



**Office of Manpower Economics**

## **Performance pay in the public sector: A review of the issues and evidence**

**Commissioned by: Office of Manpower  
Economics**

**Undertaken by:  
Graham Prentice  
University of Warwick**

**Simon Burgess  
CMPO**

**Carol Propper  
CMPO**

**November 2007**

**The work described in this report was carried out under contract as part of OME's research programme. The views and judgements expressed in this report are therefore those of the contractor and do not necessarily reflect those of OME.**

## **Executive summary**

The purpose of this report is to assess recent, robust empirical studies of public sector incentive schemes and identify significant findings and common trends. As a backdrop we note some of the schemes in operation in the UK.

Our assessment begins with the case for and against use of performance related pay, focusing on its use in the public sector and highlighting the fact that performance pay can have several effects, not all of which are beneficial.

It then reviews the empirical literature for the following groups: civil servants (including police, armed forces and judiciary), healthcare workers and teachers. In choice of evidence, we have selected the most recent and the most robust evidence, which often comes from exploitation of actual or natural experiments. We have also tried to give some idea of the size of the evidence base for each set of workers, so assessment can be made of the ‘weight’ of the evidence. However, it should be noted that overall the size of the evidence base is not extensive and much of it is not from the UK so that often we are reliant on a small number of papers.

Our first conclusion is that there is evidence that public sector workers do respond to financial schemes, particularly in the field of education and possibly also in health. Sometimes these responses are small, but often the incentive in the scheme is small too. Our second is that there is also evidence of gaming – manipulation of behaviour that uses resources and does not increase productivity - in response to schemes. Our final conclusion is that welfare calculations of the overall effect of performance related pay in the public sector are extremely rare, so that while it is possible to show that in some cases public sector workers have responded to these schemes, the overall benefits for society have not been assessed. In part, this is because such assessments are intrinsically hard to make; in other cases the schemes have not run for long enough to observe more than their impact on the employees who have been the subject of the scheme.

<b>EXECUTIVE SUMMARY .....</b>	<b>2</b>
<b>INTRODUCTION.....</b>	<b>5</b>
<b>SCHEMES CURRENTLY IN OPERATION .....</b>	<b>5</b>
<b>ISSUES .....</b>	<b>6</b>
<b>The case for performance related pay .....</b>	<b>6</b>
Productivity and efficiency .....	6
Productivity gains from recruitment – selection and sorting.....	7
<b>The case against performance related pay: general issues .....</b>	<b>8</b>
Inter-temporal aspects .....	8
<b>The case against performance related pay: particular issues for the public sector .....</b>	<b>9</b>
Multiple tasks, measurement and incentives.....	9
Multiple principals .....	10
Teams/groups in production and rewards.....	10
Public sector motivation .....	11
<b>RECENT EMPIRICAL EVIDENCE: .....</b>	<b>12</b>
<b>Why is it hard to get good (UK) evidence? .....</b>	<b>12</b>
<b>Private Sector Evidence .....</b>	<b>13</b>
<b>Civil Servants .....</b>	<b>14</b>
Summary .....	14
The US JTPA programme .....	15
International evidence .....	17
UK evidence .....	18
Police – a special case.....	20
Armed Forces .....	21
Judiciary .....	22
<b>Healthcare Workers.....</b>	<b>23</b>
Summary .....	23

US evidence .....	24
UK evidence .....	27
<b>Teachers .....</b>	<b>29</b>
Summary .....	29
US evidence .....	30
Evidence from Israel .....	33
UK evidence .....	35
<b>CONCLUSION .....</b>	<b>37</b>

## **Introduction**

The UK government has been reviewing the role of pay for performance schemes (e.g. Makinson Report, 2000) and a number of public sector incentive schemes are now in place at national and local level. But a recent report, (PWC 2006), concludes that little real progress has been made so far to link pay to performance for most public sector groups. Part of the reason for this is a policy of increasing centralisation, which is making it difficult for local council managers in particular to create pay schemes in response to local market forces.

This review first lists some schemes currently in operation in the UK. The meat of the report begins with a brief summary of the case for and against the use of financial incentive schemes, with particular focus on their use in the public sector. We then examine both UK and relevant international evidence on public sector financial incentives. The evidence section is divided along common research lines to examine civil servants, teachers and healthcare workers separately. Each main section starts by summarising the key messages from the research and contains further subdivisions dealing with international and UK evidence. Some particular issues for the armed forces, the police and the judiciary are also briefly considered. Our final section offers concluding remarks.

## **Schemes currently in operation**

Performance pay remains part of the landscape of public sector pay, but is perhaps lower down the political agenda than it was at the time of the Makinson Report. After the report, there was a hiatus when some of the energy that went into reforming performance pay went into a tougher stance on pay equality and pay ‘modernisation’ (see IDS, 2007). More recently, however, HM Treasury identified performance pay as a key objective in their pay guidelines for 2006/7. Furthermore, performance pay is a live issue in discussions around teachers pay, and much of the controversy around the new GP contract relates to the calibration of the points system for performance. One example from IDS (2007) is from the pay settlement in the British Library, where “Treasury approval [for the deal] was forthcoming because the offer was seen to

balance a strong link between performance and pay, with frozen pay scale maximums against a relatively high pay award and shortened pay scales” (p. 56). However, despite this, it remains true that for most in the public sector, the value of performance pay is modest. Some examples of this are as follows:

- In the Defence Aviation Repair Agency, “non-consolidated awards of between £50 and £150 were paid for performance” (IDS, 2007, p. 66).
- In the Department for Constitutional Affairs, “a flat-rate non-consolidated bonus of £400 was paid to all staff awarded an ‘exceeded’ performance rating” (IDS, 2007, p. 68)
- In the Foreign and Commonwealth Office, non-consolidated rewards for exceptional performance varied by salary grade between £900 and £1850. (IDS, 2007, p. 76)
- In the Home Office, top ranked performers were awarded 2% of salary. (IDS, 2007, p. 80)
- In local government, performance pay also remains rare (IDS, 2007, p. 100, 107).
- Pay for performance structures in education and health are discussed in the relevant sections below.

## **Issues**

### ***The case for performance related pay***

#### **Productivity and efficiency**

The arguments for linking pay to performance are the following. Effective financial incentives provide an opportunity to improve the productivity of public sector workers. With compensation linked to performance, employees should expend more effort, lifting the quantity and/or the quality of their output. Thus by promoting better performance internally, governments can use incentives as a means of delivering superior public services. The introduction of performance related pay can also motivate employees to pursue professional development opportunities that previously

offered little in the way of additional benefits for the individual. Productivity is therefore likely to improve both in the short run, because employees are working harder, and also in the longer run, as staff professional development generates further gains in productivity (see Lavy 2007).

In the public sector, financial incentives also project a clear message about which outcomes are valued by society, and by how much. Employees can then prioritize tasks correctly and allocate more time and effort to higher valued tasks. This is the so-called “efficiency” case for performance related pay, which demands that incentive scheme designers have a detailed understanding of the priorities of the users being served.

### **Productivity gains from recruitment – selection and sorting**

Alongside improvements in the productivity and efficiency of current public sector employees is the impact of financial incentives on the potential pool of new applicants waiting to join. Certainly, increasing the level of compensation associated with such work will increase the quantity and quality of new applicants. Offering the chance of higher pay for strong performers might tempt such candidates away from what are otherwise more attractive opportunities elsewhere. This “selection” effect helps to bring more able workers into the workforce, with obvious positive consequences for overall productivity. In fact, Lazear argues that as a source of productivity gains, the selection effects from performance-linked-pay are similar to the gains available from motivation effects, if not greater (see Lazear 2000).

Financial incentives (and penalties) could also have desirable consequences for the retention of high achievers and the turnover of low achievers over time. Besley and Machin (2006) found that the existence of a performance premium associated with being a good head teacher resulted in poor head teachers leaving their jobs rather than accepting lower salaries. Lavy (2007) notes, however, that equally poor outside opportunities might curb this kind of effect for lower level workers and more generally, the presence of (or lack of) alternative opportunities will have a significant impact on the sorting of workers between establishments.

## *The case against performance related pay: general issues*

### **Inter-temporal aspects**

Most incentive schemes remain in place for extended periods during which the performance of employees is measured and bonuses are awarded multiple times. Employers can, therefore, observe the ability of an employee from previous performance measurements and make an assessment of the difficulty of the work. Employers wishing to reduce the costs of their scheme can then realign thresholds and raise targets to restrict the distribution of financial bonuses at all performance levels.

Anticipating such changes, employees are likely to reduce their effort in earlier periods to mislead the employer into believing that the task is more difficult than it is in reality. This phenomenon is known as the “ratchet effect”, which occurs when employers give an undertaking not to modify the incentive system after it is introduced. Like free rider effects in large teams, ratchet effects have the potential to reduce the value of financial incentives schemes to employers (see Burgess and Ratto 2003).

In contrast, career concerns can work in the opposite direction to ratchet effects. Whether ratchet effects can be overcome depends upon the age of employees, their prospects and the level of uncertainty regarding their ability, amongst other factors. Supposing an employer is unaware of the true ability of an employee. Knowing this, the employee can work harder initially, to form the impression that they are a high ability worker. An employer will then observe the employee’s high effort level and form the perception that they are a high ability worker. The pay off for the employee emerges over the long run, in terms of improved job opportunities and higher future earnings. Indeed, the existence of career concerns suggests that explicit financial incentives may not be necessary to motivate younger, untested employees. Furthermore, greater uncertainty regarding employee ability and true performance, as is a feature of public sector roles, may actually strengthen career concerns effects (see Dixit 2002). Attempts to measure performance in order to link it to pay may, therefore, reduce these effects.

## ***The case against performance related pay: particular issues for the public sector***

While some of the possible issues relating to financial incentives are generic, the special nature of the public sector creates particular challenges.

### **Multiple tasks, measurement and incentives**

In many cases public services are multifaceted and, as a result, the objectives of an organisation are difficult to define and so to measure. So for example, the objectives of a school might be to provide a “good education” but this is harder to define than, say, the production of a cars or the adequate collection of garbage. Even breaking down the overall objectives into sub-components can be difficult; there is for example, much debate about what constitutes a good measure of hospital quality. This means that it can be hard to find good measures of performance and the measures adopted may be very noisy – they impart relatively little information about the effort of an employee or the organisation (Propper and Wilson 2003). As a consequence, linking rewards to the meeting of performance targets does not give effective incentives, may lead to some elements of strategic behaviour and imposes unnecessary risk on employees.

