability individuals (the “sorting” effect).

Critics counter that PRP does not work when tasks are multi-dimensional as it results in “gaming” behavior whereby effort is only allocated towards what is observed and measured which may not improve overall outcomes. A huge literature has examined the challenges in effectively designing these schemes, particularly in public sector contexts where tasks are complex and outcomes cannot be easily measured. The psychological and behavioral economics literature in addition has argued that individuals are also motivated by intrinsic concerns about the inherent social value of the job — particularly in the public sector — and that PRP which explicitly focuses on extrinsic benefits might crowd-out intrinsic motivation thereby reducing worker productivity.

In reviewing the empirical evidence, 153 studies in total, the paper is explicit in focusing largely on job types that are relevant to the public sector and characterized by task complexity and unobservability of effort thereby excluding studies that examine PRP for simple manufacturing jobs or other repeatable tasks. These jobs, borrowing James Q. Wilson’s typology, are classified as “craft” and “coping” jobs. This narrows the list to 110 studies. Reviewing these, the paper draws the following conclusions. First, a majority (93 of the 153) of all studies find some positive effect of PRP, and a majority (65 of 110) of studies of craft and coping jobs show positive findings. Limiting the analysis to craft and coping also shows while early case studies found largely inconclusive evidence on the impact of PRP on staff morale, effort, and productivity, work over the last 10 to 15 years that has been based on more systematic observational studies and experimental evaluations in the laboratory and in the field have generally found that explicit performance standards linked to some form of bonus pay can improve, at times dramatically, desired service outcomes. 46 of the 68 high quality studies of craft and coping jobs showed a positive effect of PRP.

Second, however, these more rigorous studies are usually for jobs where the outputs or outcomes are more readily observable, such as revenue collection, teaching and health care. There is simply not enough robust evidence, positive or negative, of the effect of PRP in organizational contexts that are similar to that of the core civil service, characterized by task complexity and the difficulty of measuring outcomes to reach a generalized conclusion for reform. To be specific, of the 68 high quality studies, only 2 were for contexts similar to core civil service jobs. Third, while some of these studies have shown that PRP can work even in the most dysfunctional bureaucracies, there is limited evidence in developing country contexts (only 10 high quality studies), with considerable discretionary as opposed to rule-based behavior and with significant politicization that can negatively affect the overall credibility and legitimacy of the incentive scheme. Fourth, several observational studies identify problems with unintended consequences and gaming of the incentive scheme, although it is unclear whether the gaming

results in an overall decline in productivity compared to the counterfactual. Finally, few studies follow up PRP effects over a long period of time, leaving the possibility that the positive findings may be due to Hawthorne Effects⁴, and that gaming behavior may increase over time as employees become more familiar with the PRP scheme and learn to manipulate it.

This literature review focuses almost entirely on the individual incentive effects of PRP as this is what the bulk of the literature has emphasized. It should be noted that there are, at least in the policy literature, potential agency-level and public sector wide effects of PRP that have to date been underexplored in the academic literature. PRP can provide a mechanism for conveying to staff the increasing expectations of agency performance, in effect changing the agency-level culture of performance (“this is how hard we work around here”) through the multitude of individual performance appraisal discussions (Marsden 2004). PRP can also have wider public sector impacts on the fiscal sustainability of the wage bill, public sector pay competitiveness, and social objectives, such as gender equity. PRP may contribute to cost containment by limiting pay increases to less costly performance bonuses (Marsden and French 1998). Some studies have also suggested that PRP may have a gender dimension as men are more likely to see the arrangement as fair and reasonable. Finally, PRP in the OECD countries been introduced in the context of other changes in pay policy, in particular moves towards differentiated pay arrangements whereby pay for similar jobs varies across government agencies and occupational groups, and the delegation of pay setting authority from central human resource ministries to line ministries and agencies (see Box 1). What remains unexplored are the potential linkages between these reforms and whether or not PRP caused or was the effect of these other changes in pay policy.

Box 1: The elements of pay flexibility

OECD countries have move towards flexible pay arrangements in the public sector which is in essence some combination of three key features, or design elements:

- **Performance-related pay:** Enabling pay to differ for civil servants doing the same job by linking a portion of the civil servants’ pay to individual or group effort or performance with performance targets set either at the individual level and/or the group level.
- **Differentiation:** Pay differences within and across government ministries, departments, and agencies for civil servants doing the same job - based on the need to attract and retain qualified staff for those jobs, or to persuade staff to accept new working arrangements. The need for special incentives for attraction or retention can be based on labor market conditions and/or cost of living in the localities where agencies operate or as a function of the specific skills that the agency competes for in the labor market, based for example on job evaluations and labor market surveys. Differentiation can affect: (i) all staff paid by the entity, with the result that there are agency-specific pay scales; or (ii) particular occupational groups or cadres; (iii) specific individuals, with scarce but vital skills; or (iv) specific locations entailing particular hardship.
- **Delegation:** Transferring authority over pay setting and human resource management from a central civil service agency to ministries, agencies, or departments. The transfer may entail transferring from the central to the entity level (or a mixture): pay negotiations (if any); setting the overall wage bill; and design of pay scales.

While the three dimensions of pay individualization, differentiation and delegation describe analytically distinct elements of pay flexibility, reform examples in the real world often combine all elements. There are three reasons why components of pay flexibility tend to go together: a shared set of assumptions that underpin them; a complementarity in the objectives that they are aiming for; and sequencing. First, many of these reforms took place in the context of New Public Management that emphasized giving managers the authority to manage in exchange for tighter accountability standards. This move reflected wider changes in the economy where there has been a significant move away from industry and sector-wide negotiating arrangements with significant government involvement towards more local and agency-specific arrangements.

⁴ Hawthorne Effect is the phenomenon whereby a subject modifies his or her behavior simply in response to the fact that they know they are being studied, and not because of the particular treatment of interest.

Second, the three components of pay flexibility likely complement each other in producing outcomes at the individual, agency, and public-sector wide level. For example, differentiated pay-setting is widely assumed to allow the agency to set pay at a level which is appropriate for the given task in the specific labor market in which the agency operates. However, to the extent that it also disrupts sector-wide pay negotiations, allowing government as employer to limit pay increases to the minimum necessary to maintain performance without the “leveling up” effect of a common pay framework, it can also have some effect in reducing pressures on the aggregate public sector wage bill. Delegated pay setting by giving managers authority to engage their staff in determining how best to keep the agency functioning within the hard constraints is also likely to have some effect on the culture of performance at the agency level.

There may also be a logical sequencing between the components. For example, delegation can lead to PRP since if pay-setting authority and accountability for results have been transferred to agency managers, they will look for mechanisms that maximize the likelihood that the results will be achieved at minimum cost. Delegation also automatically leads to differentiation to the extent that “local supplements will be used as a means of topping up the earnings of those professional staff in demand from private sector employers rather than of those for whom the state acts as a near monopsonist employer” (Grimshaw 1998a, p.7)

This paper is organized as follows. The next section details the main theoretical arguments for and against PRP and their hypothesized impact on staff effort and productivity. The bulk of the paper — sections 3 to 5 — focuses on the empirical evidence. The presentation of the empirical evidence is done with the aim of differentiating the findings by both the methodology of the study and the relevance of the study to different types of public sector jobs, such as the core civil service, tax administration jobs, teaching, and health sector jobs. Section 6 summarizes the main findings and the final section points to the under-explored agency and public sector wide potential effects of PRP and provides suggestions for future research.

It should also be noted here what this paper is not about. It is clear that pay policy is only one factor affecting staff incentives and public sector performance; other human resource management considerations including recruitment, personnel management, training, and organizational management practices are undoubtedly important. The paper reviews the empirical evidence from research that is limited to PRP, assuming that all else is constant; this limitation is both to keep the task manageable and in order to examine a variable which has a particular relationship with other aspects of human resource reform. It cannot reach a conclusion that when PRP “works” that it is more important than other variables. These are important issues which merit a different study.

2. Theoretical debates

Theoretical debates on PRP have been evolving in the context of private businesses (Prendergast 1998, 1999), the general public sector (Dixit 1999; Burgess and Ratto 2003; Perry, Mesch, et al. 2006) and on specific occupations such as teachers (Neal 2011). The theoretical arguments can be roughly divided between early psychological theories on human motivation and training, popular in public administration research, on the one side and core economic theories on incentive structures and principal-agent problems on the other, with behavioral economics building a bridge between the two.

Expectancy and reinforcement theory

Public administration research on PRP usually relies on what is often called “expectancy” (Vroom 1964; Porter and Lawler III 1968) and “reinforcement” theory (Skinner 1969; Luthans 1973). Expectancy theory builds on psychological insights about repeated behavioral patterns and learning under positive and negative stimuli. In its simplest form the theory suggests that explicit incentives in the form of

performance pay work under two conditions: First, employees need to believe that increased effort leads to increased performance, second increased performance leads to desired outcomes and is recognized by management. If the two conditions are met, employees form a behaviorally salient expectation about a future reward and adjust their work effort upwardly. Reinforcement theory stresses the effect of cultivating a behavioral norm of high work effort through reinforcing behavior with positive rewards. Apart from the direct link between performance and individual rewards, advocates in the field of public administration highlight secondary effects associated with performance pay: it helps to recruit and retain highly-skilled and/or motivated staff who presumably would do better under such an arrangement; increases the awareness for organizational goals by defining explicit performance standards; weakens the power of public sector unions; makes managers more responsible; signals core organizational goals to outside actors; increases the link between individual and organizational job goals; reduces the overall wage bill by moving away from automatic pay increases; and leads to an increase in overall job satisfaction through the individual recognition of employee efforts (Marsden 2004; OECD 2005b; Marsden 2009).

Critics of performance pay point out that the two conditions of expectancy theory are not always met and it is in principle difficult to design performance pay schemes that work as intended. Furthermore, critics argue humans do not always approach work effort and the assessment of salary in an entirely rational way and thus invalidate simple theoretical models based on rational-actor assumptions. In addition, measuring performance in the public sector is often fraught with difficulty. Many core public servants perform services on a daily basis that are hard to measure or are non-measurable, or produce outputs that are not market-priced. For example, early critics of test score based school and teacher evaluations argued that teacher performance cannot be neatly summarized by mechanic student test scores and such practice usually invites behavior that contradicts the overall goals of the teaching profession (Murnane and Cohen 1986). Interestingly, these criticisms were already leveled against early forms of test-based performance pay in British schools in the 19th century (Gratz 2009). Using explicit and objective performance measures can induce tunnel vision, myopia and measure fixation (Propper and Wilson 2003). Additionally, there is a lack of clarity about who determines and evaluates performance, and civil servants often work in large teams under the supervision of multiple managers, complicating the attribution of performance and responsibility of evaluation. Some authors see as a necessary condition the presence of high levels of trust and transparency between employees and management, to avoid arbitrary implementation and worker dissatisfaction (Kellough and Lu 1993).

An influential article (Kerr 1975) identified scenarios in which well-intended incentive schemes ended up favoring behavioral responses by employees that fulfill performance criteria, but taken to the extreme contradict overall organizational goals and standards. A different strand of criticism focuses on other motivations underlying public servants' effort. Apart from pure monetary rewards as expressed in salary, civil servants, it is argued, are motivated by notions of altruism, prosocial behavior and commitment to institutional goals (Perry and Hondeghem 2008), which are seen to compete with or even stand in conflict with explicit monetary incentives.

Incentive and principal-agent theory

Several surveys have synthesized the theoretical literature on pay incentive systems for the private sector (Prendergast 1998, 1999), for the public sector (Dixit 1999; Burgess and Ratto 2003) and specialized occupations such as teaching (Neal 2011). The most basic argument for incentive pay is founded in a simple microeconomic principal-agent model of labor relations, in which a principal (the employer) wants to induce an agent (the employee) to perform a certain task. Such principal-agent relationships are commonly affected by two problems (Dixit 1999): moral hazard and adverse selection.

Moral hazard describes a scenario in which the agent's actions affect the principal's payoffs, but the action is not directly observable to the principal. This situation arises naturally in the workplace setting (public or private), i.e. the employee's effort at work is not directly observable, but influences productivity and outcomes, which the employer cares about. Contracts that tie observable outputs, which are correlated with unobservable effort, to desired pay incentives can mitigate inefficiencies in the principal-agent relationship from the perspective of the employer. Offering fixed pay contract gives the employer little leverage to influence employee effort after hiring decisions have been made. The incentive problem is exacerbated if employees are hard to fire. Bonus or merit pay schemes are therefore one way of designing incentive schemes that address moral hazard.

In the case of adverse selection the agent has access to private and valuable information at the time of contract signing. To induce the agent to reveal this private information, the principal must offer attractive contract terms. Adverse selection in the public sector plays an important role in civil service recruitment, where low and high-skill applicants are hard to distinguish based on public information. Public agencies need to offer contracts that induce high-quality applicants to apply and deter low quality applicants from misrepresenting their qualifications. Merit pay systems are argued to alleviate this sorting problem and attract higher quality personnel that expect to perform well under system of merit pay, while traditional fixed salary scales are seen to attract low grade applicants (Delfgaauw and Dur, 2008).

Avoiding moral hazard and adverse selection are often used to advocate for forms of performance pay in the public sector. Such incentive schemes fundamentally require the ability to measure some relevant outputs, design a scheme that properly links unobserved actions to outcomes and offer bonuses that induce agents to increase effort. Incentives work best if the agent's actions are tightly linked to observable outcomes, i.e. when the random noise is not overpowering the incentive effects. Incentives schemes are also affected by the risk-aversiveness of employees. Since an incentive scheme links outcomes that are only partially under the control of the agent, making final pay outcome-dependent decreases the utility of risk-averse employees, who usually demand an upward adjustment of average pay to compensate for the increase in risk. Even with very simple models, the optimality of the incentive scheme is sensitive to important design aspects, like the schedule of bonuses (linear, stepwise or other) and depends on the particularities of the employee's task.

A well-known criticism of performance pay arrangements is that when tasks are multi-dimensional incentivizing only some tasks that are observable and measurable will not necessarily improve overall outcomes, but rather lead to a substitution of effort allocation from the unobservable to the observable tasks, which under some circumstances — depending on whether the different tasks are complements or substitutes, and the nature of the functional relationship between the tasks and the outcomes — can even lead to worse outcomes (Holmstrom and Milgrom 1991). For example, the task of teaching can involve both instruction based on sound curricula and coaching on test-taking strategies, and poorly designed incentive schemes can encourage teachers to re-allocate effort to the latter and away from the former (“teaching to the test”) to the detriment of human capital accumulation. Since it is generally hard to accurately measure and evaluate any aspect of public sector jobs, it is hard to devise effective civil service schemes that have the intended consequences with regard to final outcomes.

The problem of selecting appropriate performance measures to address this problem has spawned its own theoretical and empirical debates (Courty, Heinrich, et al. 2005). A problem related to the multi-tasking argument deals with the issue of gaming or cheating incentive systems. Typical examples are outright manipulation of results, cream-skimming, i.e. the manipulative selection of clients to improve program effects (Heckman, Heinrich, et al. 1997) or even the provision of high-caloric food to students during test days (Figlio and Winicki 2005). The problem of gaming performance standards is equally relevant in a

dynamic context, requiring ongoing adjustments by the principal (Courty and Marschke 2003). Luck during early periods of the evaluation period can induce increased slack for the rest of the time, a form of incentive gaming that has been found to be empirically relevant, as an influential study of Navy recruiters has shown (Asch 1990). To counteract excessive gaming of incentive schemes, it has been suggested, in the context of student test scores and teacher merit pay, to use evaluation systems independently from output measurements, i.e. in the case of teacher evaluations, to measure teacher contributions in tests different in form and content from instruments designed to track overall student and school progress or national exams (Neal 2011). Additionally, relative performance schemes in which employees are ranked against each other, potentially in a formal tournament setting are much harder to manipulate (Barlevy and Neal 2011; Neal 2011).

However, (Marsden 2009) notes that the gaming is not necessarily restricted to ingenious behaviors on the part of staff. Managers required to implement performance-pay arrangements might conclude that, in the bargaining with their staff concerning efforts and results, it is tempting and doubtless easier to: “collude with their subordinates: to go through the motions and fill in the forms for goal setting and appraisal, but not to worry about the reality” (Marsden 2009, p.5)

As outlined above, incentive schemes do not necessarily have to reward individual performance, but can focus on rewarding teams. Rewarding team performance can have certain advantages, ranging from reduced evaluation costs, to avoiding harmful competition between employees. However, basing rewards on team outputs can also lead to problems of free-riding where some team members willfully reduce their efforts in the expectation of relying on the work of others. The strength of free-riding problems depends on the size of the team and internal monitoring and punishment norms (Dixit 1999).

Picking the correct size of bonus brings its own challenges. Small bonuses will have little incentive effects and fall short of expectations, while large bonuses can lead to employees to treat incentive schemes as pure lotteries, especially if outcomes are strongly stochastic (e.g. student test scores (Neal 2011)), and to encourage cheating.

A reverse problem can arise if employees have to deal with multiple principals, a common feature of public service hierarchies. If different principals value different outputs, have different information available and have little ability or incentive to coordinate, separately designed incentive schemes are likely to fail (Dixit 1999).

Behavioral economics —intrinsic versus extrinsic motivation

Building on the argument that worker motivation is also driven by intrinsic concerns behavioral economists have advanced a line of argument that casts additional doubt on the feasibility of performance pay arrangements, specifically in public service settings. Introducing explicit monetary incentives for employees with strong intrinsic motivation can have the effect of crowding-out these intrinsic affects, i.e. workers change their perception about the organizational goals and values, leading to an overall reduction of effort. Even small changes in the pecuniary reward structure can induce a change in attitudes, switching from seeing the task as a partially voluntary contribution to a low paid contract service without any ownership stake. Crowding-out can be especially salient if performance pay is introduced using antagonistic framing and can stifle creativity and collaboration (Frey and Osterloh 1999). While the theory of intrinsic motivation has been proposed by psychologists, formal treatments of the trade-off between extrinsic and intrinsic motivation have been developed (Kreps 1997; Benabou and Tirole 2003, 2006) and integrated in the wider context of public goods provisions (Besley and Ghatak 2004).

The debate around the weight of intrinsic motivation within overall incentives has been crystallized in the debate led by Le Grand about whether public service workers are “knaves or knights” (Le Grand 2003). Le Grand argues that post-war public administration theory in the UK (and in Europe more generally) saw public servants as public-spirited altruists, a misleading interpretation of reality which was recognized in, although not adequately redressed by, the 1980s introduction of various New Public Management reforms.

On the opposing side of the argument, (Pink 2009) has developed the critique of monetary and other extrinsic incentives into a broader theory, hypothesizing that they are both counterproductive, as they frequently undermine intrinsic incentives, and unnecessary, as intrinsic incentives can be harnessed and used to maximize individual productivity. His theory suggests that tasks can be constructed to: (i) maximize an individual’s sense of *autonomy* (drawing inter alia on the “self determination theory” expounded in the overview paper by (Ryan and Deci 2000) and the cross-country attitude survey undertaken by (Chirkov, Ryan, et al. 2003)); (ii) *mastery* (continuous incremental learning and improvements rather than distant targets citing inter alia (Sauermann and Cohen 2008) who use data on over 11,000 industrial scientists and engineers to show that intrinsic motives, particularly the desire for intellectual challenge, appear to benefit innovation more than extrinsic motives such as pay); and (iii) *purpose* (drawing inter alia on (Niemic, Ryan, et al. 2009) who show from a follow up study of 246 students who graduated from two US colleges that those who met intrinsic aspirations for personal growth, close relationships, community involvement, and physical health had better scores for psychological satisfaction and health than those who pursued attainment of the extrinsic aspirations for money, fame, and image.

Another psychological argument, known as the “Yerkes-Dodson law”, highlights the phenomenon of “choking under pressure” (Ariely, Gneezy, et al. 2009). If individual salary is subject to high-stakes pressure, individuals experience increased arousal and shift from automatic to controlled behavior, a narrowing of attention and a pre-occupation with the reward, all lessening the chances of success. The argument is that performance has an “inverse U” relationship with the level of the incentive payment, with performance improving at low and moderate levels of incentive payments as compared to no payments, but then being worse at very high levels of payment compared to moderate, low, or even no payments.

Lastly, behavioral economists have identified the possibility of satisficing instead of maximizing behavior. Employees might exert effort until a certain minimum level of reward is reached and then substitute additional labor supply for increased leisure or idle time. A study of New York cab drivers identified the prevalence of satisficing behavior and questions the effectiveness of performance pay schemes, if the satisficing threshold is hit quickly (Camerer, Babcock, et al. 1997).

The proponents of “self-determination theory” clearly have a strong and intuitively attractive point to make, but it is not evident that the case for intrinsic incentives overwhelms consideration of extrinsic incentives. Three areas of uncertainty hang over conclusions in this area. First, autonomy, mastery and purpose are more feasible in some tasks than in others. Second, intrinsic motivation can take two forms — being motivated by the inherent nature of the job or being motivated to earn the respect of one’s peers — and group-based performance schemes, to the extent that they encourage teamwork, could increase intrinsic motivation. Finally, the theory is largely addressing the moral hazard problem inherent in principal-agent relationships, suggesting that staff can sense that they are being treated instrumentally through the use of extrinsic incentives and so will be motivated to cheat. There are different implications for the self-selection/sorting argument if the general population can be divided between predominantly intrinsically and extrinsically motivated people, since under those circumstances an explicit system of performance pay will attract extrinsically motivated applicants to the civil service.

3. Organizing the empirical evidence: “Craft” and “Coping” jobs

To summarize, the main argument for PRP is that linking incentives to inputs, outputs and outcomes can: (i) ameliorate moral hazard, inducing more effort, including the selection of better working methods to the extent that the individual has some autonomy; and (ii) address adverse selection by encouraging high ability individuals who will do better under a performance pay scheme to join the agency and similarly discouraging low ability individuals. Both of these linkages suggest contemporaneous and over-time effects of incentive schemes. Comparatively little research focuses on the self-selection or sorting effects of competitive performance schemes; the larger part of the literature deals with the stringent context conditions necessary for performance incentives to engender increased effort.

These contextual factors can be disaggregated into two categories: First, variables that are characteristics of the job that the individual performs and second, variables that are characteristics of the technical design of the performance scheme itself. The key “job” variables identified in the literature are:

- *Measurability of the goal or outcome from the job:* In some jobs outcomes are easier to measure than on others. Examples are production work on the factory assembly line or sales jobs, to be contrasted with managerial jobs in line ministries or large private bureaucracies;
- *Multi-dimensionality of the actions that produce the outcome:* The overall outcome of the job may depend on a single or multiple actions or activities. Then performance pay can create perverse outcomes by encouraging effort allocation towards the actions that positively impact the performance measure being used but negatively impact the overall outcome;
- *Observability and measurability of the actions that produce the outcome* (Wilson, 1989): Most public sector jobs involve multi-dimensionality of tasks; however, what distinguishes some tasks from others is whether or not the actions are observable. As James Q. Wilson noted, in some jobs it is much easier to measure whether or not an action was performed — e.g. health safety regulations were drafted — than whether or not the outcome, improved occupational safety, occurred. As (Dixit 1999) notes, the principle terrain of incentive theory are those jobs where actions are not observable;
- *Controllability of the outcomes* (Bruns et al, 2011): The extent to which the outcome is a function of the efforts of the individual or is influenced by other factors beyond the individual’s control.

Some of the main design choices in incentive schemes are:

- *Predictability of the incentive* (Bruns et al 2011): the probability of the agent receiving the incentive if the measured outcome is achieved. If the probability is either close to 0 or 1, then the incentive will have no impact. The incentive is clearest for example, for piece-work settings, common in sales or blue-collar settings in the private sector;
- *The size and nature of the incentive payment:* While the incentive effect should theoretically increase with the size of the bonus, the Yerkes-Dodson Law points to the consequences of very large bonuses. Large bonuses also can create incentives for cheating. Group based schemes could encourage team work but also result in free-riding behavior;
- *The nature of the performance evaluation and the performance standard:* Whether this should be an objective evaluation based on quantitative performance targets or a subjective evaluation, what is the benchmark against which it is evaluation, and who should be responsible for the evaluation.

The general criticism is that performance pay cannot be implemented in the public sector because political difficulties in selecting appropriate design features to address the complexity of the job variables — for example, giving bonuses to everyone thereby rendering it a completely predictable salary supplement and not an incentive payment. However, this criticism glosses over the many different organizational contexts within the public sector. For example the outcomes of schools and tax authorities are more measurable than those of central policy or administrative units in ministries and departments. Moreover, many large private sector bureaucracies also approximate public bureaucracies on some of these variables — while there is the profit “bottom line”, the controllability of outcomes can be low and the tasks that generate the outcome can be multidimensional. In examining the empirical evidence therefore, it is very important to distinguish between these different contexts in order to better understand under what conditions these schemes can or cannot work.

This paper borrows and slightly modifies Wilson’s typology (Table 1) to organize the empirical evidence so as to present these contextual nuances.⁵ Jobs can be characterized by whether or not the job’s outputs are easily measurable and whether or not the actions in the job to produce the output, or the internal production process, are observable. The matrix provides a framework within which to organize the empirical evidence by job type, with the simplifying assumption that jobs with multiple dimensions are located within the cell that represents the most complex of those dimensions. The top left box describes “Production Jobs” in which outcomes are easily measurable, the production process consists of repeatable, mechanical tasks that are observable to an outside monitor, and controllability is likely to be high. Typical examples are manufacturing factory-floor jobs, sales jobs, and municipal services like garbage collection. If the production process is not directly observable, but outputs remain measurable, such jobs are termed “Craft Jobs”. With recent advances in measuring learning outcomes, teaching can be classified as a public service in which the exact process of production is hard to fix, but, at least to a certain degree, desired outcomes are quantifiable. Similarly some of the outcomes of healthcare, particularly preventative services like child immunization, are also more measurable. Other examples include tax collection, job placement services, and auditing.

In the bottom row are “Procedural Jobs” and “Coping Jobs”. Both are characterized by difficult to measure outcomes, but again differ in the observability of the production process to an outsider. Procedural jobs like the military have clearly defined inputs, whereas administrative jobs in general policy units of the central government neither produce easily measurable outputs, nor have transparent production processes. Coping jobs present the most challenging functional contexts for PRP.

The general literature on performance pay is quite positive in relation to production jobs. Much of the earlier literature on incentive payments in the private sector looked at the impact of piece-rate compensation on productivity in production organizations involving manufacturing jobs or on similar repetitive tasks. (Stajkovic and Luthans 2003) conducted a meta-analysis of 72 empirical studies that investigate the impact of financial, as well as other, incentives on various measures of performance in organizational settings. These studies were conducted in a variety of private organizational settings, and

⁵ Wilson had originally used this framework to classify organizations and not jobs, the implicit assumption being that organizations were homogenous in the tasks that they performed. This adaption of the framework to jobs does not change the logic of the typology, and is consistent with the proposal made by (Pritchett and Woolcock 2004) to classify decision-making functions in the public sector according to the discretion inherent in the task (vs. simple rule-following) and the number of the transactions necessary to deliver the service. Tasks which are necessarily discretionary and transaction intensive are “practice” tasks in Pritchett’s logic and “coping” tasks in Wilson’s view – but essentially refer to the same type of job – and, as Pritchett notes: “(t)he provision of key, discretionary, transaction-intensive services through the public sector is the mother of all institutional and organizational design problems” (Pritchett and Woolcock 2004, p.196).

included both industrial and service sector organizations. The study showed that financial incentives alone improved task performance by 23 percent, whereas financial incentives combined with social recognition and positive feedback from superiors increased task performance by 45 percent.

Table 1: James Q. Wilson's classification of job types

		Actions or internal production process of the job	
		Observable	Not observable
Outputs from the job	Relatively easily measurable	<p>Production job: Simple repetitive stable tasks, specialized skills.</p> <p>Examples: Manufacturing, sales, simpler municipal services (garbage collection).</p>	<p>Craft jobs: Application of general sets of skills to unique tasks, but with stable, similar outcomes.</p> <p>Examples: Auditing; revenue collection; teaching; medical practice; Job placement work</p>
	Not easily measurable	<p>Procedural job: Specialized skills; stable tasks, but unique outcomes</p> <p>Examples: Military</p>	<p>Coping job: Application of generic skills to unique tasks, but outcomes cannot be evaluated in absence of alternatives</p> <p>Examples: Administration; managerial jobs in large private sector organizations</p>

A well-known study by (Lazear 2000) uses individual-level worker data from a glass company and estimates a 44% increase in productivity after switching to a piece-rate salary schedule. In a structurally similar setting, the introduction of piece-rate pay in a Canadian tree planting business has also been found to strongly affect worker productivity and profits (Paarsch and Shearer 1999).

However, production jobs, as well as procedural jobs, are of limited relevance to the public sector. Rather, given that the loci for incentive schemes, following the principal-agent literature, are mostly in contexts where actions are not easily observable but outcomes may or may not be, the focus of this paper will be on jobs in the right hand column of Table 1. Note that some of the studies of the private sector that have focused on incentives for rank-and-file “knowledge workers” in large organizations can also be classified in this right hand column and will be included in the discussion.

Methodological approaches

The vast empirical literature on PRP utilizes a range of methodological approaches. Early studies on performance pay in the public sector were largely qualitative case studies. As elements of civil service reform and incentive schemes were introduced in OECD countries throughout the 1980s and 1990s, academics and policy experts wrote initial studies and reports that summarized the main descriptive feature of reform attempts, and chronicled the reform process and implementation. Overall evaluation of the success of incentive schemes was based on qualitative impressions of practitioners. Later studies

approached questions of performance pay slightly more systematically, comparing several reform cases with each other and using convenience samples of employees and senior management, subject to performance pay, to collect data on self-assessed motivation and satisfaction with newly introduced performance pay. While attempting to provide a systematic evaluation of the successes and failures of performance pay reform, to a large extent these studies rely on weak research designs. Despite their descriptive value, inferences based on the single or comparative evaluation of reform cases, without the use of counterfactuals, cannot establish the existence or absence of potential program effects. Furthermore, utilizing convenience samples and using survey instruments that only measure self-assessed motivation suffer from selection and perception bias problems, while ignoring the information on actual outcomes.

Although the tenor of this initial wave of research was rather discouraging about the results on the effectiveness and popularity of performance pay among staff, incentive schemes for the public sector did not lose their appeal to policy makers (Marsden 2009). An intense empirical debate on teacher incentive was sparked in the context of local American school reforms and a national debate on test-score based school accountability. A series of papers by education economists used the opportunity of local reform attempts to collect data on samples of schools and students subject to new performance pay arrangements and comparable control populations. By measuring detailed public service outcomes, student's grades and test scores, drop-out rates and attendance records, plus detailed information on bonus programs, the quantitative evaluation of this new data allowed a more stringent test of the competing theoretical arguments. The advantage of these studies compared to prior work is their ability to disentangle the effect of teacher effort from other factors that also determine student outcomes.

While an improvement, many of these quantitative observational studies and similar work on private companies and the public service still fall short of an ideal research design for causal inference on program effects. The gold standard for program evaluation is the use of randomized-controlled trials (RCT), in which treatment assignment to subjects is randomized and unrelated to other observable and unobservable characteristics. This randomization allows the estimation of the treatment effect by a comparison of treated and control units. Observational studies on the other hand rely on treatment and control groups created not by controlled random assignment, but produced by social processes, often related to the research question at hand. Issues of selection bias and confounding factors can undermine the internal validity of such studies.

Utilizing the power of a randomization framework, several behavioral economists have used laboratory experiments to test hypotheses with regard to incentive schemes. Laboratory experiments offer at least two distinct advantages over observational studies: the researcher can use randomization to ensure unbiased and consistent estimation of treatment effects; and researchers can design their experiment to directly relate to theoretical questions at hand. The review of the theoretical literature has identified the importance of design details and the plethora of possibilities when it comes to performance pay arrangements. Observational studies have to rely on bonus pay schemes that have been implemented in real life, which do not necessarily cover all interesting variations and often mix different elements that conflate theoretical questions of interest. Being able to design a clean and tailored experiment to trace the effect of intrinsic versus extrinsic motivation in a linear bonus pay scheme for example is a huge advantage which laboratory experiments offer. A wave of articles in the late 1990s and 2000s explores various issues of performance pay. While offering certain advantages over observational studies, laboratory experiments often use notoriously small samples and student subjects that share few characteristics with actual workers or public servants. Furthermore, laboratory experiments can hardly ever replicate real work place settings or offer bonus schemes that remotely approach bonus sizes common even in only moderately incentivized performance pay schemes. This raises concerns from a sampling perspective and the representativeness of the subject pool, as well as the comparability of

laboratory treatments and real-world bonus programs. Researchers generally confer low external validity to laboratory experiments and caution against the isolated interpretation of results derived from a single experiment.

The most recent attempts at addressing the issue of proper causal inference on performance pay and increasing the representativeness of results are RCTs. In an RCT researchers are able to randomize key features of an actual policy program that services the actual population of interest. The advantage of such a field experiment is the similarity of the target population, the structure of the incentive program, paired with the randomization of treatment. Although field RCT are time and resource intensive studies, several teams of researchers have implemented similar studies in different contexts around the world, considerably adding to the empirical understanding of performance pay.

4. The empirical literature reviewed

Given that the literature has analyzed incentive pay for a broad set of jobs and utilizing an array of methodological approaches, the following presentation of the empirical literature in this paper is sorted by (a) the nature of the job, using Wilson’s typology and (b) the methodological approach. In total 153 empirical studies of PRP (see the Appendix for the full list) were considered in this review, of which 110 are for craft and coping jobs (Table 2). The research to date on the subject has largely focused on advanced countries — in the review 127 studies are in OECD contexts, and only 26 are in developing country settings. The literature has also focused largely on craft jobs and production jobs, with no experimental studies to date on coping jobs.

Table 2: Studies by country environment, methodology, and job type

Country and methodology	Types of Jobs					Total
	<i>Production jobs</i>	<i>Procedural jobs</i>	<i>Coping jobs</i>	<i>Craft jobs</i>	<i>Unclassified</i>	
OECD study	27	0	15	72	13	127
Observational	14	0	15	59	13	101
Field RCT	7	0	0	13	0	20
Lab. experiment	6	0	0	0	0	6
Developing country study	1	0	1	22	2	26
Observational	0	0	1	15	2	18
Field RCT	0	0	0	6	0	6
Lab. experiment	1	0	0	1	0	2
Total	28	0	16	94	15	153

Observational studies

Public sector coping jobs

Observational studies about performance-related pay for public sector coping jobs suggest somewhat limited effectiveness of this component of pay flexibility for these activities — they also focus overwhelmingly on OECD settings.

A series of OECD reports and associated discussion papers chronicle the type and extent of pay related civil service reforms in advanced industrialized countries (OECD 1993, 1996, 1997b; Kim 2002; Burgess and Ratto 2003; OECD 2004a, 2005a, b; Perry, Mesch, et al. 2006; Ketelaar, Manning, et al. 2007; Rexed, Moll, et al. 2007; OECD 2008, 2009; Perry, Engbers, et al. 2009). Generally, these refer to coping jobs in that they refer to, or imply that they are referring to, managers with complex tasks.

(Cardona 2007) reviews incentive programs in the US, particularly the Performance Management and Recognition System, the UK's Inland Revenue Service performance scheme and similar attempts in Australia. The study documents several common issues in the implementation of performance pay: employees are hardly ever scored less than satisfactory in their evaluations, bonus systems were designed so that only very few employees actually received any payments and the majority of staff found the system de-motivating and inciting jealousies. (Straberg 2010) highlights the problem of perceived unfairness following the introduction of performance pay in an OECD country, although also notes that there was no empirical linkage between pay justice perceptions and workplace behaviors. Managers equally found little positive changes resulting from the introduction of performance pay. As context, multiple studies confirm the political and operational difficulties of successfully introducing any major program of pay reform within the public service (World Bank 1999; Kiragu and Mukandala 2003; Independent Evaluation Group 2008).

Brudney and Condrey present a pair of typical studies drawing on non-representative surveys of federal managers and government employees in the US, that were subject to performance-pay arrangements (Condrey and Brudney 1992; Brudney and Condrey 1993). While largely descriptive in nature, they find that 17% of managers report an increase in motivation, but prior attitudes about merit pay affect this result. It is unclear though whether the documented increase of motivation is solely an effect of performance pay, since no control units are used to rule out the influence of other possible confounders or if self-reported motivation is in any way linked to actual work outputs. Similar work on performance pay for civil servants in the US state of Georgia also finds highly critical opinions of staff members with regard to explicit evaluation and selective bonus pay (Kellough and Nigro 2002). More recent studies using survey data of US city managers finds higher satisfaction of employees when performance pay is used (Stazyk 2010). The research does however highlight the complexity of the issue as it suggests that there is a crowding-out effect on extrinsic motivation for staff with a distinctively high level of intrinsic motivation — but ultimately this does no harm to effort or to job satisfaction.

(Dowling and Richardson 1997) evaluate the effect of performance pay on UK National Health Service managers, a study, since it focuses on management rather than physicians, that is more closely related to core public administration jobs than healthcare jobs. Using self-reported data from a survey, they find a modest positive effect of pay incentives on manager motivation and effort.

Whether PRP should be used is one thing — when and where it can be used is another. (Dahlstrom and Lapuente 2009) theorizes, using transaction cost economics, that when senior officials share a career path with elected politicians and thus are unlikely to act impartially, then PRP is less often used as those officials are unable to provide credible commitments to their staff. Why would one work hard to achieve a goal if it transpires that the goalposts might be arbitrarily moved for political reasons at the last moment? They find empirical support for this prediction.

Public sector craft jobs

Observational studies about performance-related pay for public sector craft jobs are more optimistic concerning their effectiveness. There is more evidence to draw on from developing countries, particularly in relation to teaching and health care.

Tax administration, job placement

Revenue authorities are examples of public sector agencies that can be classified as craft organizations where outputs — number of audits conducted and tax fines collected — are more easily measurable and there is a clearer link between the efforts of, for example, individual tax auditors and revenue collection.

A good example of detailed pay PRP and high-powered performance pay incentives comes from the federal Brazilian tax collection agency. In 1988 the Brazilian government created a bonus program for tax officials that rewarded the identification of tax violations. Base salary was augmented on a monthly basis on an individual and group basis. The group reward was calculated based on the relative performance of one local agency versus others, with relative performance measured based on total fines collected, attainment of pre-defined quotas (total tax collection, number of inspections, collection of overdue taxes) and the size of the agency. Individual rewards were based on monthly evaluations by the direct supervisor, which combined objective performance criteria and managerial discretion and rated employees on a scale from zero to 70. Each employee that scored more than 21 points was entitled to an individual reward with the value being determined by the overall availability of funds (which are proportional to the collected fines) and the performance of co-workers. It was not unusual for total bonus payments to reach 200% of base pay.