One of the reasons that these services are complex is because they involve several dimensions, some of which are relatively easy to measure, others of which are much harder to measure. Examples of the former are school students’ performance on standardised tests and the re-admission of patients after operations. Examples of the latter are becoming culturally aware or improving the long term health of patients. This difference in measurability may mean that incentives can only be linked to the easy to measure outcomes. This may lead to an excessive focus on these outcomes at the expense of other tasks. So hospitals which receive financial incentives linked to cost reduction may have incentives to compromise on quality, while teachers monitored on student pass rates may choose to reduce the effort they put into less able students.

Exerting more effort on one task, therefore, reduces the amount of time or effort available for other tasks which are substitutes. In this way, higher incentives in

one task will drive an employee's effort away from alternative tasks. To avoid misallocation of effort by an employee, the employer will have to weaken the incentives on the more accurately measured tasks.

### **Multiple principals**

Precisely because the services are public in nature, several groups in society will be interested in their provision, cost and efficiency. The tax payer is always one party: tax payers' desires to keep down costs may conflict with those who are providing the service who are interested in high quality. Within an organisation, public service providers may serve several masters: for example, doctors get professional recognition from their research activities, but the managers (and possibly the patients) of hospitals might prefer they focused their time on other, less professionally recognised tasks, such as reducing waiting times for common procedures. The economics literature stresses that in these cases - referred to as multiple principals - financial incentives should be weaker than in the case where there is only a single principal.

### **Teams/groups in production and rewards**

Performance incentives designed to reward performance at group level have become a very popular way of rewarding good performance in the public sector (e.g the Makinson schemes in the UK). Team level financial incentives are attractive partly because they help to address concerns that individual level incentives might lower morale and, ultimately, reduce productivity in the workplace. Another reason for their prevalence is that group level performance measures are often the only ones available; measuring public sector output at an individual level can be difficult. Indeed, they are popular despite the fact that group level schemes ought to suffer from "free-rider" effects that could reduce the impact of financial incentives on performance. The free-rider problem occurs when the work of each team member complements the work of others to the extent that each person's contribution is indistinguishable. Large team sizes are more likely to suffer from free-rider effects, as are teams that are involved in the production of output that is hard to measure or quantify (as is commonly the case

in the public sector). More expensive monitoring techniques may be necessary, and this can add to the costs of implementing a financial incentive programme.

On the other hand, team incentives can help to promote peer monitoring in smaller teams, particularly when team members are mutually dependent on each other to achieve performance thresholds. Whether or not the threat posed by peer monitoring is credible will depend on how long the team is together. Team relationships are built up over time and repeated interactions amplify the costs of weak workers to teams. Thus under team-based financial incentives, co-operation and interdependence could provide a motivation to exert peer sanctions.

### **Public sector motivation**

Much as career concerns provide an internal motivation to exert extra effort even in the absence of financial incentives, public sector (intrinsic) motivation can have a similar impact, this time independent of an employee's age and the level of uncertainty over job performance. Public sector workers may care about the outcomes or mission of the public organisation and gain satisfaction, indeed, intrinsic compensation, from knowing that their work is contributing to it. Having internalised the goals of their organisation, police officers may care about the safety of citizens, nurses and doctors about the health of their patients and teachers about the attainment of their pupils, to the extent that they obtain welfare from seeing their user's needs being met. It is possible that external financial incentives could overwhelm public service motivation, since it suggests to the employee that their employer recognises no association between output and effort other than a pure, market relationship.

There are two recent developments to this argument. Francois (2000) suggests that intrinsically motivated employees should actually work best when incentives are small or even absent and employers commit not to divert any surpluses or public sector "profits" away from the organisation's mission. Besley and Ghatak (2005) develop this further, arguing that if public sector organisations post missions during their recruitment process, the natural sorting of applicants will do the job of financial incentives. On the other hand, financial incentives may help to focus effort on other organisation goals which could have been neglected if employers relied on public service motivation alone. This brings us back to the efficiency argument in

which financial incentives help well meaning, intrinsically motivated employees to prioritise tasks in the “right” way.

## Recent empirical evidence:

### *Why is it hard to get good (UK) evidence?*

While there is a large literature on financial incentives in the private sector, particularly at CEO level, actual empirical evidence forms only a relatively small part of this. In the public sector, much less research into financial incentives has been undertaken, and empirical evidence is particularly scarce. In order to present a more comprehensive, informed view, it is necessary to consider the UK literature (where it does exist) alongside evidence from financial incentive programmes operating in other countries. Even so, large gaps in the research remain, particularly with respect to the armed forces, the prison service, nursing and midwifery, and at senior levels. The table below highlights some of the gaps in the empirical literature and specifically, the lack of substantial UK evidence.

The table details the number of empirical studies on public sector financial incentive programmes found by profession and region dated 1999 or later:

	Teachers	Civil Servants	Doctors	Police	Judiciary	Military
UK	2	2	3			
US	7	5	7	4	1	
Other countries	8	1			2	

This lack of evidence partly arises because of the inability or unwillingness of governments to carry out experiments within the public sector: changes are either introduced for all service deliverers or all eligible recipients, or they are not introduced at all. This means that it is difficult to net out the effect of a scheme from other changes that may have taken place at the same time as the scheme. So 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 (Burgess and Ratto 2003).

In selecting evidence, therefore, we have focused on the more robust analyses that is taken from either actual experiments or from so-called ‘natural experiments’: situations in which those subjected to performance related pay can be compared to an appropriate, unaffected, control group. This method has the added advantage that characteristics such as “ability”, which may have a bearing on the results but which, due to their nature, are unobservable, can be factored out using the data. The disadvantage of many such experiments is that participants are often not chosen at random and so are not representative of their sector. This remains a common problem throughout much of the empirical evidence on public sector financial incentives.

### ***Private Sector Evidence***

Prendergast (1999) remains the most recent, comprehensive review of robust evidence on incentives in both private and public sectors. It concludes that:

- There is strong evidence that workers do respond to incentives;
- Job contracts, in which the type of incentive being used by the employer is specified, frequently utilize a different type of incentive structure to the one predicted as optimal by economic theory. The implication of this is that some of the economic theory of incentives remains unconfirmed by empirical evidence;
- Evidence of incentive programmes that reward relative performance improvements is also inconclusive;
- The requirements of empirical literature, that output and quality is easy to observe and measure, mean that very little evidence exists for the majority of jobs in which performance is measured using subjective performance evaluations. This represents a substantial gap in the literature.

A robust private sector case study comes from Lazear (2000) who analyses the response of 3000 employees of Safelite Glass Corporation (a large US auto windscreen installation firm) to a new contract that links workers' pay to the number of windscreens they install. As the new contract was phased in gradually between 1994 and 1995, Lazear is able to use individual level data covering a 19 month period to control for outside influences on worker productivity. He estimates the switch from hourly wages to output-linked (piece-rate) pay resulted in a 44% gain in output per worker on average. He found that incentive effects were responsible for approximately half the 44% gain in output, while the other half of the gain was attributed to selection effects. Further analysis suggests the large selection effect came as a result of a reduction in employee turnover among high performing employees already employed by Safelite, and from the recruitment of more able workers attracted by the potential to earn higher wages. The scheme resulted in a 10% wage increase for workers. The firm addressed quality issues by adopting a policy whereby the worker responsible for a poor job (as indicated by a windscreen that broke) had to replace the windscreen without receiving wages for it. As a result of this policy, Lazear actually identified a slight improvement in quality after the incentive scheme was introduced.

Other studies by Paarsch and Shearer (2000), Shearer (2003) and Bandiera, Barankay and Rasul (2005) all suggest private sector firms in similar industries (tree planting and fruit picking) witness gains of a similar order of magnitude to those found by Lazear (2000).

## *Civil Servants*

### **Summary**

- All the studies examining incentive schemes for civil servants evaluate programmes that reward work at group or team level.
- There is strong evidence that civil servants do respond to financial incentives, even when benefits are not directly salary enhancing.

- One UK study suggests that team incentives might be more effective in smaller groups in which performance can be easily monitored by colleagues or peers.
- In several US studies, strategic behaviour was identified amongst civil servants in incentive programmes. One paper suggests that this kind of behaviour is more likely when performance is measured inaccurately or less accurately relative to other performance measures.
- UK evidence comes from programmes which were only in place for a limited time period of one year or less. The short duration of these schemes makes it unlikely that the novelty of incentives will have worn off. The long term response to incentives may not, therefore, be the same as short term impacts suggest.
- Most studies have not considered whether financial incentives represent good value for money overall, although one US paper provides evidence that strategic behaviour by civil servants was placing identifiable costs on society. The full welfare implications of financial incentives for civil servants, therefore, remain unclear.
- One further complication is the need to identify appropriate performance incentives which let civil servants prioritise tasks in the “right” way. This problem is particularly acute for the police and armed forces.
- The judiciary is one group for whom financial incentives are traditionally considered inappropriate.

### **The US JTPA programme**

A rich source of evidence comes from the US Job Training Partnership Act (JTPA), a federal employment and training programme which promotes desirable employee conduct using agency level financial incentives rather than outright regulation. In this scheme, the US Department of Labor awards bonuses directly to the training centres and these bonuses cannot be used to supplement employee salaries. Thus while socially motivated employees may obtain benefits from better facilities and observing higher staff numbers, individual performance is not rewarded directly. The actual size

of the reward depends on an agency's performance according to a range of measures, which have varied over the life of the programme and from State to State. Performance was originally measured in terms of cost per participant as well as participant wage rates, employment status, and earnings up to three months after they graduated. Bonuses have augmented the operating budget of agencies by 7% on average (Courty and Marschke 2004).

Common findings from empirical studies of the JTPA programme are summarised in Barnow and Smith (2004). They first find mixed evidence overall of 'cream-skimming', the practice of selecting applicants most likely to help agencies meet their performance targets. Heckman and Smith (2004) look at 16 agencies and, by decomposing the selection process, find that lack of awareness of scheme eligibility rather than cream-skimming by agency staff is responsible for driving most of the observable differences between eligible candidates and selected participants. Examining one agency in detail, Heckman, Smith and Taber (1996) find evidence that agents are deliberately choosing *less* employable candidates despite the presence of performance related financial incentives. Barnow and Smith (2004) suggest that regression-based methods to adjust the performance measures used in the schemes for case mix have been somewhat successful at reducing the incentives to cream-skin attractive applicants.

Second, in contrast to the inconclusive results on cream-skimming, there is substantial evidence of strategic behaviour, in which agents improve measures of performance without increasing their actual performance. Courty and Marschke (1996, 1997 and 2004) look at the choice of termination date made by agency workers, on which day the key performance measurements are taken. Under the scheme, agents can choose any date up to 90 days from the end of training to terminate a participant and record their employment and salary details. Since bonuses are awarded on the basis of annual data, agents can engage in strategic reporting to spread good and poor performance over multiple years. This behaviour rewards agencies because bonus payments are target based, with no additional benefits associated with exceeding targets.

In their study based on 16 agencies, Courty and Marschke (2004) first show that agencies delay for longer the termination date of programme participants if they remain unemployed throughout the 90 day period. Secondly, they identify an annual pattern in their results, in which unemployed participants are terminated earlier in the last three months than those finishing within the first 9 months of the programme year. Then thirdly, looking at their findings for the last three months alone, they notice that both very good years and bad years are associated with earlier programme terminations for unemployed participants. This is what we would expect to find if agents were engaged in strategic behaviour, as agencies which are either unlikely to reach their target performance, or have generously exceeded it, have nothing to lose by terminating unemployed participants within the current programme year.

Whether this strategic behaviour imposes welfare costs on service users and other stakeholders is also examined. Courty and Marschke (2004) compare the earnings of programme graduates from agencies with high and low levels of strategic behaviour and find high levels are associated with lower graduate earnings. Further, they observe that courses in June - the final month of the programme year - are more likely to be truncated than in other months, and that those participants taking June courses generally show a smaller wage impact from the programme. This is all evidence that strategic behaviour in this environment, far from simply being an accounting issue, is also imposing real costs on society as well.

### **International evidence**

Some of the strongest evidence that financial incentives can influence the output of civil servants comes from Khan, Silva and Ziliak (2001), who evaluate the introduction of a performance pay scheme within the Brazilian Tax Collection Authority. The scheme, initiated in 1989, rewarded both individual and group effort using objective measures of relative efficiency including the number of inspections made and the amount of fines collected from tax evaders. This was supplemented by subjective performance appraisals for individuals. The scheme paid bonuses equal to 70% of additional fines collected, and as a consequence, substantial bonuses reaching twice the mean annual salary became common. That this bonus scheme was so high powered makes this case study relatively unique.

After netting out region specific characteristics, Khan et al. (2001) evaluate the impact of the scheme from data covering three years before and three years after performance pay was implemented. Overall, they estimate that the bonus scheme led to a 75% increase in the rate of growth of fines per inspection. At the same time, they also identify substantial regional variation, with responses ranging from 19% to 145%. The authors suggest that diverse management techniques resulted in some regions targeting wealthier sources (such as corporations), more aggressively. This prompts Khan et al. to issue a warning that high powered incentive schemes such as this risk encouraging extortion in their workforce and may have significant social costs as well as clear financial benefits. Unfortunately, limited data prevents them from examining these social costs further.

### **UK evidence**

Here, the evidence is limited to only a handful of studies. Burgess et al (2004) examine the impact of an incentive programme piloted in Jobcentre Plus, a UK government employment and benefits agency. Like the US JTPA programme, the UK Job centre pilot was team-based, with performance assessed at team level. In all, 90 teams took part, consisting of between five and 39 offices ranged in size from 264 to 1535 people. Unlike the US scheme however, bonuses were paid to workers directly, with each team member receiving a 1% increase in salary for each performance target met. Meeting all five quality and quantity targets was associated with an additional 2.5% increase in salary, so workers in a high performing team could benefit from a salary rise of 7.5% in total. Crucially, quantity was measured in terms of job entry points (JEP) achieved, with for example, the successful placement of a jobless lone parent attracting twelve points, compared to one point for an already-employed service user. Further points were awarded if the service user was still in employment after four weeks.

Burgess et al. (2004) focus on three aspects of the Jobcentre Plus incentives scheme: first, the extent to which financial incentives influence the behaviour of public sector workers; second, the consequences of designing an incentive scheme that rewards team rather than individual effort, and finally, whether a wide variation

in the accuracy of performance measures can influence the behaviour and effort of workers in the public sector.

Looking at job placement performance, Burgess et al. find that while the overall response to incentives was close to zero, team size made a significant difference, with 'very small' and 'small' teams showing 10% increases in output compared to evidence of reduced output for 'large' teams. They interpret this as suggesting that peer performance monitoring, more easily done in smaller teams, is effective at reducing the problems associated with team-based incentives. Furthermore, for smaller teams, the small additional cost of the incentive scheme compares favourably with a general pay rise or an increase in staff numbers. Splitting performance into quantity and less easily measured quality components, the scheme appeared to result in improvements in the quantity but not the quality of services offered. This suggests that measurement accuracy does indeed have an important impact on worker's strategic behaviour in this context.

In another UK study, Burgess et al. (2005) evaluate two different team-level financial incentive schemes in operation at HM Customs and Excise during 2002 and compare their performance to a randomly selected group that was not involved in either bonus scheme. Both teams were assessed on their performance in identical tasks, included auditing work and tax revenue collection. One team had bonus payments that paid a fixed percentage of officers' salary (ranging from £550 to £1520) while the other team had flat rate bonuses which were worth £740 for each individual. Teams were required to meet all non-incentive targets in addition to further targets on both key performance indicators in order to receive their respective bonuses.

The main aim was to investigate whether and how individuals changed their behaviour in response to team-based incentive schemes. Overall, Burgess et al. (2005) find that officers of both teams worked harder over the period than the randomly selected control group. However, only team two (the fixed bonus team) succeeded in hitting all its targets. Seeking to explain this difference, Burgess et al. find evidence that managers in team two allocated more incentive tasks to efficient officers (who were among the top 25% most productive workers before the scheme was initiated)

than managers in team one. Whether the flat bonus structure led to the strategic behaviour by managers in team two is not addressed. Hence, while all officers in both teams responded positively to the incentive programme, the team facing a uniform fixed bonus also worked smarter as a team, strategically deploying efficient officers where their effort was best rewarded.

### **Police – a special case**

Police forces represent one government bureaucracy in which it is particularly difficult to identify appropriate performance incentive measures. Providing performance incentives for some of the measurable outcomes that do exist could have unintended consequences which might reprioritise police tasks against the wishes of local communities (see FitzGerald et al. (2002) cited in Vollaard, 2003). Indeed, Prendergast (2001) offers evidence from the Los Angeles Police Department that increased performance measurement and public oversight on their own undermined officer's willingness to make arrests, which in turn led to a significant increase in crime rates.

Baicker and Jacobson (2007) examines the police response to financial incentives introduced by the US Department of Justice. Initiated in 1984 as part of the Comprehensive Crime Control Act, the programme provided police departments the opportunity to share a proportion of the proceeds of drug-related asset seizures. Actual proportions were determined by State governments, resulting in a wide variation in the fraction returned to police departments. Netting out differences in county characteristics, Baicker and Jacobsen estimated a 10% increase in the fraction returned to police departments is associated with a \$0.19 per capita increase in the value of seizures police made. States which returned almost all drug related asset seizures added a further \$0.09 per capita to this figure. Police departments retaining a higher fraction of their seizures also worked more strategically, placing greater emphasis on possession offences than drug sales, which are less likely to include a significant cash component on arrest.

While this particular scheme has been relatively successful, it remains highly unlikely that this programme design can be replicated across the full range of police duties. A further problem arises from the difficulty of finding accurate quality

indicators in order to prevent officers from meeting targets without raising productivity. Changing regional and national priorities and the need for detailed local knowledge also implies that targets will quickly become inappropriate. Subjective, qualitative performance appraisals may therefore be more effective in this context.

### **Armed Forces**

The armed forces encounter many of the same measurement problems faced by police forces. Partly as a response to this, military forces have developed a system in which promotion confers greater status and privileges, (in addition to the higher salaries and allowances typically found in civil employment). These non-financial rewards provide substantial performance incentives on their own, for as long as individuals remain employed in the armed forces. Even so, the main purpose of the promotion system in place continues to be aimed at identifying personnel able to perform well in roles associated with higher levels of responsibility, as opposed to one explicitly designed to reward performance. Hence, the link between performance and promotion is not necessarily a straightforward one – promotion depends upon being given sufficient opportunities to demonstrate ability including educational background and command responsibilities, in addition to ability itself.

Under the current system, high performing individuals have the opportunity to rise up the ranks quicker than average performers, while consistently poor performers are more likely to be discharged under an “up or out” structure. Increasing the extent to which pay is linked to rank rather than years of service is one way of making the standard armed forces reward programme more effective at providing performance incentives. Since the chances of promotion falls as military personnel approach retirement age, the incentive effect associated with promotion is also likely to fall unless the incremental increases in pay associated with the highest ranks are large enough to counter the increased risk. Thus, while no evidence of a performance related pay schemes exists for armed forces, elements of a merit pay programme are present or can be made more explicit within the current compensation framework. Asch (1993) and Kosiak (2005) discuss current issues in US military pay.

## **Judiciary**

Financial compensation policy for the judiciary is designed to remove all external pressures in order to allow judges to act independently. This independence is a vital component of any democratic political system, and any compensation policy that adds external pressures could, therefore, be inappropriate. Addressing the US system, Choi and Gulati (2004) propose the creation of a set of objective performance measures which can then be used to rank judges in order to identify candidates suitable for promotion – up to Supreme Court level. This induces incentives for judges to improve their performance ranking. Their justification for such a tournament is that the present system is excessively politicised and improved ways of measuring and considering judicial quality need to be considered. Examples of possible performance criteria include the publication rate of opinions, their citation rates by other courts, citation rates by the Supreme Court, dissent rates, case productivity etc. It is possible that equivalent objective performance measures do exist for the UK, but identifying appropriate measures and their weightings could be difficult and any strategic behaviour by judges other than productivity improvements in judicial quality could have serious consequences.

Ramseyer and Rasmusen (1997) look at all Japanese judges who started work between 1961 and 1965 in order to identify key determinants of career success and the extent of political influence in the Japanese judicial system. The Secretariat grades the productivity and quality of all judges and uses this performance information to allocate jobs of greater or lesser prestige and pay. The system enables the Secretariat to reward ability and provide performance incentives for judges who begin their careers unproven and at a comparatively young age. While Ramseyer and Rasmusen do not examine whether judges respond to financial incentives, they do find evidence that judges responded to some incentives which were politically biased. The incumbent political party was, therefore, able to use the system in place to influence the judiciary without overtly intervening in the decision-making process.