(Kahn, De Silva, et al. 2001) found that this incentive scheme resulted in a 75% increase in fines per inspection. At the same time, they also found substantial regional variation, with responses ranging from 19% to 145%. The authors do caution that diverse management techniques resulted in some regions targeting wealthier sources (such as corporations), more aggressively, which points to the potential negative effects of such high powered incentive schemes which may encourage extortion. Unfortunately, limited data prevented the authors from examining these social costs further.

Another study of revenue authorities is by (Bertelli 2006), which explores the potential tradeoff between intrinsic and extrinsic motivation in the Internal Revenue Service in the US. The IRS implemented a paybanding system that imposed high-powered performance incentives on supervisors, but not on non-supervisory personnel. Using data from the 2002 Federal Human Capital Survey, the author showed that the incentive scheme crowded in intrinsic motivation at the lowest pay levels, and crowded out at the highest levels.

(World Bank 2001) used survey data from revenue departments in 14 low, middle and high income countries, and detailed case study evidence from 7 of those, to review the effectiveness of bonus and salary supplement systems as a means to enhance effectiveness in revenue departments. They concluded that the “circumstantial evidence” suggests that bonus systems do indeed seem to have an impact on organizational effectiveness. They note that in a number of countries the introduction of bonus systems have had a measurable impact on recruitment and retention of employees. However, they note that the success of bonus systems relies heavily on “legitimacy”, i.e. the internal and external “acceptance” of the bonus system.

A set of studies of performance incentives in a similar organizational context is that of the US Job Training Partnership Act (JTPA). Under the Act, 620 semi-autonomous training centers were responsible for implementing job training programs for the indigent, and were given financial incentives tied to labor market outcomes — employment status, earnings — of the trainees. These bonuses were given to the training centers thereby augmenting their budgets but could not be used to supplement staff salaries. (Courty and Marschke 2004) find evidence of the prevalence of gaming among the agency staff in the choice of termination date of the training for the participants, which while increasing organizational bonuses imposed a cost to the participants in terms of earnings. Similar effects have been found by related studies of the program (Heckman, Heinrich, et al. 1997).

An early quantitative observational study of performance pay in the public service was implemented by (Asch 1990), who collected data on the behavior of Navy recruiters, subject to a point-based performance system. The incentive consisted of a point-scheme for the quality of recruited candidates, a fixed time frame for evaluation and a minimum threshold of points needed to qualify for a bonus. Asch shows that

the incentive scheme did increase the effort of recruiters and led to the recruitment of more high-quality candidates, but also induced recruiters to exhibit a form of gaming behavior, i.e. increasing recruiting efforts early in the cycle and once an expectation of reaching the bonus level was achieved, a reduction of effort for the remaining evaluation time followed.⁶

Teaching

The literature on performance pay for teachers, an essentially craft job, also shows a variety of research designs. Some researchers use qualitative case studies (Murnane and Cohen 1986) or perception surveys of teachers subject to performance pay to measure effects on performance (Heneman III and Milanowski 1999; Kelley 1999). These studies often show a low degree of satisfaction with bonus systems and explicit evaluation. By contrast, studies focusing on actual outcomes often find rather encouraging results. A series of papers use the introduction of performance pay for teachers to quantitatively assess the effects on student outcomes. By now studies have utilized various data sources and structures to assess the effects of performance programs.

In the American context researchers have evaluated the effects of teacher quality on student outcomes (Goldhaber and Brewer 2000; Hanushek and Rivkin 2006; Clotfelter, Glennie, et al. 2007). Since performance pay is argued to be one important tool for attracting and retaining highly-qualified teachers, it is important to determine the effectiveness of merit systems in that regard. (Clotfelter, Diaz, et al. 2004; Clotfelter, Glennie, et al. 2008) show, using detailed data from North Carolina's schools, that accountability and performance pay systems contribute positively to retaining quality teachers. The introduction of merit pay can also be linked to student test scores, but with varying empirical robustness. (Cooper and Cohn 1997) find for a sample of over 500 South Carolina classes a positive effect of merit awards for teachers on mathematics and reading test score achievements.

Cross-sectional studies using data from the American National Educational Longitudinal Survey have been used to show a positive link between individual merit awards for teachers and student test scores (Figlio and Kenny 2007). Positive effects of performance pay have also been found in Arkansas kindergartens (Winters, Ritter, et al. 2009). To mitigate problems of causal inference, sometimes researchers can use difference-in-difference estimation, due to the geographic difference of the phasing-in of reforms (Eberts, Hollenbeck, et al. 2002; Atkinson, Burgess, et al. 2004). While (Eberts, Hollenbeck, et al. 2002) find no effects on student test scores and even slightly negative effects on other outcomes, their analysis relies on student-level data from only two schools. A more thorough and systematic difference-in-difference analysis by Atkinson et al. finds clear positive effects of performance pay for British schools. By utilizing particular features of Tennessee's Career Ladder System and the Project STAR field experiment, (Dee and Keys 2004) are able to link teachers' quality assessments, as expressed in the career ladder grouping, to student test scores. They find that the official career ladder system had only mixed success in rewarding teachers with the highest test score gains, but nonetheless teachers with merit awards had positive effects on student's math scores. They found however no statistically significant effects on reading scores.

⁶ An early observational study which revealed significant gaming is noted in (Wilms and Chapleau 1999) who note that performance-based pay began in the UK in about 1710, with salaries based on test scores in reading, writing and arithmetic. The rationale was that it would help keep students from poor families in school, where they could learn the basics. In reality, the incentives encouraged teachers to narrow the curricula to include only easily assessed subjects, and cheating by both inspectors and teachers made the system ultimately untenable. The system was dropped in the 1890s. A similar scheme was introduced briefly in Canada in 1876, but it ran into similar difficulties and was terminated around the same time.

Several studies from the context of American school reform also document the role of unintended side effects of explicit accountability programs. Large-scale testing of students as part of the 2001 No Child Left Behind Act ties student test scores to important resource allocations for schools, giving schools incentives to improve student learning, but also to increase pure test-taking ability or to engage in outright cheating (Jacob and Levitt 2003; Jacob 2005). Quite surprisingly, (Figlio and Winicki 2005), using daily lunch menu data from a random sample of 23 school districts in Virginia, show that even the caloric content of school lunches was adjusted upwardly to improve cognitive ability on test days.

An interesting study of private schools in India assesses the role of teacher unionization on student outcomes (Kingdon and Teal 2008). While not explicitly evaluating performance pay, unionization of teachers represents an increase in job security and uniform, higher pay, without being linked to explicit performance standards. (Kingdon and Teal 2008) find strong negative effects of unionization on student outcomes, utilizing a within-pupil across subject variation fixed effects design. The study by (Ladd 1999), mentioned above, on school accountability in Dallas, uses panel data and finds positive effects of merit pay on student performance and dropout rates.

A set of observational studies (Lavy 2008, 2009) uses data from an Israeli policy experiment with tournament based teacher competition for bonuses. Using regression discontinuity and difference-in-difference designs to approximate random treatment assignment, the study shows significant gains in student achievements. The studies also assess potential mechanisms for the link between performance pay and improved test scores, identifying a change in teaching methods, enhanced after-school teaching and increased teacher responsiveness over test-score manipulation as important causal channels.

A comprehensive study using cross-national data on performance systems in schools and PISA test scores also finds an positive association between pay-for-performance type reforms, improved teacher quality and student test scores (Woessman 2010).

Health care jobs

A fairly large literature has dealt with the role of performance pay in improving health care delivery and services. General literature reviews (Petersen, D., et al. 2006; C. and N 2009) reflect the diversity in empirical approaches and the strong focus on OECD country experiences. In particular performance incentives in the UK NHS system and several U.S. insurance systems have received attention in the research literature.

A series of studies has evaluated the potential effects of financial incentives for primary care physicians. The British NHS introduced performance-pay elements into the remuneration of family practitioners in 2004. (Doran, Fullwood, et al. 2006) use data on over 8000 family practices and evaluate the effect of performance pay on patient outcomes and find overall high performance in the first year of the incentive scheme, but also evidence of “gaming” through the exclusion of patients. (Campbell, Roland, et al. 2005; Campbell, Reeves, et al. 2007) also analyze the role of financial incentives in a stratified random sample of British general practices focusing on care for coronary heart disease, asthma and type 2 diabetes, finding a substantial effect of financial incentives introduced in 2004. (Steel, Maissey, et al. 2007) find positive effects on asthma and hypertension treatment, as do (Vaghela, Ashworth, et al. 2009).

(Chalkley, Tilley, et al. 2010), using evidence derived from a natural experiment in the UK publicly funded dental care system, analyzed the effects of using incentive pay that provided explicit rewards for increased service provision against the alternative of offering an employment-like relationship. They found that dentists who were moved from quasi-employment arrangement to an activity-based incentive contract increased their activity in the publicly funded service by 26%. They also found evidence of

considerable variation between suppliers, which suggests that factors such as an individual's intrinsic motivation, professional standards, and preferences were important moderators of financial incentives.

In the U.S. context, several insurance providers and health maintenance organizations have experimented with elements of financial incentives for care providers in various states. An early study of health maintenance organizations (HMO) managers' views on financial incentives found mixed support for the effectiveness of performance pay (Hillman, Pauly, et al. 1991) in the eyes of managers.

More recently, several studies have found fairly positive effects of performance incentives on health services, patient outcomes and satisfaction. (Safran, Rogers, et al. 2000) use a cross-sectional study of Massachusetts adults to assess the effects of various health-maintenance organizations (HMO) and their specific contract elements on primary care. One of the results links financial incentives for physicians to patient satisfaction. An evaluation of performance pay pilot program for physicians found meaningful improvements for diabetes patients compared to the control group (D. and Horrigan 2005). Small positive or mixed effects were also found studies by (Amundson, Solberg, et al. 2003; Casalino, Gillies, et al. 2003; McMenamin, Schauffler, et al. 2003; Levin-Scherz, DeVita, et al. 2006; Coleman, Reiter, et al. 2007; Felt-Lisk, Gimm, et al. 2007; Mandel and Kotagal 2007; Young, Meterko, et al. 2007).

(Rosenthal, Frank, et al. 2005) analyze a natural experiment, comparing quality improvements in two physician groups in the U.S. from 2001 to 2004. They find improvements in cervical cancer screenings but not other outcomes, largely rewarding practices with a high baseline performance. In a cross-sectional sample of primary care physicians that contracted with Medicaid managed care organizations in 2002 in California found a partially positive effect of incentive pay on STD care (Pourat, Rice, et al. 2005).

(Lindenauer, Remus, et al. 2007) analyze the effects of public reporting and pay-for-performance in hospital care in a Medicare/Medicaid demonstration project. Hospitals participating in the performance scheme show a significant improvement in overall measures of patient care quality, including care for heart failure, acute myocardial infection and pneumonia by up to 16%, compared to the control group. In a related, but patient-level study (Glickman, Ou, et al. 2007) evaluate the largest pay-for-performance pilot project in the U.S., finding no conclusive effects for several treatments and patient outcomes.

Similarly, (Pearson, Schneider, et al. 2008) find that pay-for-performance elements in physician contracts in Massachusetts did not add any significant gains above and beyond secular improvement in a time period from 2001-2003. In a study of public community health centers in Houston, (Gavagan, Du, et al. 2010) also find no effects of performance pay, while (Chung, Palaniappan, et al. 2010) find no effects for primary care physicians in California. Mirroring the results found in other areas, while financial incentives can improve particular behavioral responses of staff members, it is difficult to design an incentive scheme that does not also produce unintended consequences and rewards unwanted behavior. (Shen 2003) provides evidence of gaming and selection effects of financial incentives for substance abuse care providers. (Li, Hurley, et al. 2011) utilize a natural experiment in Ontario, assessing the effect of performance-related pay on physician behavior and targeted primary care provision. They do find positive results for some, but not all financial incentives, providing a cautionary message with regard to the potential impact of performance pay.

Importantly, all these studies evaluate performance pay for specific, easily measurable tasks, in a highly institutionalized environment with powerful monitoring capabilities. Observational studies dealing with performance pay in the health care sector in the developing country context are much fewer in number and can draw on fewer large-scale experiences. Despite the lack of widespread use of incentive pay in the developing world, incentive focused reforms for doctors and nurses have received some attention. (Vujicic 2009) outlines the potential benefits of performance pay for health services staff, but also

identifies the risks of supplier-induced demand and cost-explosions. He also correctly distinguishes between performance pay within health care units and the general contracting-out of health services to NGOs. While contracting of services though often includes overall performance targets and entails the introduction of staff performance incentives within contracted facilities, this review focuses explicitly only on performance pay elements in health care facilities.⁷

Studies focusing on staff-level performance pay in low and middle income countries generally find positive results, but largely illustrate the lack of systematic findings and evidence. (McNamara 2005) discusses overall six cases of payment for quality in the health services sector across developed and developing countries, with cases in Nicaragua and Haiti having had a positive effect. The Nicaraguan reform efforts though combined decentralization of decision-making authority, increased local accountability with explicit performance agreements, and while being judged to have led to an overall improvement (Jack 2003), it is hard to disentangle the effects of each reform element.

Similarly, in a recent study by (Witter, Zulfikur, et al. 2011) a pay-for-performance arrangement in a NGO-led health project in the Battagram district of Pakistan was evaluated and found to have improved general services provision, but with an unclear effect of the performance-based elements. The study highlights though the weak link between bonus pay and performance, as well as the low amount of monetary incentives in relative terms.

A study of health care reform efforts in two Rwandan districts shows that the use of performance elements paired with increased autonomy seems to offer a viable and cost-effective way to improve health care delivery (Meessen, Musango, et al. 2006). In a later study, (Meessen, Kashala, et al. 2007) evaluate the performance of 15 health centers in Kabutare, Rwanda. They document a sharp increase in staff productivity after the introduction of output-based bonuses. (Soeters, Habineza, et al. 2006) highlight the potential applicability of the Rwandan experience in sub-Saharan Africa more generally.

Similarly, efforts to improve health services provision in Haiti using performance-based payment for NGOs in a USAID pilot project showed encouraging effects on immunization coverage and organizational behavior (Eichler, Auxila, et al. 2001).

A recent book by (Eichler and Levine 2009) outlines the general argument for the use of explicit incentive schemes in the provision of health services in developing countries. Apart from demand-driven financial incentives through conditional cash transfer programs, they discuss the role of performance-pay and contracting-out of services provision. They review evidence from various contexts, in particular experiences with contracting NGOs in Afghanistan and project evaluations in Haiti and Nicaragua, overall advocating the increased use of financial incentives.

Private sector: Craft or coping jobs

Observational studies on performance-related pay in private sector craft or coping jobs are generally suggestive that PRP has a positive impact – but highlights the importance of careful design.

There is a newer and growing empirical literature that has looked at incentive payments in private organizational settings that are closer to that of ministries and departments, characterized by low measurability and low controllability of tasks. These organizational settings are akin to craft or, in some

⁷ (Loevinsohn and Harding 2005) review the success of ten contracting-out projects in the developing world, finding largely encouraging results.

cases, coping organizations. Analyses of incentive schemes in such organizational settings also exhibit a large variation in the research designs. (Beer and Cannon 2004) study the failure of 13 incentive plans at Hewlett Packard using interviews and internal documents, finding that managers abandoned the programs due to the perceived costs. Providing slightly more comprehensive evidence, in a worldwide survey of 205 top managers (Beer and Katz 2003) document the weak support of incentive schemes among management. Both studies echo some concerns identified in similar public sector studies, but suffer from weaknesses in their research design.

A number of quantitative observational studies improve upon prior work by using more representative samples. (Belfield and Marsden 2003) use panel data from a large UK work place survey, which include both piece-rate jobs and “knowledge work” jobs in which performance is based on achievement of previously agreed goals, and find strong effects of individual pay-for-performance, but only conditionally on the monitoring regime. Another study uses the British Household Panel Survey to distinguish the productivity and sorting effects of performance pay, finding that jobs with performance-related pay attract workers of higher ability and induce workers to provide greater effort (Booth and Frank 1999).

(Blasi, Freeman, et al. 2008) econometrically examine the relationship between various forms of profit sharing and stock options on staff turnover, absenteeism, effort, and other productivity measures. The analysis is based on two large surveys of private sector firms, and finds statistically significant positive linkages between these shared capitalism schemes and perceptions of workplace performance such as turnover, loyalty, and worker effort. These incentives have the strongest impact when combined with other organizational variables such as competitive wages, training, and employee involvement in workplace policies.

(Hochberg and Lindsey 2010) is one of the few empirical examinations of the impact of stock options on company rank-and-file on firm performance (as opposed to the impact of options on top executives, on which there is a large literature). Since stock options are a group incentive based on overall company performance, incentives to free-ride are very high. However, alternative literature also suggests that non-executive compensation may also increase cooperation and encourage mutual monitoring among co-workers. Using a large database for a broad set of firms, and explicitly controlling for endogeneity, the study shows that stock options exert a positive effect on firm performance. The study also finds that this effect is higher on smaller firms, consistent with the free-riding hypothesis, and also higher in firms with higher growth opportunities where the monetary incentive is higher.

(Aboody, Johnson, et al. 2007) also examine the impact of executive and non-executive stock options on firms’ operating performance, based on an empirical investigation of a sample of 1300 firms and contrasting between firms that re-priced their options to make them an attractive financial incentive versus those that did not. The study finds that while firms that repriced their options had a larger increase in operating income and cash flows compared to non-repricers, this impact was entirely due to executive stock options. The repricing of non-executive stock options had no impact, consistent with the free-rider argument, and in contrast to the (Hochberg and Lindsey 2010).

Some studies from the private sector have also highlighted other conditional variables, such as strength of social networks, that impact pay incentives. (Bandiera, Barankay, et al. 2005) use data from a farm labor operation to compare the effects of piece-rate versus relative incentive schemes, showing that individual piece-rates enhance effort irrespective of social relations among workers, whereas the effects of relative incentive schemes, which impose negative externalities among workers, is diminished when workers have stronger social ties. On the other hand, if incentives explicitly recognize team efforts, team rewards can outperform individual piece-rate wages, as one study on team work in a garment plant has shown

(Hamilton, Nickerson, et al. 2003). Data from a large group incentive scheme at Continental Airlines even found positive effects in the presence of strong free-rider incentives (Knez and Simester 2001).

Experimental studies

Among experimental studies a distinction has to be made between laboratory and field experiments, with each having their own advantages and disadvantages. Laboratory experiments are characterized by their strong control over experimental design and treatment specification, but often lack the ability to use representative subjects or simulate convincingly real-world settings. Field experiments combine the randomized assignment of treatment with actual real-world programs and participants, but are often more constrained in their design choices, face stronger exogenous pressures to succeed and are resource-intensive.

Meta-studies

(Jenkins, Mitra, et al. 1998) provided the first meta-analysis of the psychological literature on the impact of financial incentives on performance and concluded, based on analysis of 47 studies that these incentives resulted in a 12% improvement in performance quantity and a negligible effect on performance quality. The study however had several limitations as it did not distinguish between types of incentive programs, complexity of tasks, and differences in organizational settings.

Building on the Jenkins study, (Condley, Clark, et al. 2003) conducted a meta-analysis of 64 field and laboratory experiments, as well as observational studies, on the impact of monetary and non-monetary individual and group incentives on performance. The criteria for inclusion of studies in the analysis was that studies had to (a) have a use a control group or a pre-treatment measure of average performance; (b) involve the use of incentives to enhance performance; and (c) and report some statistical data. The studies included private sector settings, public sector, as well as laboratory experiments with college students. Within the public sector, however the vast majority were of schools, with only one study in a core government setting, thereby greatly limiting the generalizability of the findings to the core civil service. Importantly, the analysis also distinguished between studies that looked at cognitive (38 studies) versus mechanical manual tasks (26 studies), and the measurability of the tasks, looking at both quantitative and qualitative performance targets. The results of the meta-analysis showed that employees and other research participants who received performance incentives achieved an average 22% increase in work performance. The findings were irrespective of the settings, although again it should be noted that there was only one study of performance incentive in a government agency. Monetary incentives were found to be more effective than non-monetary gifts, and group-based incentives were significantly more effective (48% increase in performance) than individual incentives (19% increase). While the incentives worked for both cognitive and physical tasks, the gains were higher for the latter (30% increase compared to 20% increase). No significant difference was found based on the measurability of the tasks.

(Weibel, Rost, et al. 2009) conduct a meta analysis of 46 high quality empirical studies published in the fields of economics and psychology and covering both simple and complex tasks, and with both quantitative and qualitative outcome measurements. Overall the study finds a statistically significant and positive effect of pay for performance on performance; however, and in contrast to the above, the findings were significantly positive for simple tasks and smaller, but still statistically significant, negative effect for complex tasks. The authors argue that this negative effect is due to the reduction in intrinsic motivation brought about by the incentive scheme.

Laboratory experiments

Building on these psychological studies, a number of behavioral economists have done laboratory experiments to explore the different aspects of performance pay, including the functional relationship between bonus sizes and performance, the incentive and sorting affects of performance pay, the impact of different types of bonuses, and the possible tradeoffs between extrinsic and intrinsic motivation.

(Ariely, Gneezy, et al. 2009), explore the effect of bonus size on performance in laboratory experiments using subjects in the US and India, with 24 and 87 participants respectively. Participants had to solve cognitive tasks under time pressure and were incentivized with bonuses that varied from small to large relative to their normal pay. They found evidence for an “inverse-U” relationship between bonus size and performance, with the “choking-under-pressure” effect where bonuses at very high levels lead to a worsening of performance compared to bonuses at low and moderate levels.

An experiment with 115 Australian students that tried to distinguish the potential incentive and sorting effect of performance pay found supportive evidence for both hypotheses (Cadsby, Song, et al. 2007). In addition they found not only that low productivity subjects were less likely to sort into pay-for-performance jobs, but also that subjects with higher levels of risk-aversion avoided pay-for-performance, suggesting important unintended side effects. The experimental comparison of piece-rate, team rewards and relative performance schemes found that piece-rate systems and team rewards overall induce similar effort levels (free-riding in team rewards is compensated by higher effort contributions). Effort was higher, but also more variable in tournament-based reward systems. The experiment also revealed that subjects' attitudes about the varying reward systems differed widely (van Dijk, Sonnemans, et al. 2001). (Straberg 2010) showed in an empirical study concerning the perceived impact of performance-related pay in Sweden, that men were much more likely to see the arrangement as fair and reasonable.

Tackling the problem of multi-dimensionality of many tasks, (Fehr and Schmidt 2004) conduct an experiment with university students to understand the effects of varying bonus schemes on effort provision on two distinct tasks, only one of which is contractible. They find that simple piece-rate contracts lead to a focusing on the contractible task, while bonus arrangements designed to be more encompassing and to explicitly address the multi-tasking problem also induce participants to spend time on the second task.

The issue of extrinsic versus intrinsic motivation and how performance pay can change the perception of salary arrangement is the subject of a laboratory experiment by (Gneezy and Rustichini 2000). They use high school and university students in Israel and offer them different size of bonuses for specific tasks. The results suggest that subjects showed higher levels of productivity when offered large rewards, but small awards led to worse performance than offering no monetary reward at all. This suggests the importance of framing of performance pay — if bonuses adequately communicate the importance of performing assigned tasks well compared to the overall goals of the organization, they can work, but if bonuses trigger a change in evaluation of the worker relationship, crowding-out of intrinsic motivation can worsen productivity.

An interesting laboratory experiment recruits future teachers in India to assess the possibility of gaming effects under performance pay. The experiment assesses teacher efforts when rewards are a function of average student test scores. In a situation with strong social heterogeneity and prejudice, teachers might focus on assisting high-status students, while neglecting lower caste pupils. The experiments reveals that poorly designed incentive plans lead to such a misallocation of teacher effort, which produces unequal

distribution of effort across student groups, but properly designed incentives can mitigate such behavior (Jain and Narayan 2011).

Public sector craft jobs

The evidence from experimental studies for public sector craft jobs is basically similar to that from observational studies – there is generally a positive impact but the limitations are also similar, in that the evidence is largely from OECD countries with the significant exception of teaching and health care where several significant experiments have been undertaken in developing countries.

Tax administration, job placement

(Burgess, Propper, et al. 2010) use a randomized controlled trial (RCT) to examine the impact of a pilot team-based incentive scheme introduced in 2002 in Her Majesty's Customs and Excise (HMCE), the indirect tax assessment and collection agency of the UK government. Each team consisted of a small number of tax offices, and ranged from 150 to 280 workers; there were two treatment teams that received two different types of incentive payments — one a bonus that was a fixed percentage of the officer's salary and the other a flat rate bonus — and a control team. The incentive scheme for both the treatment teams consisted of meeting a set of targets on revenue collection and conduct of audits, with an average bonus size of approximately 3% of annual salary. If the target was met then all staff in the team received the bonus. The authors' use detailed data from the HMCE's performance management system and personnel records to show that the tax yield increased for both the treatment teams relative to the control group, and that these increases were due to more time spent auditing that resulted in the recovery of greater tax revenue. The study also found that the strategies of the two incentivized teams were different, with the managers in team 2 allocating more incentive tasks to efficient workers than managers in team 1. Whether the flat rate bonus structure contributed to this task allocation was not examined.

(Burgess, Propper, et al. 2011) also did another RCT to examine the impact of a team-based incentive scheme introduced in a large UK public agency, Jobcentre Plus, which is tasked with placing the unemployed into jobs and administering welfare benefits. Each team consisted of a district, which had several offices, ranging from 250 to 1500 staff. There were 17 treatment districts and 73 control districts. The incentives were based on achieving both quantitative and qualitative targets — number of individuals placed in jobs, customer and employer service, and reducing benefit calculation error and fraud. The job placement target was weighted by the types of clients who found employment (e.g. highest weights to an unemployed single parent) with extra points if the person retained employment. While individuals worked in offices, the targets were set at the district level, which consisted of many offices that operated independently of each other, a potential flaw in the schemes design. The study was explicitly designed to assess the impact of incentives given multi-dimensionality of tasks, as well as possible free-riding given the nature of the team incentive.

The study's findings were that while overall there was little difference between the treatment group and the control group on job placement, in smaller teams (fewer offices per district, and smaller offices) the incentives resulted in 10% greater job placements than in the control group, with the effect declining for larger districts such that there was a significant negative impact on productivity for the largest quartile of districts. This suggests considerable free-riding behavior in larger teams. The authors also found that none of the quality measures were significant in the treatment group, irrespective of team size, suggesting that measurement problems were important.

Teaching

A number of field experiments have evaluated the impact of performance pay for teachers on reducing absenteeism and improving learning outcomes. The findings are generally mixed. In a study, (Duflo, Hanna, et al. 2010) show that random assignment for monitoring and financial incentives for teachers in rural India led to a strong reduction of teacher absenteeism and increased students' test scores by approximately 0.17 standard deviations. Units of observation for the study were single-teacher schools run by an NGO. The NGO selected 120 schools for testing the monitoring and incentive program, 60 which were randomly assigned to the treatment group. In the treatment groups, teachers had to use tamper-proof cameras to document their classroom presence at the start and end of each school day. Teacher salary was then made a function of the days in attendance, ranging from 50% to 130% of the pay in the fixed wage control group.

(Kremer and Chen 2001) by contrast show that subjective monitoring arrangements by an individual in the institutional hierarchy (like the headmaster of a school) may not work in developing country settings because the monitor might shirk, try to avoid confrontation, or collude with the workers. In Kenya, the Early Childhood Education Project offered substantial material incentives to teachers (bicycles) with good attendance as reported by their supervisor. Yet, the study found no effect of this program on absences as there was considerable cheating. In every school, the headmaster reported sufficient attendance for the teacher to receive the prize; however, when the research team independently verified absence through unannounced visits in both treatment and comparison schools, they found that the absence rate was actually exactly at the same high level in treatment and in comparison schools. This and Duflo's study suggest that impersonal, external monitoring by a camera coupled with a clear, credible, and automatic threat of punishment and promise of reward was the key design feature for program success.

Studies of performance pay linked to student outcomes are also similarly mixed. A field experiment in 50 Kenyan schools linking teacher salaries to student test scores failed to find lasting effects (Glewwe, Ilias, et al. 2010). Teacher attendance did not improve; teachers did not adjust their teaching methods or conduct more preparation sessions. Students in treated schools did perform better during the program duration, but these gains did not carry beyond the study period.

A field experiment conducted in NYC public schools also failed to find statistically significant effects of team incentives for teachers on student outcomes (Fryer 2011). In 2007 New York City launched a pilot program of financial incentives to teachers in 400 low-performing schools with the goal of improving student outcomes. There were explicit eligibility criteria for schools to be part of the pilot program; about half were randomly selected to receive treatment, which consisted of a school bonus of \$3,000 per staff member if certain standards are reached. The study did not find any effect of the financial incentives on teacher or student behavior. The surprising null finding is seen to be potentially produced by a lack of strong individual incentives, driven by the small size of the average bonus (4% of annual teacher salary), free-riding of teachers within schools and the overall complexity of the incentive scheme. A related study that also assesses the effects of the NYC group incentive program on classroom activities and teacher turnover and qualification, apart from test scores and teacher effort, similarly finds no effects (Goodman and Turner 2010). A three-year experimental evaluation of the Project on Incentives in Teaching (POINT) in Metropolitan Nashville schools also found no significant effects of bonus incentives on student test scores (Springer, Ballou, et al. 2010).

On the other hand a large-scale field experiment in a representative sample of 300 government-run rural primary schools in India found that bonus pay linked to the mean improvement of student test scores in an independent learning assessment led to a statistically significant and substantively meaningful

improvement of student outcomes (Muralidharan and Sundararaman 2009). In the treatment group scores were higher by 0.28 standard deviations in math tests and by 0.16 standard deviations in language tests, across “conceptual” and “mechanical” parts of the test. They also find positive spill-over effects to subjects not part of the official student assessment.

Health sector

A number of randomized-controlled trials have been implemented to determine the role of performance pay on health worker productivity, patient treatment and outcomes. Similar to studies on healthcare relying on observational data, the majority of studies assess these questions in the context of OECD health care systems. (Prentice, Burgess, et al. 2007) point out that improving quality rather than quantity of output is the primary focus for many, if not all, performance pay schemes implemented in the healthcare sector.

One of the first studies to employ an experimental design was work on performance incentives for nursing homes (Norton 1992). He finds that nursing homes assigned to the treatment group show better resident health outcomes and shorter stays.

(Kouides, Bennett, et al. 1998) implement a randomized-controlled trial, offering a randomly selected set of primary care physicians financial incentives based on influenza immunization rates of the elderly, as part of a Medicare demonstration project. Doctors in the treatment group were eligible to receive a \$0.80 payment per shot for an immunization rate of 70% and \$1.60 for each shot, if an immunization rate of 85% was attained. The experiment finds a difference of 7% in the immunization rate between treatment and control groups.

On the other hand, (Hillman, Ripley, et al. 1998; Hillman, Ripley, et al. 1999) use two RCT designs to incentivize cancer screenings for women of age 50 and above and pediatric immunizations, respectively. In both studies the authors document no significant difference between treatment and control groups. Similarly, a RCT implemented by (Grady, Lemkau, et al. 1997) finds no clear effects of financial incentives on mammography referrals by primary care physicians.

Contrastingly, a set of studies (Fairbrother, Hanson, et al. 1999; Fairbrother, Siegel, et al. 2001), also focusing on pediatric immunizations, finds that performance incentives increased immunizations rates by several percentage points compared to the control group. A randomized field trial at the clinic-level found that financial incentives improved treatment of smoking cessation outcomes (Roski et al. 2003). Work on performance pay for cognitive services interventions by pharmacists also finds positive effects (Christensen et al. 2000).

To our knowledge, the only two available randomized-controlled trials on performance pay in health care in a low income country are a study by Basinga et al. (2010) in Rwanda and a study by Singh (2010) in India. Basinga et al. use an RCT design to evaluate performance pay in Rwandan primary health care centers. The authors took advantage of a sequenced roll-out of the scheme across Rwandan health care facilities, collecting data on child preventive care and prenatal delivery. To isolate the performance-pay effect from a general increase in resources, comparison facilities received an equivalent increase in their budgets. The study uses information from 166 facilities and 2158 households. They find large effects on all central outcome measures, but with particularly striking effects for services with the highest payoffs and smallest necessary staff effort.

(Singh 2010) treated three groups of mothers and staff providing child care and nutritional advice to them in Chandigarh, India: in one group the workers received with performance pay; in a second group the workers had no performance pay but the women that they worked with were separately given factual information about nutrition; and the third group received both treatments. The study found that children's weights improved only in the third group compared to the control group.

It is noteworthy that nearly all studies on the health care sector so far focus on fairly narrow types of performance pay and specific, single outcome measures in preventative care, not necessarily overall multidimensional patient treatments and outcomes.

Private sector: Craft or coping jobs

The evidence from experimental studies for private sector craft or coping jobs is largely from OECD countries. In a field experiment from the private sector (Bandiera, Barankay, et al. 2006), some managers were treated with the introduction of a performance-pay system and productivity of lower-tier workers was used as an outcome measure. The study finds evidence of both an incentive and sorting effect, i.e. managers support their high-productivity workers and fire the least qualified employees. Evidence from a quasi-experiment in a private sector company found that the introduction of a new appraisal system that feeds into performance pay can improve trust in top-level management (Mayer and Davis 1999). An experimental treatment of monitoring efforts by management in a call center found that employees largely behave according to a rational-cheater model of human behavior, highlighting the importance of performance measurement, but at the same time a substantial portion of employees remains unaffected by monitoring attempts (Nagin, Rebitzer, et al. 2002).

5. Assessing the evidence

To assess the overall evidence, the 152 studies that were reviewed in this paper (see the Appendix for a list) were grouped into three categories: *positive* if their findings provide positive evidence for the effectiveness of incentive schemes;⁸ *neutral* if the study is largely descriptive or finds contradicting evidence; and *failed* if the evidence indicates no effect or negative effect of performance pay. Figure 1 shows the overall frequency of results. A majority of studies (93 out of the 152) presents supportive evidence for some form of effect of performance pay schemes, with experimental studies showing more positive findings than observational ones.

In drawing lessons however, it is important to distinguish the findings more systematically by the research quality of the concerned study. Study quality was ranked in two different ways. First, each study was assessed for its 'internal validity', or strength of the causal arguments being made, using a five-point ranking (from weak to strong) as follows:

1. no empirical study or faulty research design
2. descriptive; small sample size
3. secondary data analysis and/or descriptive data analysis; small sample size; some statistical analysis
4. quasi-experimental design; reasonable sample size; conclusions based on statistical analysis

⁸ Inevitably, there is some subjectivity in the classification of studies. Studies were rated as *positive* if there was general evidence on the basic functionality of incentive schemes, even if additional results qualify the effect, e.g. studies on crowding-out of intrinsic motivation generally still find positive effects of explicit incentives.

5. laboratory experiments; randomized controlled trial; large sample size; strong statistical analysis; strong conclusions

Second, studies were also evaluated on the dimension of “external validity”, or to what extent the causal connections drawn in the specific context of the study would remain valid if replicated in other contexts. So for example, lab experiments and RCTs offer very strong evidence about causality (high internal validity), but in a specific context — they tell us the average impact of a particular intervention in a particular location with a particular sample at a particular point in time. They are often accused of being low on external validity as the study subjects (usually college students in the case of laboratory experiments) are not representative of the general population, or in this case the population of interest (civil servants) and the requirements of the experiment imply very particular conditions that may not approximate real world settings.

Figure 1: Aggregate findings on performance-related pay

Figure 2 (left panel) shows the overall results by the measure of internal validity. The majority of high quality studies (score of 4 and above) show a positive finding, while a majority of the low quality study (score of 1 or 2) show a negative or neutral finding. When assessing the findings according to the external validity of studies (Figure 2 right panel) again more positive results are found in more externally valid studies.

Figure 2: Findings by internal and external validity

Parsing the evidence by job type, a majority of studies of craft and production jobs show positive results, while a majority of studies of coping jobs find negative or neutral results (

Figure 3 left panel). These limitations of PRP to the demanding contextual environments of coping jobs are in line with the theoretical arguments, though the total number of studies of coping jobs (16 in total in this review) are too few to allow generalizations with any degree of confidence. Analyzing studies of relevance to the public sector — i.e. craft or coping jobs only, or 110 studies in the review — by country context reveals that, counter-intuitively, the weight of the evidence is somewhat stronger for developing countries (

Figure 3 right panel). However, it must be emphasized that the number of studies for developing countries is low.

Figure 3: Findings by job type

When the quality of study (internal validity) is considered separately for studies of craft and coping jobs, again more rigorous research shows on average more positive findings.

Figure 4 illustrates the breakdown of results by study quality for relevant studies. The left panel shows that a majority of high quality studies (ranked 4 and 5 in the measure of internal validity) of craft and coping jobs show positive findings (46 out of 68) while a majority of low quality studies show negative or neutral results. However, the dearth of the evidence on coping jobs is even more apparent for high quality studies as the bulk of the literature has focused on craft jobs, in particular teaching and healthcare (right panel). The findings for craft jobs is generally more positive in these high quality studies in developing country contexts than in OECD contexts, although the number of such studies is very small and largely limited to health and education (Table 3).

6. Summary

Overall the body of evidence paints a supportive picture of performance pay in craft jobs within the public sector, but less so for coping jobs. So, as incentive theory would predict, PRP seemingly has a greater role to play in jobs where the outputs are more readily observable, such as teaching and health care jobs, than it does in more general administration. That these are jobs where the day to day actions of staff are unobservable does not seem to be an obstacle — apparently confounding, at least in the short term, the behavioral economics concern about crowding out intrinsic incentives. It is also in relation to

craft jobs that there have been more observational and experimental studies in developing countries — and generally the evidence from those settings is more positive than in OECD settings.