## *Healthcare Workers*

### **Summary**

- Improving quality is the primary focus for many, if not all, performance pay schemes implemented in the healthcare sector. This stands in contrast to civil servants' financial incentive schemes, which attempt to improve the quality **and** quantity of output.
- The evidence that doctors respond to financial incentives is inconclusive. However, much of the US evidence comes from incentive schemes which were only in place for a limited period and in nearly all cases, the financial rewards involved were small.
- There is limited evidence from US studies that a clinician's response to incentives depends upon the absolute and relative size of the financial incentives as well as the relative effort required to obtain them.
- UK performance measures suggest that the quality of healthcare delivery from GPs is generally very high. GPs have therefore been very successful in obtaining financial bonuses under the current performance pay scheme. However, this is against a backdrop of previously high levels of performance from GPs – so the extent to which this scheme has been successful is unclear.
- Both of the UK studies looking at GPs' response to incentives identify a small but significant proportion of doctors who may be deliberately misreporting performance measures to increase their financial rewards.
- The analysis from one of these UK studies also suggests that GPs are motivated by concerns for patient health as well as financial incentives.
- Unfortunately, the short duration of many incentive programmes means evidence on cost effectiveness is scarce. Care must, therefore, be taken when equating higher performance measures with actual improvements in the health of society.

## **US evidence**

A wide ranging US study by Lindenauer et al. (2007) evaluated a pilot programme of financial rewards based on rankings. This started in 2003 for two years, operating alongside a public reporting programme. Both programmes were developed to push through improvements in healthcare quality. The authors compare 207 hospitals which participated in both schemes to 406 others which have matching characteristics but were only taking part in the public reporting programme. In this programme, hospitals are ranked according to their absolute performance in 33 indicators of quality (as a percentage of patients received for that condition). Hospitals in the top decile overall received a two percent budgetary supplement while those in the second decile received one percent. Hospitals with fewer than 30 cases of a particular condition annually were excluded from related quality indicators. The scheme was partially funded by financial penalties for poor performing hospitals.

After taking into account differences in initial performance levels and volume, Lindenauer et al. (2007) attribute the financial incentive scheme with achieving improvements of 2.6% and 4.1% above that of non-incentive hospitals over the whole 2-year period. Furthermore, they found some evidence that the largest improvements were being made by what were, initially, the poorest performing hospitals. These results suggest first, that healthcare staff do respond to indirect, quite modest, financial incentives. Second, a scheme of financial incentives and penalties can elicit improvements in reported quality from hospitals which are already publicly reporting their performance indicators. Finally, they note that this scheme saw measured improvements in quality indicators for both low and high performing hospitals, even though bonuses were only paid to the top 20% of hospitals. However, the extent to which these results can be generalised for all US hospitals is limited by the fact that 1) hospitals in the non-incentive group chose to opt out from the study (possibly introducing some bias to the results) and that 2) the characteristics of the incentive scheme-hospitals are not typical of US hospitals as a whole.

Much more common than rank-based incentive schemes in the healthcare sector is performance targeting. Under targeting, performance is rewarded according to a hospital's or practice's ability to meet pre-defined thresholds. Rosenthal and

Frank (2006) conducted a survey of the US empirical literature on performance pay in late 2003. They identified one rank-based incentive scheme and six performance targeting programmes. Their general conclusion is that, as of 2003, empirical evidence that financial incentives can improve the quality of healthcare is weak. In particular they are critical of the research design of a number of studies which find significant positive effects of financial incentives. However, Rosenthal and Frank also acknowledge that bonuses may have been too small (an incentive scheme evaluated by Hilman et al. (1998) awarded top level bonuses of only \$1260 per site) and that the presence of multiple contracts placed too many competing demands on physician's work. Conditions are very different in the UK, where hospitals have fewer contracts.

A later survey of the empirical literature on performance pay in health care was carried out in November 2005 by Petersen et al. (2006). Comparing the results of 17 different US studies, they find tentative evidence that pay for performance schemes have had positive effects. As with Rosenthal and Frank (2006), however, they identify limitations to the literature which prevents them from drawing anything other than preliminary conclusions. These conclusions are:

- Financial incentives often induce better book keeping and documentation methods than actual increases in medical provision (which suggests poorly designed performance targets).
- Incentives are more effective when the potential reward is larger and the payment frequency is higher.
- Threshold based systems are failing to reward organisations which improve the most because they often have the lowest baseline performance.

Rosenthal et al's (2005) evaluation of a group-level physician threshold scheme operating in California provides an example of this type of reward distribution problem happening in practice. The QIP scheme, initiated in 2002, offers financial bonuses for each target of \$0.23 per member per month if the physician group exceeded the performance achieved by the 75th percentile in previous year. The average physician group had about 10,000 continuously enrolled members, but even so, a perfect performance in all 10 quality targets would only yield a bonus equal to 0.8% of an average group's annual revenue. The 163 practices in California were

compared over time to 42 similar practices in Oregon and Washington State which did not take part in the scheme. Looking at three quality targets for which they have pre and post programme data, Rosenthal et al. find a statistically significant improvement (a 3.6% difference) for just one of the three performance measures. Looking at which groups yield the highest bonuses, they find that 75% of bonuses have been awarded to physician groups which were initially performing above target levels, even though these groups had shown the smallest improvement.

Later on in January 2003, these same Californian practices were submitted to a new, parallel incentive scheme with approximately 10 times larger financial incentives (the HIA programme). This new HIA scheme cost six US health plans \$122.7m in 2004 and \$139.5m in 2005. Mullen et al. (2007) evaluate the impact of both schemes together using data before and after the schemes were introduced (July 2001- July 2005), comparing their results to Washington and Oregon practices which do not participate in either programme. They also examine quality indicators ignored by both schemes.

Their results show a positive response to some of the rewarded performance measures which do not demand significant additional doctor time/effort. Examples of these include cervical cancer screening and childhood immunisation rates - indicators which rely on identifying high risk groups and scheduling appointments appropriately. Indeed, for these indicators clinicians respond more positively under larger incentives. In contrast, the prescription of preferred antibiotics, which is only rewarded by the weaker QIP scheme, shows significant declines of between 2.5% and 10% when the HIA program is introduced. Among unrewarded performance measures, the response was more mixed, with two showing above trend improvements while a third indicated falling performance. Despite some variation in quality of care indicators in response to incentives, Mullen et al. (2007) do not find any evidence to suggest that overall patient health has suffered from the introduction of the QIP and HIA incentive programmes, but they do question whether the scheme was cost effective, given mixed and unpredictable results.

## **UK evidence**

2004 witnessed the introduction of a UK-wide financial incentive programme for general practitioners which was both broad-based and high powered. The government committed £3.2bn to the programme in its first year, potentially augmenting general practice (GP) budgets by 25% on top of their core funding allocations. The programme rewards the performance of family practitioners based on 146 quality indicators, measured at practice level. These indicators cover various aspects of clinical care, practice organisation and patient experiences. Rewards are weighted for difficulty, and are awarded on a sliding scale of achievement up to a maximum target, above which the bonus stays fixed. In this way, the UK pay for performance programme does attempt to address some of the reward distribution issues identified by Rosenthal et al. (2005).

Even before the introduction of the scheme, there is evidence that physicians were responding to quality initiatives. Campbell et al. (2005) follow three quality-of-care measures (coronary heart disease, asthma, and type-2 diabetes) for 42 general practices between 1998 and 2003. Using patient level data aggregated at practice level, they observe significant improvements for 22 of the 50 quality indicators relating to these conditions. Possible explanations for this improvement include advances in data recording techniques, as well as an expansion of public reporting initiatives focusing on quality. Campbell et al. predict high levels of performance measured under the new scheme, and that has since been confirmed in studies by Doran et al. (2006) and Gravelle, Sutton and Ma (2007).

Sutton and McLean (2006) examine a selection of performance measures for 60 clinical practices in north Scotland. These practices scored on average 612 points out of a maximum 650 points in 2004. Analysing which factors may be influencing quality measures they find that size and team composition are the most significant determinants of performance. According to the study, practices with more than four full-time GPs (clinicians) and practices with younger staff obtained higher quality scores. Thus larger practices performed better, a result which is seemingly unrelated to the proportion of nursing staff at a practice. Confounding concerns (and evidence from Gravelle, Sutton and Ma, 2007) that practices in more deprived areas might

suffer under the new system, those serving more materially deprived populations actually achieved higher scores. Former fund holding practices and training centres also scored better.

Two studies actually evaluate the effects of the performance programme on doctor behaviour. Doran et al. (2006) focus on 76 clinical quality indicators for 8105 GP practices (95% of practices) in the UK. They observe that the average practice earned an additional £76,200 from the scheme for the 2004-2005 reporting year, with 230 practices achieving the maximum bonus. In contrast to Sutton and McLean (2006), analysis by Doran et al. suggests that smaller practices perform better, although larger practices perform more consistently. They do, however, both agree that practices with younger doctors achieve better scores. In addition, Doran et al. consider rates of exception reporting for special cases and conditions. Doctors can report an exception for special cases, when the appliance of performance targets might lead to inappropriate treatment. They observed that levels of exception reporting were low across the UK (with a median rate of 6%), but this included 91 practices (1.1%) which reported exceptions for more than 15% of their patients. Furthermore, exception reporting was also an important determinant of performance, with 1% higher rates associated with an increase in measured performance of 0.31%. A possible explanation for this result is that these practices have been using exception reporting to obtain higher bonuses without actually improving outcomes, referred to as 'gaming' behaviour in the economics literature.

Gravelle, Sutton and Ma (2007) investigate the effects of exception reporting further in their study, which examines a sample of 916 Scottish GP practices under the same pay for performance programme. The authors note that the incentive programme measures performance as a ratio of patient treatment divided by condition prevalence. So a practice performing below the maximum threshold in any one quality indicator can increase their reward bonus by:

- Increasing the rate of exception reporting,
- Reducing reported levels of condition prevalence in the local population or
- Increasing the amount of patients appropriately treated.

Perversely, under the NHS incentive scheme, over-performing practices also have an incentive to report higher prevalence rates. Gravelle, Sutton and Ma follow 65 clinical quality indicators for each practice over the 2004/2005 and 2005/2006 reporting years in an attempt to identify any changes in doctor response to incentives.

Their results show that some practice characteristics, such as the proportion of female clinicians or younger clinicians were positively associated with higher quality scores and higher prevalence rates, a somewhat unexpected result. The authors also calculate that GPs chose to treat 12.4% more patients than were necessary to maximise their financial rewards, suggesting clinicians are motivated by more than just financial incentives. However, in other analysis, GP practices that exceeded maximum performance thresholds during the first year are significantly more likely to increase reported prevalence rates the following year. Practices which performed below the maximum thresholds in 2004/2005 are also found to have increased their exception reporting for 2005/2006. These results suggest that, while most doctors are behaving quite altruistically under the new NHS incentive scheme, a small proportion of doctors have been taking advantage of the system, reporting higher performance than was actually delivered in order to enhance their financial reward.