Among the observational studies, work analyzing the introduction of performance pay in the private sector finds nearly uniform support for the effectiveness of explicit incentive schemes. Observational studies of PRP for craft jobs in the public sector highlight that measurability of effort and output is more difficult in that environment and present slightly more mixed findings, but it seems the better output measurement and performance assessment are, the better performance pay works. The study of Brazilian

Figure 4: Findings for craft and coping tasks by research quality and country context

Table 3: Findings of high quality craft and coping studies by sector and country context

	Craft jobs				Coping jobs	
	Education	Health	Tax	Other	Public	Private
OECD	19	27	2	7	1	1
Positive	12	16		6	1	1
Negative or neutral	7	11	1	1	0	0
Developing country	6	4	1	0	0	0
Positive	4	4	1			
Negative or neutral	2	0	0			
Total	25	31	3	7	1	1

tax collectors is a good example of public sector work in which performance pay can be successfully implemented. Equally, Lavy's study of Israeli teacher competition (Lavy 2009) shows the ability of a well-designed performance scheme to function properly. At the same time, several observational studies identify problems with unintended consequences, generically subsumed under “gaming” the incentive scheme, that can run counter to the original intentions of the reforms. With the current state of evidence though it remains unclear whether incidents of gaming have a net negative effect in the presence of increased productivity. Furthermore, while explicit incentive schemes certainly increase the opportunity for gaming, standard civil service arrangements have their own unintended incentive effects, i.e. employees will engage in behavior that increases the chances of easy work assignments or promotions and it is simply unknown whether existing forms of gaming are worse than similar behavior under performance pay.

Moving to studies that attempt to fulfill the gold standard of experimental design, the evidence overall again speaks in favor of the potential utility of performance pay for craft jobs. Comparing various laboratory experiments, the results suggest that indeed explicit performance incentives can work, but the studies employ easily measurable performance indicators and use fairly unrepresentative subject pools.

Both concerns should caution policy makers against accepting the results independently of other research. On the other hand, similar results have been found across a varied set of experimental settings, test locations and subject pools and overall findings do resonate with the observational literature, improving overall credibility and external validity.

The strongest form of evidence comes from field experimental studies for craft jobs that neatly address concerns of internal and external validity. Here, evidence is somewhat more mixed. Several studies of teacher incentive programs found no or transient effects of bonus pay systems in the context of US schools but in the developing world, evidence has been more positive. The discrepancy between teacher incentives in the developed and developing world could on the one hand stem from the relative magnitude of incentives compared to normal salary or on the other hand come from higher marginal effects in the education production function in developing countries. Many factors enter the production of education, all of which are likely lacking in many developing country schools. Improving one input aspect, e.g. teacher presence and effort, could have conceivably larger marginal effects than the same input improvement in a developed country school.

However, three limitations of these research findings loom large. First the number of high quality studies overall is limited, and the number for developing countries is particularly low. This is particularly the case for coping jobs reflective of the work of the core public administration for which there is at this stage simply not enough evidence on whether or not PRP can work. This small pool of research studies makes it hard to disagree with the observation that: “much of the early evidence is not very robust. Consequently, a meta-analysis in which it is concluded that the “bulk” of the evidence shows this or that cannot be provided as — despite the strong positions taken by the proponents or opposers of attempts to incentivise the public sector — there is no bulk” (Prentice, Burgess, et al. 2007, pp.12-13).

Second, the studies inevitably tend to look at the impact of PRP in the short term. Is it possible that the positive results are the result of Hawthorne Effects or new incentives which degrade over time as attitudes towards work change? It is generally accepted that “(p)erformance related pay requires a long list of supportive local conditions before it stands a good chance of working as intended” (Pollitt and Dan 2011b, p.46) and it seems possible that the supportiveness of those conditions may only emerge over the longer term. In particular, to fully explore the persistent concern about crowding-out of intrinsic motivation, long term studies are necessary to assess how the design of incentive schemes and how they are communicated to employees minimizes the risk of employees changing attitudes toward their task.

Third, existing research exclusively focuses on the mechanical aspects of performance pay reforms, i.e. how to measure performance, how to design the incentive scheme and how large the bonus should be. No study explicitly considers contextual factors that can affect the operation of even the most well-designed incentive schemes. Most glaringly, the role of politicized bureaucracies has not been addressed properly. One, we need a better understanding of how politics can subvert flexible pay reforms and create important unintended side-effects and it is worth investigating how performance pay can potentially be used as a tool in de-politicizing the civil service and increase transparency and meritocracy. Interestingly, the move away from individualized pay in various area of the civil service, to uniform salary scales, was partly driven by the prevalence of patronage politics and clientelistic linkages that permeated the civil service in 19th century Western countries. Research could draw on this rich historical data and contemporary experience of civil service reform in developing countries to inform our understanding of performance pay. Apart from the influence of political parties or influential patrons, the role of civil service unions needs more attention. While most unions oppose in principle performance pay as a threat to union power, a better understanding of unions in the reform initiation and implementation could vastly improve future reform attempts. Countries vary dramatically in their wage bargaining structures and the strength of

unions, introducing performance pay in one scenario is likely to have vastly different effects than in others.

Further studies will need to broaden the coverage of developing countries and review the evidence over a longer time scale. If they continue to show a broadly positive association between PRP and performance in craft jobs, they will also need to shed light on a number of important design questions concerning individual versus group rewards, size of bonus, and types of performance measures used. They will also need to unpack the political process underlying reform initiation, implementation and sustained support for reforms — are the ideological or other pressures for PRP at odds with the long term fine-tuning which is undoubtedly necessary to maintain any positive results?

Pending that further research, the guidance for practitioners seems to be that PRP for craft jobs is a feasible possibility for performance improvement and has some evidence behind it — but that careful design and piloting, accompanied by a technical willingness and a political ability to change direction is key. For coping jobs however, there is at this stage simply not enough robust evidence to draw any conclusions either way.

This review has examined almost exclusively the effects of PRP on effort as that has been the focus of much of the theoretical and empirical literature. However, there is a small policy literature that also points to potential agency level and broader public sector wide effects that merit further more rigorous empirical examination. PRP's contribution to agency level productivity through the sorting channel — attracting higher quality workers who are likely to do better in this scheme — has already been discussed, and there is some fairly robust evidence of this in developed country contexts (Booth and Frank 1999; Bandiera, Barankay, et al. 2006; Cadsby, Song, et al. 2007). Individuals sorting into jobs with performance pay are on average higher educated, more qualified, less risk-averse and male. In a comprehensive laboratory experiment that features the self-selection of participants into performance schemes and also compares fixed wage, piece-rate and tournament-based rewards, researchers found clear incentive effects for variable payment schemes, driven by sorting of individuals. Individual-level characteristics like self-assessment, risk preferences, gender and social preferences systematically predict sorting decisions (Dohmen and Falk 2007).

In consequence, it is possible that PRP would have a clearer effect on individual incentives if the sorting effect had longer to work its way through the system. To evaluate this effect however, future studies need a longer time-frame to assess changes in recruitment and work-place behavior.

It is important to note that in OECD countries, performance-related pay has been introduced at a time of increasing “work intensification” through control measures such as the use of ICT to track individual production and working habits (Green 2001; Beynon, Grimshaw, et al. 2002). These are somewhat hard drivers of changes in what Marsden refers to as the “effort bargain” (Marsden 2004) — the implicit understanding between staff and management about “how hard we work around here”. PRP can offer an additional and less coercive point of entry into renegotiating this bargain. (Marsden 2004) cites two examples of changes in working practices that were sought in hospital management and in the tax service, both in the UK. In both cases, the task of implementing the changes fell to managers who were under pressure from their staff to be lenient with work assignments and generous with pay increases. In both cases, he argues, that individual incentives were only a modest part of the function of performance-related pay. Its real contribution was “to enable management to redefine the established performance norms in their organization, and then to obtain effective compliance with those norms, with the explicit or tacit agreement of as many employees as possible” (Marsden 2004, p.351). In sum, by incrementally ratcheting up the performance expectations through the many thousands of performance appraisal discussions, the informal agency working culture was changed.

PRP can also be considered as an element in overall cost containment as it provides the employer with the additional tactical option of proposing that pay increases should only be provided in the form of (less costly) enhanced performance bonuses (Marsden and French 1998). (French 2005) identifies the significant additional control that performance-related pay gave managers within the UK Inland Revenue in the early 1990s, allowing them to cease automatic cost-of-living allowances, largely restrict additional pay increases to the performance component of pay, combining the latter with a de facto forced curve for allocating performance-based rewards. (French 2005) also highlights how individual negotiations within a performance-related pay scheme in the UK allowed managers to sever the previous linkages between the level of complexity of the tasks required of staff and their grade. Under the performance-related pay scheme, targets were negotiated individually and not by grade.

Finally, at least in OECD countries, PRP has usually been accompanied by other changes in pay policy, of which the two noteworthy ones are differentiation - creating pay differences within and across government agencies for civil servants doing the same job, based on the need to attract and retain qualified staff for these jobs; and delegation of pay-setting authority away from a central civil service agency to ministries, agencies, or departments. The linkages between PRP and pay differentiation and delegation remain underexplored. For example, it is possible that delegation is reported to lead to PRP since managers have an incentive to look for mechanisms that maximize the results can be achieved at minimum cost.

Appendix A: List of empirical PRP studies reviewed

	Authors	Year	Type	Quality	Effect	Country context	Type of Job
1	Aboody et al	2010	observational	4	positive	OECD	Coping
2	Amundson et al	2003	observational	3	positive	OECD	Craft
3	Ariely et al.	2008	lab experiment	5	positive	Developing	Production
4	Asch	1990	observational	4	positive	OECD	Craft
5	Atkinson et al.	2004	observational	4	positive	OECD	Craft
6	Ballou	2001	observational	3	neutral	OECD	Craft
7	Bandiera, Barankay and Ras	2005	observational	4	positive	OECD	Production
8	Bandiera, Barankay and Ras	2006	field experiment	5	positive	OECD	Production
9	Basinga et al	2010	field experiment	5	positive	Developing	Craft
10	Beaulieu and Horrigan	2005	observational	4	positive	OECD	Craft
11	Beer and Cannon	2004	observational	2	failed	OECD	Production
12	Beer and Katz	2003	observational	3	positive	OECD	Production
13	Belfield and Marsden	2003	observational	4	positive	OECD	Production
14	Bender and Elliott	2003	observational	4	positive	OECD	Unclassified
15	Bertelli	2006	observational	4	neutral	OECD	Craft
16	Blasi et al	2008	observational	4	positive	OECD	Craft
17	Booth and Frank	1999	observational	4	positive	OECD	Production
18	Brudney and Condrey	1993	observational	3	positive	OECD	Coping
19	Burgess et al	2010	field experiment	5	positive	OECD	Craft
20	Burgess et al	2004	observational	4	positive	OECD	Craft
21	Burks, Carpenter and Goett	2009	field experiment	5	positive	OECD	Production
22	Cadsby, Song and Tapon	2007	lab experiment	5	positive	OECD	Production
23	Camerer et al.	1997	observational	4	neutral	OECD	Production
24	Campbell et al	2007	observational	4	positive	OECD	Craft
25	Campbell et al	2005	observational	3	positive	OECD	Craft
26	Cardona	2007	observational	1	failed	OECD	Unclassified
27	Casalino et al	2003	observational	4	positive	OECD	Craft
28	Chalkley et al	2010	observational	4	positive	OECD	Craft
29	Christensen et al	2000	field experiment	5	positive	OECD	Craft
30	Chung et al	2007	observational	3	failed	OECD	Craft
31	Clarketal	1995	observational	4	neutral	OECD	Craft
32	Clotfelter et al.	2004	observational	4	failed	OECD	Craft
33	Clotfelter et al.	2008	observational	4	positive	OECD	Craft
34	Clotfelter, Ladd and Vigdor	2007	observational	4	positive	OECD	Craft
35	Cohn	1996	observational	2	neutral	OECD	Craft
36	Coleman et al.	2007	observational	4	positive	OECD	Craft
37	Condrey and Brudney	1992	observational	3	neutral	OECD	Coping
38	Cooper and Cohn	1997	observational	4	positive	OECD	Craft
39	Courty and Marschke	2003	observational	4	positive	OECD	Craft
40	Daley	1993	observational	3	positive	OECD	Coping
41	Dee and Keys	2004	observational	4	positive	OECD	Craft
42	Dohmen and Falk	2007	lab experiment	5	positive	OECD	Production
43	Doran et al.	2006	observational	1	positive	OECD	Craft
44	Dowling and Richardson	1997	observational	4	positive	OECD	Coping
45	Duflo, Hanna and Ryan	2007	field experiment	5	positive	Developing	Craft
46	Eberts, Hollenbeck and Stor	2002	observational	4	failed	OECD	Craft
47	Eichler and Levine	2009	observational	4	positive	Developing	Craft

	Authors	Year	Type	Quality	Effect	Country context	Type of Job
48	Eichler, Auxila and Pollock	2001	observational	2	positive	Developing	Craft
49	Eriksson and Villevall	2004	lab experiment	5	positive	OECD	Production
50	Fairbrother et al	1999	field experiment	5	positive	OECD	Craft
51	Fairbrother et al	2001	field experiment	5	positive	OECD	Craft
52	Fehr and Goette	2007	field experiment	5	positive	OECD	Production
53	Fehr and Schmidt	2004	lab experiment	5	positive	OECD	Production
54	Felt-Lisk et al	2007	observational	4	neutral	OECD	Craft
55	Figlio and Kenny	2007	observational	4	positive	OECD	Craft
56	Figlio and Winicki	2005	observational	4	neutral	OECD	Craft
57	Fryer	2011	field experiment	5	failed	OECD	Craft
58	Gavaghan et al	2010	observational	4	failed	OECD	Craft
59	Gerhart and Milkovich	1990	observational	4	positive	OECD	Production
60	Glewwe, Ilias and Kremer	2010	field experiment	5	neutral	Developing	Craft
61	Glickman et al	2007	observational	4	neutral	OECD	Craft
62	Gneezy and Rustichini	2000	lab experiment	5	positive	OECD	Production
63	Goldhaber and Brewer	2000	observational	4	positive	OECD	Craft
64	Goodman and Turner	2010	field experiment	5	failed	OECD	Craft
65	Grady et al	1997	field experiment	5	failed	OECD	Craft
66	Gratz	2009	observational	2	neutral	OECD	Craft
67	Grimshaw	1998	observational	1	failed	OECD	Unclassified
68	Hamilton, Nickerson and O	2003	observational	4	positive	OECD	Craft
69	Heckman, Heinrich and Smi	1997	observational	4	neutral	OECD	Production
70	Heneman and Milanowski	1999	observational	3	neutral	OECD	Craft
71	Hillman et al	1998	field experiment	5	failed	OECD	Craft
72	Hillman et al	1999	field experiment	5	failed	OECD	Craft
73	Hillman et al	1991	observational	3	neutral	OECD	Craft
74	Hochberg and Lindsey	2010	observational	4	positive	OECD	Craft
75	Ingraham	1993	observational	1	failed	OECD	Unclassified
76	Jack	2003	observational	2	positive	Developing	Craft
77	Jacob	2005	observational	4	positive	OECD	Craft
78	Jacob and Levitt	2003	observational	4	neutral	OECD	Craft
79	Jain and Narayan	2011	lab experiment	5	positive	Developing	Craft
80	Kahn, Silva and Ziliak	2001	observational	4	positive	Developing	Craft
81	Kelley	1999	observational	2	neutral	OECD	Craft
82	Kellough and Nigro	2002	observational	3	failed	OECD	Unclassified
83	Kellough and Selden	1997	observational	3	neutral	OECD	Unclassified
84	Kerr	1975	observational	1	failed	OECD	Unclassified
85	Ketelaar, Manning and Turk	2006	observational	3	neutral	OECD	Unclassified
86	Kim	2002	observational	2	neutral	OECD	Unclassified
87	Kingdon and Teal	2008	observational	4	positive	Developing	Craft
88	Kiragu and Mukandala	2003	observational	2	neutral	Developing	Unclassified
89	Knez and Simester	2001	observational	4	positive	OECD	Production
90	Koretz	2002	observational	1	failed	OECD	Craft
91	Kremer and Chen	2001	field experiment	5	failed	Developing	Craft
92	Kouides et al	1998	field experiment	5	positive	OECD	Craft

	Authors	Year	Type	Quality	Effect	Country context	Type of Job
93	Ladd	1999	observational	4	positive	OECD	Craft
94	Lavy	2008	observational	4	positive	OECD	Craft
95	Lavy	2009	observational	4	positive	OECD	Craft
96	Lazear	2000	observational	4	positive	OECD	Production
97	Lazear	2003	observational	1	positive	OECD	Production
98	Le Grand	2007	observational	2	positive	OECD	Unclassified
99	Leford, Lawler and Mohrma	1995	observational	1	neutral	OECD	Production
100	Levin-Scherz et al.	2006	observational	4	positive	OECD	Craft
101	Li et al	2011	observational	4	neutral	OECD	Craft
102	Lindenauer et al	2007	observational	4	positive	OECD	Craft
103	Mandel and Kortagal	2007	observational	3	positive	OECD	Craft
104	Marsden	2004	observational	3	neutral	OECD	Unclassified
105	Marsden	2009	observational	1	neutral	OECD	Unclassified
106	Marsden and French	1998	observational	3	neutral	OECD	Unclassified
107	Mayer and Davis	1999	field experiment	5	positive	OECD	Production
108	McMenamin et al	2003	observational	4	positive	OECD	Craft
109	McNamara	2005	observational	2	neutral	Developing	Craft
110	Meessen et al	2006	observational	4	positive	Developing	Craft
111	Meessen, Kashala and Musi	2008	observational	3	positive	Developing	Craft
112	Milkovich and Wigdor	1991	observational	2	positive	OECD	Production
113	Muralidharan and Sundarar	2009	field experiment	5	positive	Developing	Craft
114	Murnane and Cohen	1986	observational	1	failed	OECD	Craft
115	Nagin et al.	2002	field experiment	5	neutral	OECD	Production
116	Norton	1992	field experiment	5	positive	OECD	Craft
117	Odden and Kelley	2002	observational	2	positive	OECD	Craft
118	OECD	1996	observational	2	neutral	OECD	Unclassified
119	OECD	1997	observational	2	neutral	OECD	Unclassified
120	OECD	2005	observational	2	neutral	OECD	Unclassified
121	OECD	2008	observational	2	neutral	OECD	Unclassified
122	Paarsch and Shearer	1999	observational	4	positive	OECD	Unclassified
123	Pearson et al	2008	observational	4	failed	OECD	Craft
124	Pires	2007	observational	2	failed	Developing	Unclassified
125	Pourat et al	2005	observational	4	positive	OECD	Craft
126	Reid	1992	observational	2	positive	Developing	Unclassified
127	Rexed et al.	2007	observational	2	neutral	OECD	Unclassified
128	Rogers and Vegas	2009	observational	1	positive	Developing	Craft
129	Rosenthal et al	2005	observational	4	neutral	OECD	Craft
130	Roski et al	2003	field experiment	5	positive	OECD	Craft
131	Safran et al	2000	observational	4	positive	OECD	Craft
132	Schick	1998	observational	1	failed	OECD	Unclassified
133	Shaw, Gupta and Delery	2002	observational	4	positive	OECD	Production
134	Shearer	2004	field experiment	5	positive	OECD	Production
135	Shen	2003	observational	4	failed	OECD	Craft
136	Singh	2010	field experiment	5	positive	Developing	Craft
137	Soeters and Griffiths	2003	observational	2	positive	Developing	Craft
138	Soeters et al	2006	observational	3	positive	Developing	Craft

	Authors	Year	Type	Quality	Effect	Country context	Type of Job
139	Springer et al.	2010	field experiment	5	failed	OECD	Craft
140	Stajkovic and Luthans	2001	field experiment	5	positive	OECD	Production
141	Stazyk	2010	observational	4	positive	OECD	Unclassified
142	Steel et al	2007	observational	3	positive	OECD	Craft
143	Straberg	2010	observational	3	neutral	OECD	Craft
144	Streib and Nigro	1993	observational	3	neutral	OECD	Unclassified
145	Vaghela et al.	2009	observational	3	positive	OECD	Craft
146	van Dijk, Sonnemans and va	2001	lab experiment	5	positive	OECD	Production
147	Vegas	2005	observational	3	positive	Developing	Craft
148	Vegas and Umansky	2005	observational	3	positive	Developing	Craft
149	Willems, Janvier and Hende	2006	observational	2	neutral	OECD	Unclassified
150	Witter et al	2011	observational	3	neutral	Developing	Craft
151	Woessmann	2010	observational	4	positive	OECD	Craft
152	World Bank	2001	observational	3	positive	Developing	Craft
153	Young et al	2007	observational	4	positive	OECD	Craft

Appendix B: List of High Quality Studies of Craft and Coping Jobs

EDUCATION

Country Context: OCED

Type of study: Observational

Findings		
<i>Positive</i>	<i>Neutral</i>	<i>Negative</i>
1. Atkinson et al. (2004)	1. Figlio & Winicki (2005)	1. Clotfelter et al. (2004)
2. Clotfelter et al. (2008)	2. Jacob & Levitt (2003)	2. Eberts et al. (2002)
3. Clotfelter, Ladd, & Vigdor (2007)		
4. Cooper & Cohn (1997)		
5. Dee & Keys (2004)		
6. Figlio & Kenny (2007)		
7. Goldhaber & Brewer (2000)		
8. Jacob (2005)		
9. Ladd (1999)		
10. Lavy (2008)		
11. Lavy (2009)		
12. Woessman (2010)		

Country Context: OCED

Type of study: Experimental

Findings		
<i>Positive</i>	<i>Neutral</i>	<i>Negative</i>
No studies	No studies	1. Fryer (2011)
		2. Goodman and Turner (2010)
		3. Springer et al (2010)

Country Context: Developing

Type of study: Observational

Findings		
<i>Positive</i>	<i>Neutral</i>	<i>Negative</i>
1. Kingdon & Teal (2008)	No studies	No studies

Country Context: Developing

Type of study: Experimental

Findings		
<i>Positive</i>	<i>Neutral</i>	<i>Negative</i>
1. Duflo et al (2007)	1. Glewwe et al (2010)	1. Kremer & Chen (2001)
2. Muralidharan & Sundararaman (2009)		
3. Jain & Narayan (2011)		

HEALTH

Country Context: OCED

Type of study: Observational

Findings		
<i>Positive</i>	<i>Neutral</i>	<i>Negative</i>
1. Beaulieu & Horrigan (2005)	1. Clark et al (1995)	1. Gavaghan et al. (2010)
2. Campbell et al. (2007)	2. Felt-Lisk et al (2007)	2. Pearson et al. (2008)
3. Casalino et al (2003)	3. Glickman et al (2007)	3. Shen (2003)
4. Chalkley et al (2010)	4. Li et al (2011)	
5. Coleman et al (2007)	5. Rosenthal et al (2005)	
6. Levin-Scherz et al (2006)		
7. Lindenauer et al (2007)		
8. McMenamin et al (2003)		
9. Pourat et al (2005)		
10. Safran et al (2000)		
11. Young et al (2007)		

Country Context: OCED

Type of study: Experimental

Findings		
<i>Positive</i>	<i>Neutral</i>	<i>Negative</i>
1. Fairbrother et al (2001)	No studies	1. Grady et al (1997)
2. Fairbrother et al (1999)		2. Hillman et al (1998)
3. Kouides et al (1998)		3. Hillman et al (1999)
4. Norton (1992)		
5. Roski et al (2003)		

Country Context: Developing

Type of study: Observational

Findings		
<i>Positive</i>	<i>Neutral</i>	<i>Negative</i>
6. Eichler & Levine (2009)	No studies	No studies
7. Meesen et al (2006)		

Country Context: Developing

Type of study: Experimental

Findings		
<i>Positive</i>	<i>Neutral</i>	<i>Negative</i>
8. Basinga et al (2010)	No studies	No studies
9. Singh (2010)		

REVENUE ADMINISTRATION

Country Context: OECD

Type of study: Experimental or observational

Findings		
<i>Positive</i>	<i>Neutral</i>	<i>Negative</i>
1. Burgess et al (2010)	1. Bertelli (2006)	No studies

Country Context: Developing

Type of study: Experimental or observational

Findings		
<i>Positive</i>	<i>Neutral</i>	<i>Negative</i>
1. Kahn et al (2001)	No studies	No studies

OTHER (Job placement, recruitment, private sector)

Country Context: OECD

Type of study: Experimental or observational

Findings		
<i>Positive</i>	<i>Neutral</i>	<i>Negative</i>
1. Asch (1990)	1. Heckman et al (1997)	No studies
2. Burgess et al (2004)	2.	
3. Courty & Marschke (2003)	3.	
4. Christensen et al (2000)	4.	
5. Blasi et al (2008)	5.	
6. Hochberg & Lindsey (2010)	6.	

COPING JOBS (Health managers, private sector)

Country Context: OECD

Type of study: Experimental or observational

Findings		
<i>Positive</i>	<i>Neutral</i>	<i>Negative</i>
1. Aboody et al (2010)	No studies	No studies
2. Dowling & Richardson (1997)		

References

- Aboddy, D., N. Johnson and R. Kasznik (2007), 'Employee Stock Options and Future Firm Performance: Evidence from Option Repricings', Journal of Accounting and Economics, 50, 74-92.
- Ahmad, S. and R. G. Schroeder (2003), 'The Impact of Human Resource Management Practices on Operational Performance: Recognizing Country and Industry Differences', Journal of Operations Management (21), 19-43.
- Amundson, G., L. I. Solberg, M. Reed, E. M. Martini and R. Carlson (2003), 'Paying for Quality Improvement: Compliance with Tobacco Cessation Guidelines', Joint Commission Journal on Quality and Safety, 29 (2), 59-65.
- Andrews, M. (2008a). Are One-Best-Way Models of Effective Government Suitable for Developing Countries? Harvard Kennedy School, Cambridge, Mass.
- Andrews, M. (2008b). Good Government Means Different Things in Different Countries Kennedy School of Government, Cambridge, MA <http://web.hks.harvard.edu/publications/getFile.aspx?Id=324>.
- Antwi, J. and D. Phillips (2011). Wages and Health Worker Retention: Evidence from Public Sector Wage Reforms in Ghana. World Bank, Washington DC.
- Ariely, D., U. Gneezy, G. Lowenstein and N. Mazar (2009), 'Large Stakes and Big Mistakes', Review of Economic Studies (76), 541-469.
- Arrowsmith, J. and P. Marginson (2008). Wage Flexibility. European Foundation for the Improvement of Living and Working Conditions, Dublin.
- Asch, B. J. (1990). Navy Recruiter Productivity and the Freeman Plan. RAND Corporation Santa Monica, CA
- Atkinson, A., S. Burgess, B. Croxson, P. Gregg, C. Propper, H. Slater and D. Wilson (2004). Evaluating the Impact of Performance-Related Pay for Teachers in England (Working Paper No.04/113). Centre for Market and Public Organisation, Bristol UK.
- Atkinson, A. B. (1999). Is Rising Inequality Inevitable? A Critique of the Transatlantic Consensus. UNU World Institute for Development Economics Research, Helsinki.
- Atkinson, J. and N. Meager (1986), Changing Patterns of Work: How Companies Achieve Flexibility to Achieve New Needs, London, National Economic Development Organisation.
- Balfour, D. and B. Wechsler (1996), 'Organisational Commitment: Antecedents and Outcomes in Public Organisations', Public Productivity and Management Review, 29, 256-277.
- Bandiera, O., I. Barankay and I. Rasul (2005), 'Social Preferences and the Response to Incentives: Evidence from Personnel Data', Quarterly Journal of Economics, 120 (3), 917-962.
- Bandiera, O., I. Barankay and I. Rasul (2006). Incentives for Managers and Inequality among Workers: Evidence from a Firm Level Experiment (Discussion Paper No. 2062). Institute for the Study of Labor Bonn, Germany.
- Barber, L., S. Hayday and S. Bevan (1999). From People to Profits (Ies Report 355). Institute for Employment Studies, London.
- Bardhan, P. (2002), 'Decentralization of Governance and Development', Journal of Economic Perspectives 16 (4), 185-205.
- Bardhan, P. and D. Mookherjee (eds) (2006), Decentralization and Local Governance in Developing Countries. A Comparative Perspective, Cambridge Mass., MIT Press.
- Barlevy, G. and D. Neal (2011). Pay for Percentile (Working Paper 17194). National Bureau for Economic Research, Cambridge, Mass.
- Beer, M. and M. D. Cannon (2004), 'Promise and Peril in Implementing Pay-for-Performance', Human Resource Management, 43 (1), 3-48.
- Beer, M. and N. Katz (2003), 'Do Incentives Work? The Perceptions of a Worldwide Sample of Senior Executives', People and Strategy, 26 (3), 30-44.

- Belfield, R. and D. Marsden (2003), 'Performance Pay, Monitoring Environments, and Establishment Performance', International Journal of Manpower, 24 (4), 452-489.
- Benabou, R. and J. Tirole (2003), 'Intrinsic and Extrinsic Motivation', Review of Economic Studies (70), 489-520.
- Benabou, R. and J. Tirole (2006), 'Incentives and Prosocial Behavior', American Economic Review, 96 (5), 1652-1678.
- Bender, K. A. and R. F. Elliott (1997), 'Decentralization and Pay Reform in Central Government: A Study of Three Countries', British Journal of Industrial Relations, 35 (3), 447-475.
- Bender, K. A. and R. F. Elliott (2003), Decentralised Pay Setting. A Study of the Outcomes of Collective Bargaining Reform in the Civil Service in Australia, Sweden and the UK, Farnham, UK, Ashgate.
- Bertelli, A. M. (2006), 'Motivation Crowding and the Federal Civil Servant: Evidence from the U.S. Internal Revenue Service', International Public Management Journal, 9 (23).
- Besley, T. and M. Ghatak (2004). Competition and Incentives with Motivated Agents (Working Paper). London School of Economics and Political Science, London.
- Bevan, G., K. Sisson and P. Way (1981), 'Cash Limits and Public Sector Pay', Public Administration, 59, 379-398.
- Beynon, H., D. Grimshaw, J. Rubery and K. Ward (2002), Managing Employment Change: The New Realities of Work, Oxford, Oxford University Press.
- Blasi, J. R., R. B. Freeman, C. Mackin and D. L. Kruse (2008). Creating a Bigger Pie? The Effects of Employee Ownership, Profit Sharing, and Stock Options on Workplace Performance. NBER, Cambridge, Mass.
- Booth, A. L. and J. Frank (1999), 'Earnings, Productivity, and Performance-Related Pay', Journal of Labor Economics, 17 (3), 447-463.
- Brown, W., S. Deakin, M. Hudson, C. Pratten and P. Ryan (1998). The Individualisation of Employment Contracts in Britain. Centre for Business Research, University of Cambridge, Cambridge, UK.
- Brown, W. and D. Marsden (2010). Individualisation and Growing Diversity of Employment Relationships London School of Economics and Political Science, London.
- Brudney, J. L. and S. E. Condrey (1993), 'Pay for Performance: Explaining Differences in Managerial Motivation', Public Productivity & Management Review, 17 (2), 129-144.
- Bruns, B., D. Filmer and H. A. Patrinos (2011), Making Schools Work: New Evidence on Accountability Reforms, Washington DC, World Bank.
- Burgess, S. and P. Metcalfe (1999). Incentives in Organisations: A Selective Overview of the Literature with Application to the Public Sector. University of Bristol, CMPO and CEPR, Bristol.
- Burgess, S., C. Propper, M. Ratto, S. Scholder and E. Tominey (2010), 'Smarter Task Assignment or Greater Effort: What Makes the Difference on Team Performance?', Economic Journal, 120 (547), 968-989.
- Burgess, S., C. Propper, M. Ratto and E. Tominey (2011). Incentives in the Public Sector: Evidence from a Government Agency (Working Paper No. 11/265). Center for Market and Public Organization, Bristol UK.
- Burgess, S. and M. Ratto (2003), The Role of Incentives in the Public Sector: Issues and Evidence (Working Paper), Bristol, UK, Centre for Market and Public Organisation.
- C., E. and P. N (2009), 'Performance-Based Payment: Some Reflections on the Discourse, Evidence and Unanswered Questions', Health Policy and Planning, 1 (7).
- Cadsby, C. B., F. Song and F. Tapon (2007), 'Sorting and Incentive Effects of Pay-for-Performance: An Experimental Investigation', Academy of Management Journal, 50 (2), 387-405.
- Camerer, C., L. Babcock, G. Loewenstein and R. Thaler (1997), 'Labor Supply of New York City Cabdrivers: One Day at a Time', Quarterly Journal of Economics and Philosophy, 111, 407-441.
- Campbell, S. M., D. Reeves, E. Kontopantelis, E. Middleton, B. Sibbald and M. Roland (2007), 'Quality of Primary Care in England with the Introduction of Pay for Performance', New England Journal of Medicine 357, 181-190.

- Campbell, S. M., M. Roland, E. Middleton and D. Reeves (2005), 'Improvements in the Quality of Clinical Care in English General Practice: Longitudinal Observational Study', British Medical Journal, 331 (1121-3).
- Cardona, F. (2007). Performance-Related Pay in the Public Service in OECD and EU Member States. OECD SIGMA, Paris.
- Casalino, L., R. Gillies, S. Shortell, J. Schmittiel, T. Bodenheimer, J. Robinson, T. Rundall, N. Oswald, H. Schauffler and M. Wang (2003), 'External Incentives, Information Technology, and Organized Processes to Improve Health Care Quality for Patients with Chronic Diseases', Journal of the American Medical Association, 289 (4), 434-441.
- Castaño, R., R. Bitran and U. Giedion (2004). Monitoring and Evaluating Hospital Autonomization and Its Effects on Priority Health Services Abt Associates, Bethesda, MD.
- Chalkley, M., C. Tilley, L. Young, D. Bonetti and J. Clarkson (2010), 'Incentives for Dentists in Public Service: Evidence from a Natural Experiment', Journal of Public Administration Research and Theory, 207-223.
- Chaudhury, N., J. Hammer, M. Kremer, K. Muralidharan and F. H. Rogers (2006), 'Missing in Action: Teacher and Health Worker Absence in Developing Countries', Journal of Economic Perspectives, 20 (1), 91-116.
- Chirkov, V. I., R. M. Ryan, Y. Kim and U. Kaplan (2003), 'Differentiating Autonomy from Individualism and Independence: A Self-Determination Theory Perspective on Internalization of Cultural Orientations and Well-Being', Journal of Personality and Social Psychology, 84, 97-110.
- Chomitz, K. M., G. Setiadi, A. Azwar, N. Ismail and Widiyarti (1997). What Do Doctors Want?: Two Empirical Estimates of Indonesian Physicians' Preferences Regarding Service in Rural and Remote Areas. World Bank, Washington DC.
- Christensen, T. and P. Lægreid (eds) (2011), The Ashgate Research Companion to New Public Management, Farnham, UK, Ashgate.
- Chu, K.-y. (ed.) (1991), Public Expenditure Handbook : A Guide to Public Expenditure Policy Issues in Developing Countries, Washington DC, IMF.
- Chung, S., L. P. Palaniappan, L. M. Trujillo, H. R. Rubin and H. S. Luft (2010), 'Effect of Physician-Specific Pay-for-Performance Incentives in a Large Group Practice', American Journal of Managed Care, 16 (2), 35-42.
- CIPD (2006). Working Life: Employee Attitudes and Engagement. Chartered Institute of Personnel and Development, London.
- Clotfelter, C., R. A. Diaz, H. Ladd and J. Vigdor (2004), 'Do School Accountability Systems Make It More Difficult for Low-Performing Schools to Attract and Retain High-Quality Teachers?', Journal of Policy Analysis and Management, 23 (2), 251-271.
- Clotfelter, C., E. Glennie, H. Ladd and J. Vigdor (2007). How and Why Do Teacher Credentials Matter for Student Achievement? (Working Paper 2). National Center for Analysis of Longitudinal Data in Educational Research, Washington DC.
- Clotfelter, C., E. Glennie, H. Ladd and J. Vigdor (2008), 'Would Higher Salaries Keep Teachers in High-Poverty Schools? Evidence from a Policy Intervention in North Carolina', Journal of Public Economics, 92, 1352-1370.
- Cohen, A. (1991), 'Career Stage as a Moderator of the Relationship between Organisational Commitment and Its Outcomes: A Meta-Analysis', Journal of Occupational Psychology, 64, 253-268.
- Cohen, A. (1993), 'Age and Tenure in Relation to Organisational Commitment: A Meta-Analysis', Basic and Applied Social Psychology, 14, 143-159.
- Coleman, K., K. L. Reiter and D. Fulwiler (2007), 'The Impact of Pay-for-Performance on Diabetes Care in a Large Network of Community Health Centers', Journal of Health Care for the Poor and Underserved, 18 (4), 966-983.
- Common, R. (1998), 'Convergence and Transfer: A Review of the Globalisation of New Public Management', International Journal of Public Sector Management, 11 (6), 440-448.

- Condley, S., R. Clark and H. Stolovitch (2003), 'The Effects of Incentives on Workplace Performance: A Meta-Analytic Review of Research Studies', Performance Improvement Quarterly, 16 (3), 46-63.
- Condrey, S. E. and J. L. Brudney (1992), 'Performance-Based Managerial Pay in the Federal Government: Does Agency Matter?', Journal of Public Administration Research, 2 (2), 157-174.
- Cooper, S. T. and E. Cohn (1997), 'Estimation of a Frontier Production Function for the South Carolina Educational Process', Economics of Education Review, 16 (3), 313-327.
- Courty, P., C. Heinrich and G. Marschke (2005), 'Setting the Standard in Performance Measurement Systems', International Public Management Journal, 8 (3), 1-27.
- Courty, P. and G. Marschke (2003), 'Dynamics of Performance-Measurement Systems', Oxford Review of Economic Policy, 19 (2), 268-284.
- Courty, P. and G. Marschke (2004), 'An Empirical Investigation of Gaming Responses to Explicit Performance Incentives', Journal of Labor Economics, 22 (1), 23-56.
- Cutler, T. and B. Waine (2005), 'Incentivizing the Poor Relation: 'Performance' and the Pay of Public Sector 'Senior Managers'', Competition and Change, 9 (1), 75-87.
- D., B. N. and D. R. Horrigan (2005), 'Putting Smart Money to Work for Quality Improvement', Health Services Research, 40 (5), 1318-1334.
- Dahlstrom, C. and V. Lapuente (2009), 'Explaining Cross-Country Differences in Performance-Related Pay in the Public Sector', Journal of Public Administration Research and Theory, 20, 577-600.
- Dee, T. S. and B. J. Keys (2004), 'Does Merit Pay Reward Good Teachers? Evidence from a Randomized Experiment', Journal of Policy Analysis and Management 23 (3), 471-488.
- Delfgaauw, J. and R. Dur (2008), 'Incentives and Worker's Motivation in the Public Sector', The Economic Journal, 118, 171-191.
- Dell'Aringa, C. and N. Lanfranchi (1999), 'Pay Determination in the Public Service: An International Comparison', in R. Elliott and C. Lucifora (eds.) Public Sector Pay Determination in the European Union, Basingstoke, MacMillan Press, pp 29-70.
- Dell'Aringa, C., C. Lucifora and F. Origo (2007), 'Public Sector Pay and Regional Competitiveness: A First Look at Regional Public-Private Wage Differentials in Italy', The Manchester School, 75 (4), 445-478.
- Dewatripont, M., I. Jewitt and J. Tirole (1999a), 'The Economics of Career Concerns, Part 1: Comparing Information Structures', Review of Economic Studies, 66, 183-198.
- Dewatripont, M., I. Jewitt and J. Tirole (1999b), 'The Economics of Career Concerns, Part 2: Application to Missions and Accountability of Government Agencies', Review of Economic Studies, 66, 199-217.
- Dickens, W. T. and L. F. Katz (1987). Interindustry Wage Differences and Industry Characteristics (NBER Working Paper No. W2014). National Bureau of Economic Research, Washington DC.
- Dixit, A. (1999), 'Incentives and Organization in the Public Sector. An Interpretative Review', The Journal of Human Resources, 34 (4), 696-727.
- Dixit, A. (2002), 'Incentives and Organizations in the Public Sector: An Interpretative Review', Journal of Human Resources, 37 (4), 696-727.
- Dohmen, T. and A. Falk (2007). Performance Pay and Multi-Dimensional Sorting - Productivity, Preferences and Gender (Working Paper). Institute for the Study of Labor, Bonn, Germany.
- Doran, T., C. Fullwood, H. Gravelle, D. Reeves, E. Kontopantelis, U. Hiroeh and M. Roland (2006), 'Pay-for-Performance Programs in Family Practices in the United Kingdom', New England Journal of Medicine, 355, 375-384.
- Dowling, B. and R. Richardson (1997), 'Evaluating Performance-Related Pay for Managers in the National Health Service', The International Journal of Human Resource Management, 8 (3), 348-366.
- Duflo, E., R. Hanna and S. P. Ryany (2010). Incentives Work: Getting Teachers to Come to School. MIT (Department of Economics and J-PAL) and the Kennedy School of Government, Cambridge, Mass.