## ***Teachers***

### **Summary**

Incentive programmes in the education sector can typically be divided into three categories:

1. Merit pay schemes reward pupil performance on a subjective basis at individual teacher level. Performance evaluations are often undertaken by school head teachers or (trained) external observers and financial rewards are allocated on perceptions of teaching quality.
2. Performance related pay schemes reward individual teacher performance based on objective measures of pupil performance. This includes specific indicators such as measurements of pupil progress and pupil performance in end-of-year examinations.

3. Finally, school level incentive programmes distribute bonuses to schools on the basis of performance measured at school level. Depending on the scheme, school performance can be measured in absolute terms - often against a performance threshold - or relative to other schools (although the latter requires adjustments to account for differences between schools).
  - There is strong evidence that teachers do respond to financial incentives. Several studies suggest that this response does not universally affect all students: most of the improvement appears to come from previously weak students performing better under such schemes.
  - All studies suggest that directly rewarded outcomes improve under school and teacher level incentive schemes. The evidence on unrewarded outcomes is, however, inconclusive.
  - One study suggests that subjective, teacher-based merit pay schemes tend to be much more effective if rewards are distributed selectively. Another study comparing the impact of a group versus an individual level performance pay scheme finds that the individual level scheme is more cost effective.
  - Nearly all studies evaluate a sample of schools which are unrepresentative or specially selected to take part. This means that care must be taken in applying the results of studies more generally.
  - From a welfare perspective, there is still little understanding of the processes of change within schools that have adopted a financial incentive scheme, and subsequent effects on staff morale. Neither is there an understanding of the full costs and benefits of these schemes.

### **US evidence**

Owing to a substantial amount of freedom over compensation policy at State level, there are US examples for all three types of teacher incentive scheme. In a cross-sectional study, Figlio and Kenny (2006) examine the relationship between school

performance and the presence of a merit pay scheme for 502 schools across the US. Using a rich dataset of school and pupil characteristics to net out any pupil based variation between schools, they are able to associate the presence of one or more merit pay schemes at a school with a 1.3 to 2.1 point increase in test scores. While this is a small effect in absolute terms, Figlio and Kenny's analysis suggests that the presence of a merit pay scheme has three times the impact that maternal education has on pupil performance (a widely recognised determinant of education). Importantly, they obtain this result by narrowing their definition of a merit pay scheme to more selective programmes only. Indeed, when they expand their definition to include both selective and non-selective merit pay schemes, they do not observe any significant impact. This suggests that only relatively selective merit pay schemes are correlated with higher pupil performance. Unfortunately, the cross-sectional nature of their dataset prevents Figlio and Kenny from proposing anything more than associations and correlations.

One of the most sophisticated US programmes was Tennessee's Career Ladder evaluation system, which rewarded performance, based upon a range of sources and evaluations, with progress up a rigid career ladder. Each new level was associated with salary rises and professional development opportunities. Achieving level I would have added \$1000, while a level III teacher would have received a salary supplement of \$7000. Dee and Keys (2004) evaluate the Tennessee merit pay scheme at the same time as a separate experiment was randomly allocating pupils and teachers to their class types. As the assignment of students and teachers to classes was random, this should eliminate any significant correlation between teacher type and student ability which would invalidate Dee and Key's results. The authors compare the performance of pupils in classes taught by merit pay-based teachers with those taught by teachers not participating in the scheme.

Using data from a representative selection of 79 state schools over 4 years from 1985, Dee and Keys (2004) find that pupils taught by career ladder teachers had 3% higher math scores than pupils taught by non participating teachers. Compared to other factors affecting pupil performance, this increase is equivalent to between 40% and 60% of the gains associated with small class sizes. Their results for math were

statistically significant, but this was not the case for reading scores even though they were 2% higher for merit scheme teachers. Further analysis shows that higher pupil achievement in math was particularly associated with merit pay teachers on the lower ranks of the Tennessee career ladder. Dee and Keys conclude that while the merit pay scheme did significantly improve performance in maths, it is often opposed by teacher unions because of the difficulty in accurately assessing teacher quality.

Eberts, Hollenbeck and Stone (2002) compare the outcomes of a mixed merit/pay for performance scheme implemented at one Michigan high school to another school with similar characteristics which maintained a fixed compensation policy. The mixed programme paid a 12%-12.5% bonus quarterly to teachers who successfully retained more than 80% of students at the end of the quarter. Furthermore, on the basis of very high performance (averaging 4.65/5) in student evaluations of teaching quality over the year, teachers could become eligible for a 5% salary supplement and an extra 10% increase in their retention bonus.

The authors examine changes in attendance rates and exam performance using data from both the incentive school and the non-incentive school before and after the mixed programme was initiated. In the school which implemented the mixed incentive scheme, Eberts, Hollenbeck and Stone find a 40% increase in retention rates after the programme was introduced. However, for the incentive school, they observe a 0.53 point drop in exam performance over the whole period, which is greater than the 0.37 point decline measured in the non incentive school. This suggests that the incentive programme resulted in the retention of more low-achieving students. This kind of response is, therefore, strong evidence that teachers do respond to directly rewarded performance measures, but it is also a warning that the effects of unintended consequences may mean that even closely related outcomes can suffer if not rewarded directly.

A US example of a group level incentive scheme comes from Dallas, Texas. The Dallas scheme issued bonuses to the top 20% of schools ranked on the basis of pupil improvement. Compared to some other group level incentive programmes, the Dallas scheme was relatively sophisticated:

- Choosing a wide range of objective performance measures including drop out rates limited the scope for teachers to concentrate all their efforts on a single exam.
- The programme tried to reward effort equally by netting out many socio-economic factors that could affect pupil's learning ability.
- Once these other factors have been taken into account, schools were ranked according to the overall improvement that their students had made across all subjects, compared to the previous year.

Dallas issued \$1000 bonuses to each teacher and manager in the top 20% of schools and \$500 bonuses to each manual worker. The scheme also paid a \$2000 activity budget supplement to the school itself.

Ladd (1999) measures the impact of the scheme from its inception in 1991 for four years, looking specifically at changes in the math and reading pass rates at Dallas schools compared to similar schools in nearby cities. Pass rates are a relatively strong measure of student outcomes because school performance was initially very poor. Compared to pass rates in other unreformed city schools, Ladd finds reading outcomes improve between 10% and 15% more in Dallas schools over the 4 year period, while for maths, the improvement ranges from 14% to 17%. Interestingly, results appear to rise over time. Ladd then examines the pupil performance of specific racial groups and finds that nearly all the improvement comes from Hispanic and white student scores. Black student scores do not show any significant response to the incentive scheme, and Ladd's results are unable to identify any reasons for this.

### **Evidence from Israel**

A similar school level scheme that, like the Dallas programme, rewarded relative improvements in student performance was launched in some 62 religious and secular Israeli schools in 1995. The scheme distributed a reward budget of \$1.4m to the top third of schools in rank order. Once again, performance was measured by a wide range of criteria including dropout rates. In contrast to the Dallas scheme, however, a school's rank determined the absolute size of the bonus awarded. Once awarded, the scheme required successful schools to distribute 75% of their bonus to all teaching

staff proportionate to their salaries, while the other 25% remained with the school. As an example of the size of rewards available, bonuses in the 1997 year ranged from \$13,250 to \$105,000 per school. Using a detailed dataset, Lavy (2002) measures the impact of the scheme on pupil achievement and also compares this scheme to a resources-based programme that provided 22 schools with \$1.2m of additional teaching resources.

Lavy finds evidence of significantly higher student achievement in secular schools two years after the programme was introduced, while the results from religious schools suggest a positive and significant impact after just one year. Indeed religious schools appear to have responded better to incentives in both years. Two years after the programme was introduced, average subject scores improved by 1.75 points for secular schools but by 3.1 points for religious schools. Dropout rates, also included in the incentive scheme, fell by 0.5% after two years for both secular and religious schools, reaching 5.5% and 3.2% respectively in 1997. Lavy next examines the compositional background of students who demonstrate the greatest improvement after the incentive scheme was introduced. He found that while the group incentive programme had little overall effect on the proportion of students passing the key matriculation certificate, those from poorer families observed a large performance improvement. Finally, compared to the incentive scheme, Lavy estimates that the impact of the resources-base programme was 50-70% greater in magnitude for credit units achieved, average pupil scores and the proportion of students taking the final matriculation exam. However, on other measures including science credits achieved, dropout rates and matriculation pass rates, he found that the resources programme elicited no significant response, in contrast to the incentive programme. Weighing up the impact of both schemes, their relative costs, and the number of schools involved, Lavy declares the incentive scheme to be more cost effective.

Having identified a relatively successful group level incentive scheme, Lavy (2004) moves on to evaluate a teacher level incentive programme. 48 Israeli high schools participated, selected on the basis of previous poor performance. In this case, the size of reward for each teacher depended upon their class's improvement in exams. After netting out various socioeconomic pupil characteristics, all 207 English

teachers and 237 maths teachers taking part were ranked according to their class scores. Teachers with positive scores received a bonus, with successively higher quartiles eligible for larger financial bonuses. Teachers in the highest quartile received a bonus worth 25% of average teacher salary. To reduce the potential for strategic behaviour, students who did not take the matriculation exams were given scores of zero.

Lavy compares the performance of similar (propensity matched) students in non-incentive and incentive classes before and after the scheme was introduced in 2001. He observes an 18% (for maths) and a 10% (for English) increase in the number of exams attempted by students who were being taught by incentive teachers. Passes in these exams are vital to obtain a place at university in Israel. The programme resulted in 5.4% more students passing the pre-university matriculation exams in maths while the improvement for English was 4.2%. Breaking down the results for ability, Lavy finds nearly all of the impact of the scheme occurred within the lower two ability quartiles, the below average students. There is also some evidence that the scheme had a positive impact on the performance of students in subjects without incentives attached. For both history and biology (which were not linked to incentives), students appear to have taken more exams, but the pass rate has remained steady.

Looking specifically at the behavioural response of teachers to incentives, the results of a survey of maths and English teachers suggests that teachers increased after hours, unpaid help. In other analysis, tests intended to identify strategic behaviour by teachers manipulating the system or cheating, produce no conclusive evidence. Finally, Lavy estimates that this teacher level scheme was approximately twice as effective as the group level scheme evaluated in Lavy (2002), while costing \$100 less per student.

## **UK evidence**

In 1999, the UK government initiated a teacher level merit pay scheme for state school teachers. In some respects, the UK programme resembles the Tennessee Career Ladder programme evaluated in Dee and Keys (2004). Here, all teachers on the highest salary level were eligible, equal to approximately 75% of teachers.