- Eberts, R., K. Hollenbeck and J. Stone (2002), 'Teacher Performance Incentives and Student Outcomes', Journal of Human Resources, 37 (4), 913-927.
- Eichler, R., P. Auxila and J. Pollock (2001), 'Performance-Based Payment to Improve the Impact of Health Services: Evidence from Haiti', World Bank Institute Online Journal (April 2001).
- Eichler, R. and R. Levine (2009). Performance Incentives for Global Health: Potential and Pitfalls. Center for Global Development, Washington DC.
- Eyck, K. V. (2003). Flexibilizing Employment: An Overview ILO, Geneva.
- Fairbrother, G., K. L. Hanson, S. Friedman and G. C. Butts (1999), 'The Impact of Physician Bonuses, Enhanced Fees, and Feedback on Childhood Immunization Coverage Rates', American Journal of Public Health, 89 (2), 171-175.
- Fairbrother, G., M. J. Siegel, S. Friedman, P. D. Kory and G. C. Butts (2001), 'Impact of Financial Incentives on Immunization Rates in the Inner City: Results of a Randomized Controlled Trial', Ambulatory Pediatrics, 1 (4), 206-212.
- Farnham, D. and S. Horton (2000), 'The Flexibility Debate', in D. Farnham and S. Horton (eds.) Human Resources Flexibilities in the Public Services, London, MacMillan Press.
- Fehr, E. and K. M. Schmidt (2004), 'Fairness and Incentives in a Multi-Task Principal-Agent Model', Scandinavian Journal of Economics, 106 (3), 453-474.
- Felt-Lisk, S., G. Gimm and S. Peterson (2007), 'Making Pay-for-Performance Work in Medicaid', Health Affairs, 26 (4), 516-527.
- Fields, G. S. and H. J. Wan (1989), 'Wage-Setting Institutions and Economic Growth', World Development, 17 (9), 1471-1483.
- Figlio, D. N. and L. W. Kenny (2007), 'Individual Teacher Incentives and Student Performance', Journal of Public Economics, 91, 901-914.
- Figlio, D. N. and J. Winicki (2005), 'Food for Thought: The Effects of School Accountability Plans on School Nutrition', Journal of Public Economics, 89 (381-394).
- French, S. (2005). Performance-Related Pay in the UK Public Services: Unraveling the Contradictions. New Developments in Public Sector Pay-setting, Queens University Belfast and the UK Labour Relations Agency.
- Frey, B. S. and M. Osterloh (1999). Pay for Performance - Immer Empfehlenswert? Zeitschrift für Führung und Organisation, Münster, Germany.
- Fryer, R. G. (2011). Teacher Incentives and Student Achievement: Evidence from New York City Public Schools (NBER Working Paper 16850). National Bureau of Economic Research, Washington DC.
- Fudge, C. (1990), 'Flexibility Reconsidered: Selected Issues', in OECD (ed.) Flexible Personnel Management in the Public Service, Paris, OECD, pp 91-99.
- Gallup (2011). Employee Engagement: What's Your Engagement Ratio? Washington DC, Gallup Consulting. <http://www.gallup.com/consulting/121535/Employee-Engagement-Overview-Brochure.aspx>.
- Gavagan, T., H. Du, B. Saver, G. Adams, D. Graham, R. McCray and K. Goodrick (2010), 'Effect of Financial Incentives on Improvement in Medical Quality Indicators for Primary Care', Journal of American Board Family Medicine, 23, 622-631.
- Georgellis, Y., E. Iossa and V. Tabvuma (2011), 'Crowding out Intrinsic Motivation in the Public Sector', Journal of Public Administration Research and Theory, 21 (3), 473-493.
- Gerber, A. and N. Malhotra (2008), 'Do Statistical Reporting Standards Affect What Is Published? Public Bias in Two Leading Political Science Journals', Quarterly Journal of Political Science, 3 (313-326).
- Glewwe, P., N. Ilias and M. Kremer (2010), 'Teacher Incentives', American Economic Journal: Applied Economics, 2 (3), 205-227.

- Glickman, S., F. Ou, E. DeLong, M. Roe, B. Lytle, J. Mulgund, J. Rumsfeld, W. Gibler, E. Ohman, K. Schulman and E. Peterson (2007), 'Pay for Performance, Quality of Care, and Outcomes in Acute Myocardial Infarctions', Journal of the American Medical Association, 297 (21), 2373-2380.
- Gneezy, U. and A. Rustichini (2000), 'Pay Enough or Don't Pay at All', The Quarterly Journal of Economics, 115 (3), 791-810.
- Goldhaber, D. D. and D. J. Brewer (2000), 'Does Teacher Certification Matter? High School Teacher Certification and Student Achievement', Educational Evaluation and Policy Analysis, 22 (129).
- Goodman, S. and L. Turner (2010). Teacher Incentive Pay and Educational Outcomes: Evidence from the Nyc Bonus Program (Working Paper). PEPG Conference "Merit Pay: Will It Work? Is It Politically Viable?". Harvard Kennedy School, June 3-4, 2010
- Grady, K., J. Lemkau, L. N. and C. Caddell (1997), 'Enhancing Mammography Referral in Primary Care', Preventive Medicine (26), 791-800.
- Gratz, D. B. (2009), The Peril and Promise of Performance Pay. Making Education Compensation Work, Lanham, MD, Rowman & Littlefield.
- Green, F. (2001), 'It's Been a Hard Day's Night: The Concentration and Intensification of Work in Late Twentieth-Century Britain', British Journal of Industrial Relations, 39 (1), 53-80.
- Grimshaw, D. (1998a). National Systems of Public Sector Pay: Implications for 'Welfare Outcomes' and Economic Stability. The ESRC Labour Studies Seminars - 27th November 1998: Reinventing the State, Centre for Comparative Labour Studies, Department of Sociology, University of Warwick, Economic and Social Research Council, <http://www.csv.warwick.ac.uk/fac/soc/complabstuds/confsem/Grimshaw.htm>.
- Grimshaw, D. (1998b). National Systems of Public Sector Pay: Implications for "Welfare Outcomes" and Economic Stability. The ESRC Labour Studies Seminar. London.
- Grimshaw, D., K. Jaehrling, M. van der Meer, P. Méhaut and N. Shimron (2007), 'Convergent and Divergent Country Trends in Coordinated Wage Setting and Collective Bargaining in the Public Hospitals Sector', Industrial Relations Journal, 38 (6), 591-613.
- Groshen, E. L. (1991), 'Sources of Intra-Industry Wage Dispersion: How Much Do Employers Matter?', Quarterly Journal of Economics, 106 (3), 869-884.
- Gruening, G. (2001), 'Origin and Theoretical Basis of New Public Management', International Public Management Journal (4), 1-25.
- Hakimi, E., N. Manning, S. Prasad and K. Prince (2004), Asymmetric Reforms: Agency-Level Reforms in the Afghan Civil Service, South Asia Region: PREM Working Paper Series, Washington DC, World Bank.
- Hamilton, B. H., J. A. Nickerson and H. Owan (2003), 'Team Incentives and Worker Heterogeneity: An Empirical Analysis of the Impact of Teams on Productivity and Participation', Journal of Political Economy, 111 (3), 465-497.
- Hammer, J. S. and N. Chaudhury (2004), 'Ghost Doctors: Absenteeism in Bangladeshi Health Facilities', World Bank Economic Review, 18, 423-441.
- Hanushek, E. A. and S. G. Rivkin (2006), 'Teacher Quality', in E. Hanushek and F. Welch (eds.) Handbook of the Economics of Education, Amsterdam, North-Holland, pp Chapter 18.
- Heckman, J., C. Heinrich and J. Smith (1997), 'Assessing the Performance of Performance Standards in Public Bureaucracies', The American Economic Review 87 (2), 389-395.
- Heintzman, R. and B. Marson (2005), 'People, Service and Trust: Is There a Public Sector Service Value Chain?', International Review of Administrative Sciences, 71 (4), 549-575, <http://ras.sagepub.com/cgi/content/short/71/4/549>.
- Heneman III, H. G. and A. T. Milanowski (1999), 'Teacher Attitudes About Teacher Bonuses under School-Based Performance Award Programs', Journal of Personnel Evaluation in Education, 12 (4), 327-341.
- Hillman, A., M. Pauly, K. Kerman and C. Martinek (1991), 'Hmo Manager's Views on Financial Incentives and Quality, Health Affairs', Health Affairs, 10 (4), 207-219.

- Hillman, A., K. Ripley, N. Goldfarb, I. Nuamah, J. Weiner and E. Lusk (1998), 'Physician Financial Incentives and Feedback: Failure to Increase Cancer Screening in Medicaid Managed Care', American Journal of Public Health, 88 (11), 1699-1701.
- Hillman, A., K. Ripley, N. Goldfarb, J. Weiner, I. Nuamah and E. Lusk (1999), 'The Use of Physician Financial Incentives and Feedback to Improve Pediatric Preventive Care in Medicaid Managed Care', Pediatrics (104), 931-935.
- Hochberg, Y. V. and L. Lindsey (2010), 'Incentives, Targeting and Firm Performance: An Analysis of Non-Executive Stock Options', Review of Financial Studies, 23 (11).
- Holmstrom, B. (1982), 'Managerial Incentive Problems: A Dynamic Perspective', Review of Economic Studies, 1 (169-182).
- Holmstrom, B. and P. Milgrom (1991), 'Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design', Journal of Law, Economics & Organization, 7, 24-52.
- Hood, C. (2005), 'Public Management: The Word, the Movement, the Science', in E. Ferlie, L. Lynn Jr. and C. Pollitt (eds.) The Oxford Handbook of Public Management, Oxford, OUP, pp 7-26.
- Hood, C. and R. Dixon (2010), 'The Political Payoff from Performance Target Systems: No-Brainer or No-Gainer?', Journal of Public Administration Research and Theory, 281-298.
- Houston, D. J. (2009), 'Motivating Knights or Knaves? Moving Beyond Performance-Related Pay for the Public Sector', Public Administration Review, 69 (1), 43-56.
- Hutton, W. (2010), Fair Pay in the Public Sector: Interim Report, London, H.M. Treasury.
- Independent Evaluation Group (2008), Public Sector Reform: What Works and Why?, Washington DC, World Bank.
- Ipsos-MORI (2006). Change Management and Leadership: The Challenges for the Public Sector. Ipsos-MORI, London.
- Jack, W. (2003), 'Contracting for Health Services: An Evaluation of Recent Reforms in Nicaragua', Health Policy and Planning, 18 (2), 195-204.
- Jackson, S. E., R. S. Schuler and S. Werner (2012), Managing Human Resources, Mason, Ohio, South-Western.
- Jacob, B. A. (2005), 'Accountability, Incentives and Behavior: The Impact of High-Stakes Testing in the Chicago Public Schools', Journal of Public Economics 89, 761-796.
- Jacob, B. A. and S. D. Levitt (2003), 'Rotten Apples: An Investigation of the Prevalence and Predictors of Teacher Cheating', Quarterly Journal of Economics, 118 (3), 843-877.
- Jain, T. and T. Narayan (2011), Incentive to Discriminate? An Experimental Investigation of Teacher Incentives in India (Working Paper), Indian School of Business.
- Jenkins, G. D., A. Mitra, N. Gupta and J. D. Shaw (1998), 'Are Financial Incentives Related to Performance? A Meta-Analytic Review of Empirical Research', Journal of Applied Psychology, 83 (5), 777-787.
- Kahn, C. M., E. C. De Silva and J. P. Ziliak (2001), 'Performance-Based Wages in Tax Collection: The Brazilian Tax Collection Reform and Its Effects', The Economic Journal, 111, 188-205.
- Karlson, N. and H. Lindberg (2011). The Decentralization of Wage Bargaining. The Ratio Institute, Stockholm.
- Kelley, C. (1999), 'The Motivational Impact of School-Based Performance Awards', Journal of Personnel Evaluation in Education, 12 (4), 309-326.
- Kellough, E. J. and H. Lu (1993), 'The Paradox of Merit Pay in the Public Sector: Persistence of a Problematic Procedure', Review of Public Personnel Administration 13, 45-64.
- Kellough, J. E. and L. G. Nigro (2002), 'Pay for Performance in Georgia State Government: Employee Perspectives on Georgiagain after 5 Years', Review of Public Personnel Administration, 22 (2), 146-166.
- Kernaghan, K. (2011), 'Getting Engaged: Public-Service Merit and Motivation Revisited', Canadian Public Administration, 54 (1), 1-21.

- Kerr, S. (1975), 'On the Folly of Rewarding a, While Hoping for B', The Academy of Management Journal, 18 (4), 769-783.
- Ketelaar, A., N. Manning and E. Turkisch (2007). Performance Based Arrangements for Senior Civil Servants - OECD Experiences (OECD Governance Working Paper). Paris.
- Kim, P. S. (2002). Strengthening the Pay-Performance Link in Government: A Case Study of Korea. Governing for Performance in the Public Sector: OECD-Germany High-Level Symposium. Berlin.
- Kingdon, G. and F. Teal (2008), Teacher Unions, Teacher Pay and Student Performance in India: A Fixed Effects Approach (CESifo Working Paper No. 2428), Munich, Germany, Ifo Institute, Center for Economic Studies.
- Kiragu, K. and R. Mukandala (2003). Public Sector Pay Reform - Tactics Sequencing and Politics in Developing Countries: Lessons from Sub-Saharan Africa Pricewaterhousecoopers and University of Dar es Salaam, Dar es Salaam, Tanzania.
- Knez, M. and D. Simester (2001), 'Firm-Wide Incentives and Mutual Monitoring at Continental Airlines', Journal of Labor Economics, 19 (4), 743-772.
- Knudsen, R. and L. Pedersen (1993). Wage Determination and Sex Segregation in Employment in Denmark. Manchester School of Management, UMIST, Manchester.
- Kouides, R., N. Bennett, B. Lewis, J. Cappuccio, W. Barker and M. LaForce (1998), 'Performance-Based Physician Reimbursement and Influenza Rates in the Elderly', American Journal of Preventive Medicine, 14 (2), 89-95.
- Kremer, M. and D. Chen (2001). An Interim Report on a Teacher Attendance Incentive Program in Kenya (Mimeo). Cambridge, Mass., Harvard University.
- Kremer, M., K. Muralidharan, N. Chaudhury, J. S. Hammer and F. H. Rogers (2004), 'Teacher Absence in India: A Snapshot', Journal of the European Economic Association, 3, 2-3.
- Kreps, D. M. (1997), 'Intrinsic Motivation and Extrinsic Incentives', The American Economic Review, 87 (2), 359-364.
- Krueger, A. B. and L. H. Summers (1988), 'Efficiency Wages and the Inter-Industry Wage Structure', Econometrica, 56 (2), 259-293.
- Kumar, P., G. Murray and S. Schetagne (1999), Workplace Change in Canada: Union Perceptions of Impacts, Responses and Support Systems, Kingston Ontario, Queens University.
- Ladd, H. F. (1999), 'The Dallas School Accountability and Incentive Program: Evaluation of Its Impacts on Student Outcomes', Economics of Education Review, 18, 1-16.
- Lafuente, M. and N. Manning (2010). Executive-Legislative Authority over Public Servants' Pay: Lessons from Paraguay. World Bank, Washington DC.
- Lavy, V. (2008). Gender Differences in Market Competitiveness in a Real Workplace: Evidence from Performance-Based Pay Tournaments among Teachers (NBER Working Paper No. 14338). National Bureau for Economic Research, Washington DC.
- Lavy, V. (2009), 'Performance Pay and Teachers' Effort, Productivity and Grading Ethics', American Economic Review, 99 (5), 1979-2011.
- Lazear, E. (1989), 'Pay Equality and Industrial Politics', Journal of Political Economy, 97, 561-580.
- Lazear, E. P. (1981), 'Agency, Earnings Profiles, Productivity and Hours Restrictions', The American Economic Review, 71 (5), 606-620.
- Lazear, E. P. (2000), 'Performance Pay and Productivity', The American Economic Review, 90 (5), 1346-1361.
- Le Grand, J. (2003), Motivation, Agency and Public Policy: Of Knights and Knaves, Pawns and Queens, New York, Oxford University Press.
- Levin-Scherz, J., N. DeVita and J. Timbie (2006), 'Impact of Pay-for-Performance Contracts and Network Registry on Diabetes and Asthma: Hedis Measures in an Integrated Delivery Network', Medical Care Research and Review, 63 (1), 14S-28S.