Participants must pass an initial performance appraisal carried out by the school principal, during which both subjective and objective measures of performance are considered. Success at this stage was rewarded by a permanent £2000 salary increase and the opportunity to raise their salary further by passing additional performance reviews. Funded by a separate central government fund with no quotas or limits, success rates for the initial bonus amongst eligible teachers were very high in its first year (97% nationally). However, subsequent levels have introduced greater pay variation.

Atkinson et al. (2004) describe the details of the first introduction and evaluate its impact on an objective measure of pupil progress, "Value Added". Value Added shows the average improvement in learning outcomes for each teacher based on the improvement of their students in national exams taken at 14 and 16. Atkinson et al. (2004) follow 182 teachers teaching either maths, English or science at one of 18 UK schools. The sample has some differences from the average UK school due to necessary data demands. In an unexpected initial finding, they observe a substantial amount of variation in Value Added scores obtained by each teacher from year to year, almost comparable to average degree of variation between teachers. Their analysis, which attempts to net out any outside factors that affect student's learning ability, suggests that on average, the pupil of a merit pay teacher gains an additional half grade of Value Added above what is achieved by pupils of non merit pay teachers. They also find that most of the improvement in pupil progress is being achieved by below average students, a result that holds even when high achievers who only score the top grades (and hence hit an achievement "ceiling") are removed from the analysis. Finally, examining class assignment of pupils over time, they observe no evidence of merit pay teachers strategically picking high performing pupils.

Besley and Machin (2006) looks at the pay of UK head teachers and the performance of the schools they are associated with. They find that the market for head teachers creates a pay premium for effective leaders, which, as a direct consequence, imparts performance incentives on all head teachers.

## Conclusion

We find evidence that financial incentives can yield productivity improvements for some public sector workers. The evidence is strongest for civil servants and teachers but relatively weak for healthcare workers. Perhaps the main reason for the predominance of team based incentive schemes in the public sector is that employees may be concerned about the effects that individual level schemes could have on worker morale. Only one study has sufficient information to conclude that workers are less responsive to a team-based pay for performance programme than a scheme which rewards individual performance. There are also only a handful of studies which are able to suggest that smaller teams do appear to be more responsive to incentives than larger teams. Many of the individual level programmes implemented so far have rewarded performance relative to peers both inside and outside the immediate group. This has perhaps helped to limit any impact that such schemes might have on worker morale.

Some studies of civil servants and healthcare workers have found evidence of unintended consequences and strategic behaviour by workers. While there is evidence from one study that strategic behaviour does appear to impose real costs on society, the prevalence of this kind of behaviour may be more a reflection of bad programme design than a problem with financial incentives in the public sector per se. Using a mixture of subjective and objective performance schemes might help to reduce strategic behaviour in circumstances when output is difficult to measure.

Although some studies do attempt to make comparative value for money calculations, there are as yet no UK estimates of the full welfare cost of any incentive schemes in the public sector. One reason for this is the difficulty of measuring the impact of such schemes on recruitment. Another reason is that the period of analysis of nearly all public sector incentive schemes was too short: few schemes were in place for longer than a year.

Finally, while evidence from the UK public sector is limited, results from the UK generally support the conclusions of international studies. Nevertheless, care must be taken when generalising the results of individual studies across other countries or industries. Most empirical studies have typically relied on non-representative samples

for their results. Indeed, the individual nature of each scheme is important because incentive schemes should reflect the environment in which they are operating. Small design differences can have important consequences for the behaviour of employees, and all financial incentive schemes need to be carefully adapted to prioritise local needs.

## References

### *Theoretical Papers*

Acemoglu, Daron; Kremer, Michael; Mian, Atif (June 2003) “Incentives in Markets, Firms and Governments” NBER Working Paper 9802  
<http://www.nber.org/papers/W9802>

Akerlof, George A and Kranton, Rachel E (2005) “Identity and the Economics of Organisations” (2005) Journal of Economic Perspectives Volume 19, Number 1, Winter 2005, pp. 9-32(24)  
<http://www.ingentaconnect.com/content/aea/jep/2005/00000019/0000001/art00002>

Albano, Gian Luigi and Leaver, Clare (December 2004) “Transparency, Recruitment and Retention in the Public Sector” Mimeo, Research Centre for Economic Learning and Social Evolution <http://privatewww.essex.ac.uk/~itbara/pewg/tranpewg.pdf>

Asch, Beth (1993) “Designing Military Pay Contributions and Implications of the Economics Literature” RAND Monograph Report  
[http://www.rand.org/pubs/monograph\\_reports/MR161/](http://www.rand.org/pubs/monograph_reports/MR161/)

Asch, Beth J. and John T. Warner (2001) “A Theory of Compensation and Personnel Policy in Hierarchical Organizations with Application to the United States Military” Journal of Labor Economics, vol. 19, no. 3, 523-562  
<http://www.journals.uchicago.edu/cgi-bin/resolve?JOLE010701PDF>

Auriol, Emmanuelle; Guido Friebel, and Lambros Pechlivanos (2002) “Career Concerns in Teams” Journal of Labor Economics, volume 20 pages 289–307  
<http://www.journals.uchicago.edu/cgi-bin/resolve?id=doi:10.1086/338240>

Beaulieu, Nancy Dean and Dennis R. Horrigan (2005) “Putting Smart Money to Work for Quality” Improvement Health Services Research 40 (5p1), 1318–1334  
<http://www.blackwell-synergy.com/doi/abs/10.1111/j.1475-6773.2005.00414.x>

Benabou, Roland and Tirole, Jean (2003) “Intrinsic and Extrinsic Motivation” Review of Economic Studies Vol. 70, 489-520 <http://www.blackwell-synergy.com/links/doi/10.1111%2F1467-937X.00253>

Besley, Timothy and Ghatak, Maitreesh (2003) “Incentives, Choice And Accountability In The Provision Of Public Services” Institute of Fiscal Studies Working paper 03/08 <http://www.ifs.org.uk/wps/wp0308.pdf>

Bevan, Gwyn and Hood, Christopher. (2006) “What’s Measured Is What Matters: Targets And Gaming In The English Public Health Care System.” Public

Administration 84:3, 517–538 <http://www.blackwell-synergy.com/doi/abs/10.1111/j.1467-9299.2006.00600.x>

Casalino, Lawrence P. and Arthur Elster (2007) “Will Pay-For-Performance And Quality Reporting Affect Health Care Disparities?” Health Affairs, 26, no. 3 w405-w414 <http://content.healthaffairs.org/cgi/content/abstract/26/3/w405>

Chambers, Robert G., John Quiggin (2005) “Incentives and Standards in Agency Contracts” Journal of Public Economic Theory 7 (2), 201–228 <http://www.blackwell-synergy.com/links/doi/10.1111/j.1467-9779.2005.00201.x>

Choi, Stephen and Gulati, Mitu (Jan 2004) “A Tournament of Judges?” UC Berkeley School of Law [http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=394700](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=394700)

Delfgaauw, Josse and Dur, Robert (2006) “Incentives and Workers’ Motivation in the public sector” Tinbergen Institute Discussion Paper, Netherlands <http://ftp.tinbergen.nl/discussionpapers/04060.pdf>

Eggleston, Karen (2005) “Multitasking and mixed systems for provider payment” Journal of Health Economics Volume 24, Issue 1, Pages 211-223 [http://www.sciencedirect.com/science?\\_ob=ArticleURL&\\_udi=B6V8K-4DNRTWC-1&\\_user=121739&\\_coverDate=01%2F31%2F2005&\\_rdoc=1&\\_fmt=&\\_orig=search&\\_sort=d&\\_view=c&\\_acct=C000010018&\\_version=1&\\_urlVersion=0&\\_userid=121739&md5=31e99c2efcdb6576a687faa8f5e7cbc4](http://www.sciencedirect.com/science?_ob=ArticleURL&_udi=B6V8K-4DNRTWC-1&_user=121739&_coverDate=01%2F31%2F2005&_rdoc=1&_fmt=&_orig=search&_sort=d&_view=c&_acct=C000010018&_version=1&_urlVersion=0&_userid=121739&md5=31e99c2efcdb6576a687faa8f5e7cbc4)

Glazer, Amihai (2004) “Motivating Devoted Workers” International Journal of Industrial Organisation Vol. 22, No. 3 427-440 <http://www.mgmt.purdue.edu/centers/ijio/Accepted/2210.pdf>

Kaarbøe, Oddvar M., Trond E. Olsen (2006) “Career Concerns, Monetary Incentives and Job Design” Scandinavian Journal of Economics 108 (2), 299–316. <http://www.blackwell-synergy.com/doi/abs/10.1111/j.1467-9442.2006.00453.x>

Koch, Alexander K. and Eloic Peyrache (May 5, 2006) “Tournaments, Individualized Contracts and Career Concerns” Mimeo Royal Holloway and HEC, Paris. <http://www.bwl2.uni-bonn.de/tourncontr.pdf>

Lazear, Edward P. (2003) “Teacher incentives” Swedish Economic Policy Review Vol. 10 No. 2, p. 179-214 [http://www.ekonomiskaradet.se/Panda\\_ekonomiska/Data/Documents/sepr2003/Lazear.pdf](http://www.ekonomiskaradet.se/Panda_ekonomiska/Data/Documents/sepr2003/Lazear.pdf)

Leaver, Clare (2004) “Bureaucratic Minimal Squawk Behaviour: Theory and Evidence from Regulatory Agencies” Mimeo, Research Centre for Economic Learning and Social Evolution <http://www.nottingham.ac.uk/economics/research/seminars/papers/c.leaver-sqwk04.pdf#>

Posner, Richard A. (1995?) What Do Judges Maximize? (The Same Thing Everybody Else Does) [http://www.law.uchicago.edu/Lawecon/WkngPprs\\_01-25/15.RAP.Judges.pdf](http://www.law.uchicago.edu/Lawecon/WkngPprs_01-25/15.RAP.Judges.pdf)

Prendergast, Canice (2007) "The Motivation and Bias of Bureaucrats" American Economic Review, Vol. 97, No.1, 180-196  
<http://www.ecare.ulb.ac.be/ecare/seminars/2003-2004/papers%202003-2004/cp1912.pdf>

Ratto, Marissa, with Simon Burgess, Bronwyn Croxson, Ian Jewitt and Carol Propper, (June 2001) "Team-Based Incentives in the NHS: An Economic Analysis" CMPO Working Paper Series No. 01/37  
<http://www.bris.ac.uk/cmpto/workingpapers/wp37.pdf>

Schumacher, Johanna (September 2004) "A referee's report on the article 'Competition and Incentives with Motivated Agents' by Besley and Ghatak" Topics in Public Economics <http://www.nek.lu.se/NEKFRA/Topics/Assign04/besley.pdf>

Strunk, Bradley C. and Robert E. Hurley (May 2004) "Paying For Quality: Health Plans Try Carrots Instead Of Sticks" Center for Studying Health System Change Issue Brief No. 82 (Washington DC) <http://www.hschange.org/CONTENT/675/675.pdf>

Vollaard, Ben (2003) "Performance contracts for police forces" CPB Document No. 31 <http://www.cpb.nl/eng/pub/cpbreeksen/document/31/doc31.pdf>

### ***Empirical Papers***

Atkinson, Adele, Simon Burgess, Bronwyn Croxson, Paul Gregg, Carol Propper, Helen Slater, Deborah Wilson (2004) "Evaluating the Impact of Performance-related Pay for Teachers in England" CMPO Working paper Series No. 04/113  
<http://www.bris.ac.uk/depts/CMPO/workingpapers/wp113.pdf>

Baicker, Kate and M. Jacobson (2007) Finders Keepers: Forfeiture Laws, Policing, and Local Budgets, forthcoming in Journal of Public Economics, 2007 (also NBER Working Paper 10484) <http://www.nber.org/papers/W10484.pdf>

Bandiera, Oriana; Barankay, Iwan and Rasul, Imran (2005) "Social Preferences And The Response To Incentives: Evidence From Personnel Data" The Quarterly Journal Of Economics Vol. 120 No.3, pages 917-962

Besley, Timothy and Machin, Steven (2006) "Are Public Sector CEOs Different? Leadership Wages and Performance in Schools" Preliminary Draft - Not to be Quoted without the Authors' Permission, London School of Economics (October 2006)  
<http://www.ksg.harvard.edu/inequality/Seminar/Papers/Machin062.pdf>

Bognanno, Michael (2001) "Corporate Tournaments" Journal of Labour Economics, Vol. 19, No. 2 290-316 <http://ideas.repec.org/a/ucp/jlabec/v19y2001i2p290-315.html>

Borzaga, Carlo and Tortia, Ermanno (2006) "Worker Motivations, Job Satisfaction, and Loyalty in Public and Nonprofit Social Services" Nonprofit and Voluntary Sector Quarterly, Vol. 35, No. 2, 225-248 <http://nvs.sagepub.com/cgi/content/abstract/35/2/225>

Boyne, George A. and Chen, Alex A. (August 9, 2006) "Performance Targets and Public Service Improvement" Journal of Public Administration Research and Theory <http://jpart.oxfordjournals.org/cgi/reprint/mul007v1.pdf>

Burgess, Simon; Croxson, Bronwyn; Gregg, Paul and Propper, Carol (July 2001) "The Intricacies of the Relationship Between Pay and Performance for Teachers: Do teachers respond to Performance Related Pay schemes?" CMPO Working Paper Series No. 01/35

Burgess, Simon; Propper, Carol; Ratto, Marisa and Tominey, Emma (2004) "Incentives in the Public Sector: Evidence from a Government Agency" CMPO Working Paper Series No. 04/103 <http://www.bris.ac.uk/cmpo/workingpapers/wp103.pdf>

Chen, Qi and Wei Jiang, (June 2005) "Career Concerns and the Optimal Pay-for-Performance Sensitivity" Mimeo, Fuqua School of Business, Duke University and Columbia Business school [http://www.columbia.edu/~wj2006/career\\_concerns.pdf](http://www.columbia.edu/~wj2006/career_concerns.pdf)

Christianson, Jon B. PhD, David J. Knutson BA, Roger S. Mazze PhD (2006) "Physician Pay-For-Performance. Implementation and Research Issues" Journal of General Internal Medicine 21 (s2), S9-S13 <http://www.blackwell-synergy.com/doi/abs/10.1111/j.1525-1497.2006.00356.x>

Congressional Budget Papers (Dec. 1995) "Military Pay and the Rewards for Performance" <http://www.cbo.gov/ftpdoc.cfm?index=5307>

Courty, Pascal and Marschke, Gerald (2003) "Dynamics of Performance Measurement Systems" Oxford Review of Economic Policy, Vol. 19, No. 2 <http://oxrep.oupjournals.org/cgi/content/abstract/19/2/268>

Courty, Pascal and Marschke, Gerald (2004) "An empirical Investigation of Gaming Responses to Explicit Performance Incentives" Journal of Labour Economics, Vol. 22, no. 1, 23-56 <http://www.journals.uchicago.edu/cgi-bin/resolve?id=doi:10.1086/380402>

Crewson, P. E. (1997) "Public Service Motivation: Building Evidence of Incidence and Effect", Journal of Public Administration Research and Theory Vol.7, 499-519. <http://www.jstor.org/view/10531858/ap040028/04a00010/0>

Cullen, Julie Berry and Reback, Randall (May 2006) “Tinkering toward accolades: School gaming under a performance accountability system” NBER Working Paper 12286 [http://www.econ.barnard.columbia.edu/working\\_papers/wp0601.pdf](http://www.econ.barnard.columbia.edu/working_papers/wp0601.pdf)

Dee, Thomas S. and Benjamin J. Keys (2004) “Does Merit Pay Reward Good Teachers? Evidence from a Randomized Experiment” Journal of Policy Analysis and Management, Vol.23, No.3, 471 –488

Doran, Tim M.P.H., Catherine Fullwood, Ph.D., Hugh Gravelle, Ph.D., David Reeves, Ph.D., Evangelos Kontopantelis, Ph.D., Urara Hiroeh, Ph.D., and Martin Roland, D.M. “Does Merit Pay Reward Good Teachers? Evidence from a Randomized Experiment” Journal of Policy Analysis and Management, Volume 23, Issue 3 (p 471-488) <http://www3.interscience.wiley.com/cgi-bin/abstract/109072269/ABSTRACT>

Eberts, Randall; Hollenbeck, Kevin and Stone, Joe (2002) “Teacher Performance Incentives and Student Outcomes” Journal of Human Resources Vol. 37, No. 4 <http://www.jstor.org/view/0022166x/sp030007/03x0075w/0>

Figlio, David N. and Kenny, Lawrence (Issued in October 2006) “Individual Teacher Incentives and Student Performance” NBER Working Paper No. 12627 <http://www.nber.org/papers/w12627>

Frank, Richard G. and Meredith B. Rosenthal (2006) “What Is the Empirical Basis for Paying for Quality in Health Care?” Medical Care Research and Review, Vol. 63, No. 2, 135-157 [http://mcr.sagepub.com/cgi/content/abstract/63/2/135?ijkey=9461803f81d9768b92700f98eded99eeb9a9facf&keytype=tf\\_ipsecsha](http://mcr.sagepub.com/cgi/content/abstract/63/2/135?ijkey=9461803f81d9768b92700f98eded99eeb9a9facf&keytype=tf_ipsecsha)

Frey, Bruno S and Jegen, Reto (2001) “Motivation Crowding Theory” Journal of Economic Surveys, Vol. 15, No. 5, 589-611 <http://www.landecon.cam.ac.uk/speer/iewwp026.pdf>

Gauld, Robin (March 2007) ‘Principal-Agent Theory And Organisational Change’, Policy Studies, 28:1, 17 – 34 <http://dx.doi.org/10.1080/01442870601121395>

Gaynor, Martin, James B. Rebitzer, and Lowell J. Taylor (2004) “Physician Incentives in Health Maintenance Organizations” Journal of Political Economy, volume 112 (2004), pages 915–931 <http://www.journals.uchicago.edu/cgi-bin/resolve?id=doi:10.1086/421172>

Gielen, Anne C., Marcel J.M. Kerkhofs and Jan C. van Ours (July 2006) “Performance Related Pay and Labor Productivity” IZA Discussion Paper No. 2211 <http://ftp.iza.org/dp2211.pdf>

Glewwe, Paul; Ilias, Nauman and Kremer, Michael (November 2002) “Teacher Incentives” Mimeo, Brookings Institution, Washington DC and NBER Working paper No. 9671 <http://www.nber.org/papers/w9671>

Gravelle, Hugh; Matt Sutton and Ada Ma (2007) “Doctor behaviour under a pay for performance contract: Evidence from the quality and outcomes framework” University of York: Centre for Health Economics, Research Paper 28 <http://www.york.ac.uk/inst/che/pdf/rp28.pdf>

Grossbart, Stephen R. (2006) “What’s the Return? Assessing the Effect of “Pay-for-Performance” Initiatives on the Quality of Care Delivery” Medical Care Research and Review, Vol. 63 No. 1 [http://mcr.sagepub.com/cgi/content/refs/63/1\\_suppl/29S](http://mcr.sagepub.com/cgi/content/refs/63/1_suppl/29S)

Heckman, J., C. Heinrich, and J. Smith (2002) “The Performance of Performance Standards” The Journal of Human Resources 37: 778–811 [http://links.jstor.org/sici?sici=0022-166X\(200223\)37%3A4%3C778%3ATPOPS%3E2.0.CO%3B2-D](http://links.jstor.org/sici?sici=0022-166X(200223)37%3A4%3C778%3ATPOPS%3E2.0.CO%3B2-D)

Heinrich, C. (2002) “Outcomes-Based Performance Management in the Public Sector: Implications for Government Accountability and Effectiveness.” Public Administration Review 62: 712–725 <http://www.blackwell-synergy.com/doi/abs/10.1111/1540-6210.00253>

Heinrich, C. (2004) “Improving Public-Sector Performance Management: One Step Forward, Two Steps Back?” Public Finance and Management 4(3): 317–351 <http://www.pmrnet.org/conferences/georgetownpapers/Heinrich.pdf>

Heinrich, Carolyn J (2007) “False or Fitting Recognition? The Use of High Performance Bonuses in Motivating Organizational Achievements” Journal of Policy Analysis and Management Vol. 26 issue 2, 281–304 <http://www3.interscience.wiley.com/cgi-bin/fulltext/114177540/PDFSTART>

Jacob, Brian and Levitt, Steven (2003) “Catching Cheating Teachers: The Results of an Unusual Experiment in Implementing Theory.” Brookings-Wharton Papers on Urban Affairs p.185-209 <http://pricetheory.uchicago.edu/levitt/Papers/JacobLevittCatchingCheating2003.pdf>

Jensen, Paul H., Robin E. Stonecash (2005) “Incentives and the Efficiency of Public Sector-outsourcing Contracts” Journal of Economic Surveys 19 (5), 767–787. <http://www.blackwell-synergy.com/doi/abs/10.1111/j.0950-0804.2005.00267.x>