- Li, J., J. Hurley, P. DeCicca and G. Buckley (2011). Physician Response to Pay-for-Performance: Evidence from a Natural Experiment (Working Paper 16909). National Bureau for Economic Research, Cambridge, Mass.
- Lindauer, D. L. and B. Nunberg (eds) (1994), Rehabilitating Government, Washington DC, World Bank.
- Lindenauer, P., D. Remus, S. Roman, M. Rothberg, E. Benjamin, A. Ma and D. Bratzler (2007), 'Public Reporting and Pay for Performance in Hospital Quality Improvement', New England Journal of Medicine, 365 (5), 486-496.
- Loevinsohn, B. and A. Harding (2005), 'Buying Results? Contracting for Health Service Delivery in Developing Countries', The Lancet (366), 676-681.
- Luthans, F. (1973), Organizational Behavior, New York, NY, McGraw-Hill.
- Maguire, M. (1993), 'Pay Flexibility in the Public Sector -- an Overview', in OECD (ed.) Pay Flexibility in the Public Sector, Paris, OECD, pp 9-18.
- Mandel, K. and U. Kotagal (2007), 'Pay for Performance Alone Cannot Drive Quality', Archives of Pediatric Adolescent Medicine, 161 (7), 650-655.
- Mangham, L. (2007). Addressing the Human Resource Crisis in Malawi's Health Sector: Employment Preferences of Public Sector Registered Nurses (Esau Working Paper 18). Overseas Development Institute, London.
- Manning, N. (2001), 'The Legacy of the New Public Management in Developing Countries', International Review of Administrative Sciences, 67 (2), 297-312.
- Manning, N. and N. Parison (2003), International Public Administration Reform : Implications for the Russian Federation, Moscow, Higher School of Economics, with the World Bank.
- Marsden, D. (2004), 'The Role of Performance-Related Pay in Renegotiating The "Effort Bargain": The Case of the British Public Service', Industrial and labor Relations review, 57 (3), 350-370.
- Marsden, D. (2009). The Paradox of Performance Related Pay Systems: Why Do We Keep Adopting Them in the Face of Evidence That They Fail to Motivate? Centre for Economic Performance, London School of Economics, London.
- Mathieu, J. and D. Zajac (1990), 'A Review and Meta-Analysis of the Antecedents, Correlates, and Consequences of Organisational Commitment', Psychological Bulletin of the American Psychological Association, 108, 171-194.
- Mayer, R. C. and J. H. Davis (1999), 'The Effect of the Performance Appraisal System on Trust for Management: A Field Quasi-Experiment', Journal of Applied Psychology, 84 (1), 123-136.
- McMenamin, S. B., H. H. Schauffler, S. M. Shortell, T. G. Rundall and R. R. Gillies (2003), 'Support for Smoking Cessation Interventions in Physician Organizations: Results from a National Study', Medical Care, 41, 1396-1406.
- McNamara, P. (2005), 'Quality-Based Payment: Six Case Examples', International Journal for Quality in Health Care, 17 (4), 357-363.
- Meessen, B., J. Kashala and L. Musango (2007), 'Output-Based Payment to Boost Staff Productivity in Public Health Centres: Contracting in Kabutare District, Rwanda', Bulletin of the World Health Organization, 85 (2), 108-115.
- Meessen, B., L. K. Musango, J. and J. Lemlin (2006), 'Reviewing Institutions of Rural Health Centres: Performance Initiative in Butare, Rwanda', Tropical Medicine and International Health, 11 (8), 1303-1317.
- Milkovich, G. and A. Wigdor (1991), Pay for Performance: Evaluating Performance Appraisal and Merit Pay, Washington, DC, National Academy Press.
- Mills, Z., S. Dahal, C. Garrity and N. Manning (2011). Wage Bill and Pay Compression Summary Note. World Bank, Washington DC.
- Moynihan, D. (2008), The Dynamics of Performance Management, Washington, DC, Georgetown University Press.
- Moynihan, D. P. and S. K. Pandey (2007), 'The Role of Organizations in Fostering Public Service Motivation', Public Administration Review, 67 (1), 40-53.

- Muralidharan, K. and V. Sundararaman (2009). Teacher Performance Pay: Experimental Evidence from India (NBER Working Paper 15323). National Bureau for Economic Research, Washington DC.
- Muralidharan, K. and V. Sundararaman (2011), 'Teacher Opinions on Performance Pay: Evidence from India', Economics of Education Review, 30 (3), 394-403.
- Murnane, R. J. and D. K. Cohen (1986), 'Merit Pay and the Evaluation Problem: Why Most Merit Pay Plans Fail and Few Survive', Harvard Educational Review, 56 (1), 1-17.
- Nagin, D. S., J. B. Rebitzer, S. Sanders and L. J. Taylor (2002), 'Monitoring, Motivation, and Management: The Determinants of Opportunistic Behavior in a Field Experiment', American Economic Review, 92 (4), 850-873.
- Ndeti, D. M., L. Khasakhala and J. O. Omolo (2008). Incentives for Health Worker Retention in Kenya: An Assessment of Current Practice. Africa Mental Health Foundation and Institute of Policy Analysis and Research, Dar es Salaam, Kenya.
- Neal, D. (2011). The Design of Performance Pay in Education (NBER Working Paper 16710). National Bureau for Economic Research, Washington DC.
- Niemiec, C. P., R. M. Ryan and E. L. Deci (2009), 'The Path Taken: Consequences of Attaining Intrinsic and Extrinsic Aspirations in Post-College Life', Journal of Research in Personality, 73 (3), 291-306.
- Niskanen, W. (1973), Bureaucracy: Servant or Master, London, Institute of Economic Affairs.
- Norton, E. (1992), 'Incentive Regulation of Nursing Homes: Specification Tests of the Markov Model', in D. Wise (ed.) Topics in the Economics of Aging, Chicago, University of Chicago Press, pp 275-304.
- Nunberg, B. (1988). Public Sector Pay and Employment Reform (World Bank Working Paper). World Bank, Washington DC.
- Nunberg, B. and J. Nellis (1990), Civil Service Reform and the World Bank (World Bank Working Paper), Washington DC, World Bank.
- Nunberg, B. and R. Taliervo (2012). Making Things Worse: Do Aid Donors Undermine Civil Service Reforms? (Unpublished Manuscript). Washington DC.
- O'brien, J. and M. O'Donnell (2007), 'From Workplace Bargaining to Workplace Relations: Industrial Relations in the Australian Public Service under the Coalition Government', in M. J. Pittard and W. Phillipa (eds.) Public Sector Employment in the Twenty-First Century, Canberra, Australian national University Press.
- Odden, A. and C. Kelley (2002), Paying Teachers for What They Know and Do, Thousand Oaks, CA, Corwin Press.
- OECD (1993). Pay Flexibility in the Public Sector. OECD, Paris.
- OECD (1996), Pay Reform in the Public Service: Initial Impact on Pay Dispersion in Australia, Sweden, and the United Kingdom, Paris, OECD PUMA.
- OECD (1997a), Measuring Public Employment in OECD Countries: Sources, Methods and Results, Paris, OECD.
- OECD (1997b), Trends in Public Sector Pay in OECD Countries, Paris, OECD/PUMA.
- OECD (2004a). Trends in Human Resources Management Policies in OECD Countries. An Analysis of the Results of the OECD Survey on Strategic Human Resources, Paper Presented to the Human Resources Management Working Party. OECD, Paris.
- OECD (2004b), 'Wage-Setting Institutions and Outcomes', in J. Martin (ed.) OECD Employment Outlook, Paris, OECD, pp 127-181.
- OECD (2005a), Modernising Government: The Way Forward, Paris, OECD.
- OECD (2005b), Performance-Related Pay Policies for Government Employees, Paris, OECD.
- OECD (2008), The State of the Public Service, Paris, OECD.
- OECD (2009), Government at a Glance, Paris, OECD.
- OECD (2011a). 2010 Human Resources Management Composites: Theoretical Framework, Construction and Weighting OECD, Paris.

- OECD (2011b), Government at a Glance, Paris, OECD.
- OECD Working Party of Senior Budget Officials (2011). Restoring Public Finances. Public Governance and Territorial Development Directorate, OECD, Paris.
- Osterloh, M. and J. Frost (2002), 'Motivation and Knowledge as Strategic Resources', in B. S. Frey and M. Osterloh (eds.) Successful Management by Motivation: Balancing Intrinsic and Extrinsic Incentives, New York, Springer-Verlag, pp 27-51.
- Paarsch, H. J. and B. S. Shearer (1999), 'The Response of Worker Effort to Piece Rates', Journal of Human Resources 34 (4), 634.
- Painter, M. (2006), 'Sequencing Civil Service Pay Reforms in Vietnam: Transition or Leapfrog', Governance, 19 (2), 325-346.
- Palmer, D. (2006), 'Tackling Malawi's Human Resources Crisis', Reproductive Health Matters, 14 (27), 27-39.
- Pearson, S., E. Schneider, K. K., K. Coltin and J. Singer (2008), 'The Impact of Pay-for-Performance on Health Care Quality in Massachusetts, 2001-2003', Health Affairs, 27 (4), 1167-1176.
- Perry, J. L., T. A. Engbers and S. Y. Jun (2009), 'Back to the Future? Performance-Related Pay, Empirical Research and the Perils of Persistence', Public Administration Review, 69 (1), 39-51.
- Perry, J. L. and A. Hondeghem (eds) (2008), Motivation in Public Management: The Call of Public Service, Oxford, Oxford University Press.
- Perry, J. L., D. Mesch and L. Paarlberg (2006), 'Motivating Employees in a New Governance Era: The Performance Paradigm Revisited', Public Administration Review, 66 (4), 505-514.
- Petersen, L. A., L. D., T. U. Woodard, C. Daw and S. Sookanan (2006), 'Does Pay-for-Performance Improve the Quality of Health Care?', Annals of Internal Medicine, 145 (4), 265-272.
- Pfeffer, J. (1998a), The Human Equation: Building Profits by Putting People First, Cambridge, Mass., Harvard Business School Press.
- Pfeffer, J. (1998b), 'Seven Practices of Successful Organizations', California Management Review, 40 (2), 96-124.
- Pink, D. H. (2009), Drive: The Surprising Truth About What Motivates Us, New York, Riverhead.
- Podsakoff, P. M. and S. B. Mackenzie (1994), 'Organisational Citizenship Behavior and Sales Unit Effectiveness', Journal of Marketing Research, 31 (351-363).
- Pollitt, C. (1993), Managerialism and the Public Services, Oxford, Blackwell.
- Pollitt, C. (1995), 'Justification by Works or by Faith', Evaluation, 1 (2), 133-154.
- Pollitt, C. and S. Dan (2011a), The Impacts of the New Public Management in Europe: A Meta-Analysis COCOPS, Erasmus University, Rotterdam.
- Pollitt, C. and S. Dan (2011b), The Impacts of the New Public Management in Europe: A Meta-Analysis (COCOPS Working Paper No. 3), Brussels, European Commission.
- Porter, L. W. and E. E. Lawler III (1968), Managerial Attitudes and Performance, Homewood, IL, Dorsey Press.
- Pourat, N., T. Rice, M. Tai-Seale, G. Bolan and J. Nihalani (2005), 'Association between Physician Compensation Methods and Delivery of Guideline-Concordant Std Care: Is There a Link?', The American Journal of Managed Care, 11, 426-432.
- Prendergast, C. (1998). What Happens within Firms? A Survey of Empirical Evidence on Compensation Policies (NBER Working Paper). National Bureau for Economic Research, Washington DC.
- Prendergast, C. (1999), 'The Provision of Incentives in Firms', Journal of Economic Literature 37 (1), 7-63.
- Prentice, G., S. Burgess and C. Propper (2007), 'Performance Pay in the Public Sector: A Review of the Issues and Evidence'.
- Pritchett, L. and M. Woolcock (2004), 'Solutions When the Solution Is the Problem: Arraying the Disarray in Development', World Development, 32 (2), 191-212.

- Propper, C. and D. Wilson (2003), The Use and Usefulness of Performance Measures in the Public Sector (Cmpo Working Paper Series No. 03/073), Bristol, UK, The Centre For Market And Public Organisation.
- Rafferty, A. M., J. Maben, E. West and D. Robinson (2005). What Makes a Good Employer? International Council of Nurses, Geneva, http://www.icn.ch/images/stories/documents/publications/GNRI/Issue3_Employer.pdf.
- Rexed, K., C. Moll, N. Manning and J. Allain (2007). Governance of Decentralised Pay Setting in Selected OECD Countries (OECD Working Papers on Public Governance, 2007/3). OECD, Paris, <http://caliban.sourceoecd.org/vl=7179447/cl=20/nw=1/rpsv/cgi-bin/wppdf?file=514qdfvl56d.pdf>.
- Rosenthal, M., R. Frank, Z. Li and A. Epstein (2005), 'Early Experience with Pay-for-Performance', Journal of the American Medical Association, 294 (14), 1788-1793.
- Ryan, R. M. and E. L. Deci (2000), 'Self-Determination Theory and the Facilitation of Intrinsic Motivation, Social Development, and Well-Being', American Psychologist, 55 (1), 68-78.
- Safran, D., W. Rogers, A. Tarlov, T. Inui, D. Taira, J. Montgomery, J. Ware and C. Slavin (2000), 'Organizational and Financial Characteristics of Health Plans. Are They Related to Primary Care Performance?', Archive of Internal Medicine, 160, 69-76.
- Samaratunge, R., Q. Alam and J. Teicher (2008), 'The New Public Management Reforms in Asia: A Comparison of South and Southeast Asian Countries', International Review of Administrative Sciences, 74 (1), 25-46.
- Sauermann, H. and W. M. Cohen (2008). What Makes Them Tick? Employee Motives and Firm Innovation (NBER Working Paper No. 14443). NBER, Cambridge MA, <http://www.nber.org/papers/w14443.pdf>.
- Schick, A. (1998), 'Why Most Developing Countries Should Not Try New Zealand's Reforms', World Bank Research Observer (International), 13, 23-31.
- Shen, Y. (2003), 'Selection Incentives in a Performance-Based Contracting System', Health Services Research, 38 (2), 535-552.
- Singh, P. (2010). Performance Pay and Information: Reducing Child Malnutrition in Urban Slums. London School of Economics, London.
- Skinner, B. F. (1969), Contingencies of Reinforcement, New York, NY, Appleton-Century-Crofts.
- Soeters, Robert & Griffiths, Fred (2003) Improving government health services through contract management: a case from Cambodia, Health Policy and Planning, 18 (1), 74-83
- Soeters, R., C. Habineza and P. Peerenboom (2006), 'Performance-Based Financing and Changing the District Health System: Experience from Rwanda', Bulletin of the World Health Organization, 84, 884-889.
- Springer, M. G., D. Ballou, L. Hamilton, V.-N. Le, J. R. Lockwood, D. F. McCaffrey and M. P. B. M. Stecher (2010). Teacher Pay for Performance: Experimental Evidence from the Project on Incentives in Teaching. National Center on Performance Incentives at Vanderbilt University, Nashville, TE.
- Stajkovic, A. D. and F. Luthans (2003), 'Behavioral Management and Task Performance in Organizations: Conceptual Background, Meta - Analysis, and Test of Alternative Models', Personnel Psychology (56), 15-194.
- Stazyk, E. C. (2010). Crowding out Intrinsic Motivation? The Role of Performance-Related Pay. American University, School of Public Affairs Washington DC.
- Steel, N., S. Maisey, A. Clark, R. Fleetcroft and A. Howe (2007), 'Quality of Clinical Primary Care and Targeted Incentive Payments: An Observational Study', British Journal of General Practice, 57 (449-454).
- Stevens, M. and S. Tegemann (2004), 'Comparative Experience with Public Service Reform in Ghana, Tanzania and Zambia', in S. Kpundeh and B. Levy (eds.) Building State Capacity in Africa, Washington DC, World Bank, pp 43-86.

- Straberg, T. (2010). Employee Perspectives on Individualised Pay: Attitudes and Fairness Perceptions. Department of Psychology. Stockholm, University of Stockholm. PhD.
- Therkildsen, O., P. Tidemand, B. Bana, A. Kessy, J. Katongole, M. B. Ddiba and M. Nielsen (2007). Staff Management and Organisational Performance in Tanzania and Uganda: Public Servant Perspectives. Danish Institute for International Studies, Copenhagen, Denmark.
- Thompson, J. R. (2006), 'The Federal Civil Service: The Demise of an Institution', Public Administration Review, 66 (4), 496-503.
- Thompson, J. R. and S. L. Fulla (2001), 'Effecting Change in a Reform Context: The National Performance Review and the Contingencies of "Microlevel" Reform Implementation', Public Performance and Management Review, 25 (2), 155-175.
- Vaghela, P., M. Ashworth, P. Schofield and M. C. Gulliford (2009), 'Population Intermediate Outcomes of Diabetes under Pay-for-Performance Incentives in England from 2004 to 2008', Diabetes Care, 32 (427-9).
- Valentine, T. R. (2002). A Medium-Term Strategy for Enhancing Pay and Conditions of Service in the Zambian Public Service (Final Report). Management Development Division, Cabinet Office, Lusaka, Zambia.
- van Dijk, F., J. Sonnemans and F. van Winden (2001), 'Incentive Systems in a Real Effort Experiment', European Economic Review, 45 (2), 187-214.
- Vance, R. J. (2003). Employee Engagement and Commitment: A Guide to Understanding, Measuring and Increasing Engagement in Your Organization. Society for Human Resource Management Foundation, Alexandria, VA.
- Vandenberg, R. and C. Lance (1992), 'Satisfaction and Organisational Commitment', Journal of Management, 18, 153-167.
- Vegas, E. and I. Umansky (2005). Improving Teaching and Learning through Effective Incentives: What Can We Learn from Education Reforms in Latin America. World Bank, Washington DC.
- Vroom, V. H. (1964), Work and Motivation, Hoboken, NJ, Wiley.
- Vujicic, M. (2009). How You Pay Health Workers Matters: A Primer on Health Worker Remuneration Methods. World Bank, Washington DC.
- Wagstaff, A. and M. Claeson (2004), The Millennium Development Goals for Health: Rising to the Challenges, Washington DC, World Bank.
- Wallerstein, M. (1999), 'Wage-Setting Institutions and Pay Inequality in Advanced Industrial Societies', American Journal of Political Science, 43 (3), 649-680.
- Weber, M. (1978), Economy and Society (Vol. 2), Berkely, CA, University of California Press.
- Weibel, A., K. Rost and M. Osterloh (2009), 'Pay for Performance in the Public Sector - Benefits and (Hidden) Costs', Journal of Public Administration Research and Theory, 20 (2), 387-412.
- White, G. (2000), 'Pay Flexibility in European Public Services: A Comparative Analysis', in D. Farnham and S. Horton (eds.) Human Resources Flexibilities in the Public Services, London, MacMillan Press, pp 255-279.
- Wilms, W. W. and R. R. Chapleau (1999), 'The Illusion of Paying Teachers for Student Performance', Education Week, 19 (10).
- Winters, M. A., G. W. Ritter, J. P. Greene and R. Marsh (2009), 'Student Outcomes and Teacher Productivity and Perceptions in Arkansas', in M. G. Springer (ed.) Performance Incentives. Their Growing Impact on American K-12 Education, Washington DC, Brookings Institution Press.
- Witter, S., T. Zulfikur, S. Javeed, A. Khan and A. Bari (2011), 'Paying Health Workers for Performance in Battagram District', Human Resources for Health, 9 (23).
- Woessman, L. (2010), Cross-Country Evidence on Teacher Performance Pay (CESifo Working Paper No. 3151), Munich, Germany, Ifo Institute, Center for Economic Studies.
- World Bank (1999). Civil Service Reform: A Review of World Bank Assistance: Report No. 19211. OED, World Bank, Washington DC.

- World Bank (2001). Salary Supplements and Bonuses in Revenue Departments (Final Report). World Bank, Washington DC.
- World Bank (2004). Labor Markets in Europe and Central Asia. World Bank, Washington DC.
- World Bank (2007). What Do We Know About School-Based Management, Washington DC, World Bank.
- World Bank (2009). Pay Policy Reform: Building a Foundation for Public Sector Performance through Improved Public Sector Pay Policy by Using A "Single Pay Spine" World Bank, Washington DC.
- Yemin, E. (1993), 'Labour Relations in the Public Service: A Comparative Overview', International Labour Review, 132 (4), 469-490.
- Young, G., M. Meterko, H. Beckman and E. Baker (2007), 'Effects of Paying Physicians Based on Their Relative Performance for Quality', Journal of General Internal Medicine, 22 (6), 872-887.