Jurges, Hendrik; Richter, Wolfram F. and Schneider, Kerstin (31<sup>st</sup> May 2005) “Teacher quality and incentives: Theoretical and empirical effects of standards on teacher quality” DIW Berlin, CESifo, Munich and IZA, Bonn [http://www.mea.uni-mannheim.de/mea\\_neu/pages/files/nopage\\_pubs/pxmghpz02xuj0nav\\_91-2005.pdf](http://www.mea.uni-mannheim.de/mea_neu/pages/files/nopage_pubs/pxmghpz02xuj0nav_91-2005.pdf)

Kahn, Charles M; Emilson C. D. Silva, James P. Ziliak (2001) “Performance-based Wages in Tax Collection: The Brazilian Tax Collection Reform and its Effects” The Economic Journal Vol. 111 No.468, 188–205 <http://www.blackwell-synergy.com/doi/abs/10.1111/1468-0297.00594>

Knot, Ondrej (March 2005) "The Role of a Mission in the Public Sector: The Case of Czech and English Judiciary" Mimeo, Charles University, Prague [https://www.cerge-ei.cz/pdf/dp/DP140\\_2005.pdf](https://www.cerge-ei.cz/pdf/dp/DP140_2005.pdf)

Kosiak, Steven M. (2005) "Military Compensation: Requirements, Trends and Options" Centre for Strategic and Budgetary Assessments, Washington DC. <http://www.csbaonline.org/4Publications/PubLibrary/R.20050218.Personnel/R.20050218.Personnel.pdf>

Ladd, Helen F (1999) The Dallas school accountability and incentive program: an evaluation of its impacts on student outcomes, Economics of Education Review, Elsevier, vol. 18(1), pages 1-16, February <http://ideas.repec.org/a/eee/eoedu/v18y1999i1p1-16.html>

Lavy, Victor (2002) Evaluating the Effect of Teachers' Group Performance Incentives on Pupil Achievement, Journal of Political Economy, University of Chicago Press, vol. 110(6), pages 1286-1317, December <http://ideas.repec.org/a/ucp/jpolec/v110y2002i6p1286-1317.html>

Lavy, Victor (2003) Paying for Performance: The Effect of Teachers' Financial Incentives on Students' Scholastic Outcomes, CEPR Discussion Papers 3862, C.E.P.R. Discussion Papers. <http://ideas.repec.org/p/cpr/ceprdp/3862.html>

Lavy, Victor (2007) "Using Performance-Based Pay to Improve the Quality of Teachers" Future of Children Vol. 17 No.1 Spring 2007 Princeton/ Brookings [http://www.futureofchildren.org/usr\\_doc/7\\_05.pdf](http://www.futureofchildren.org/usr_doc/7_05.pdf)

Lavy, Victor (Issued in July 2004) "Performance Pay and Teachers' Effort, Productivity and Grading Ethics" NBER Working Paper No. 10622 <http://www.nber.org/papers/10622>

Lazear, Edward P. (2000) "Performance Pay and Productivity" The American Economic Review Vol. 90 No.5 p.1346-1361

Levin-Scherz, Jeffrey; DeVita, Nicole and Timbie, Justin (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 Vol. 63; No. 1, Supplement S14-S28 [http://mcr.sagepub.com/cgi/content/abstract/63/1\\_suppl/14S](http://mcr.sagepub.com/cgi/content/abstract/63/1_suppl/14S)

Marsden, David and Belfield, Richard (2006) "Pay for Performance when output is hard to measure: the case for performance pay for school teachers" [online] LSE Research Online <http://eprints.lse.ac.uk/850/>

Marsden, David and Ray Richardson (1992) "Motivation And Performance Related Pay In The Public Sector: A Case Study Of The Inland Revenue" Centre for  
45

Economic Performance, London School of Economics, Discussion Paper No.75  
<http://cep.lse.ac.uk/pubs/download/DP0075.pdf>

Martin G. (1994) "Performance-related Pay in Nursing: Theory, Practice and Prospect" Health Manpower Management, Volume 20, Number 5, 1994, pp. 10-17(8)  
<http://www.ingentaconnect.com/content/mcb/039/1994/00000020/00000005/art00002>

Mas, Alexandre (Issued in May 2006) "Pay, Reference Points, and Police Performance" NBER Working Paper No. 12202 <http://www.nber.org/papers/w12202>

McElduff, P, G Lyratzopoulos, R Edwards, R F Heller, P Shekelle and M Roland (2004) "Will changes in primary care improve health outcomes? Modelling the impact of financial incentives introduced to improve quality of care in the UK" Quality and Safety in Health Care 2004 No.13: 191-197  
<http://qshc.bmj.com/cgi/content/abstract/13/3/191>

Metawie, Miral and Gilman, Mark (September 22-23, 2005) "Problems With The Implementation Of Performance Measurement Systems In The Public Sector Where Performance Is Linked To Pay: A Literature Review Drawn From The UK" Department of Industrial Relations, University of Kent at Canterbury  
<http://www.kent.ac.uk/kbs/crbp/pdf/Paper%20-%20Miral%20Metawie.pdf>

Mizala, Alejandra and Pilar Romaguera (2004) "School And Teacher Performance Incentives: The Latin American Experience", Center for Applied Economics Department of Industrial Engineering, University of Chile (March 2004) Published in the international journal of educational development  
[http://www.webmanager.cl/prontus\\_cea/cea\\_2003/site/asocfile/ASOCFILE22003041\\_0154027.pdf](http://www.webmanager.cl/prontus_cea/cea_2003/site/asocfile/ASOCFILE22003041_0154027.pdf)

Mullen, Kathleen J., Frank, Richard G., and Rosenthal, Meredith B (April 2007) "Can You Get What You Pay For? Pay-For-Performance and the Quality of Healthcare Providers" Job Market Paper: Harvard University  
[http://www.people.fas.harvard.edu/~kmullen/papers/P4P\\_latest.pdf](http://www.people.fas.harvard.edu/~kmullen/papers/P4P_latest.pdf)

Muralidharan, Karthik and Venkatesh Sundararaman (2006) "Teacher Incentives in Developing Countries: Experimental Evidence from India" Job Market Paper; Department of Economics, Harvard University  
[http://econ.ucsd.edu/seminars/0607seminars/Muralidharan\\_paper.pdf](http://econ.ucsd.edu/seminars/0607seminars/Muralidharan_paper.pdf)

Paarsch, Harry J and Shearer, Bruce (2000) "Piece Rates, Fixed Rates and Incentives Effects: Statistical Evidence from Payroll Records" International Economic Review Vol. 41 No. 1 p.59-92

Pema, Elda (2005?) "Internal Labour Markets Revisited: Tournaments in Academia" Working Paper, Graduate School of Business and Public Policy, Naval Postgraduate School, Monterey, CA <http://www.msu.edu/~pemaelda/E%20Pema-Tournaments%20in%20Academia.pdf>

Petersen, Laura A, MD, MPH; LeChauncy D. Woodard, MD, MPH; Tracy Urech, BA; Christina Daw, MPH; and Supicha Sookanan, MPH (2006) “Does Pay-for-Performance Improve the Quality of Health Care?” Annals of Internal Medicine Vol. 145, No. 4 265-272 <http://www.annals.org/cgi/reprint/145/4/265>

Pink, George H., Adalsteinn D. Brown, Melanie L. Studer, Kristin L. Reiter, Peggy Leatt (2006) “Pay-for-Performance in Publicly Financed Healthcare: Some International Experience and Considerations for Canada” Healthcare Papers, 6(4) 2006: 8-26 <http://www.longwoods.com/product.php?printable=Y&productid=18260&page=2>

Prendergast, Canice (2001) “Selection and Oversight in the public sector, with the Los Angeles Police Department as an example” NBER Working Paper No. 8664 [http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=294102](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=294102)

Prendergast, Canice (1999) “The Provision of Incentives in Firms” Journal of Economic Literature Vol.37, No.1, p.7-63

Propper, Carol and Wilson, Deborah (2003) “The Use and Usefulness of Performance measures in the Public Sector” Oxford Review of Economic Policy Vol. 19 No. 2, 250-268 <http://oxrep.oxfordjournals.org/cgi/content/abstract/19/2/250>

Ramseyer, J Mark and Rasmusen, Eric B (1997) “Judicial Independence in a Civil Law Regime: The evidence from Japan” Journal of Law, Economics and Organization Vol 13, No. 2 <http://jleo.oxfordjournals.org/cgi/content/abstract/13/2/259>

Rosenthal Meredith B, Frank Richard G, Li Zhonghe and Epstein, Arnold M (2005) “Early Experience with Pay for Performance: from concept to practice” Journal of the American Medical Association, Vol. 294, No. 14, p 1788-1793 <http://jama.ama-assn.org/cgi/content/abstract/294/14/1788>

PriceWaterhouseCoopers (October 2006) “Connecting Public Sector Pay to Service Delivery” Government and Public Sector <http://www.pwc.com/extweb/pwcpublications.nsf/docid/16F22C47349E066280257218004FCC77>

Roski, Joachim Ph.D., M.P.H., Robert Jeddelloh M.D.b, Larry An M.D.c, Harry Lando Ph.D.d, Peter Hannan M.Stat.d, Carmen Hall M.S.e and Shu-Hong Zhu Ph.D (2003) “The impact of financial incentives and a patient registry on preventive care quality: increasing provider adherence to evidence-based smoking cessation practice guidelines” Preventive Medicine Volume 36, Issue 3, March 2003, Pages 291-299 [http://www.sciencedirect.com/science?\\_ob=ArticleURL&\\_udi=B6WPG-47XSXB3-3&\\_user=121739&\\_coverDate=03%2F31%2F2003&\\_rdoc=1&\\_fmt=&\\_orig=search&\\_sort=d&\\_view=c&\\_acct=C000010018&\\_version=1&\\_urlVersion=0&\\_userid=121739&md5=966f66f0ca82610ec1d8f50404c3cdc9](http://www.sciencedirect.com/science?_ob=ArticleURL&_udi=B6WPG-47XSXB3-3&_user=121739&_coverDate=03%2F31%2F2003&_rdoc=1&_fmt=&_orig=search&_sort=d&_view=c&_acct=C000010018&_version=1&_urlVersion=0&_userid=121739&md5=966f66f0ca82610ec1d8f50404c3cdc9)

Serneels, Pieter; Lindelöw, Magnus; Garcia-Montalvo, Jose and Barr, Abigail (2005) "For Public Service or Money: Understanding Geographical Imbalances in the Health Workforce" World Bank Policy Research Working Paper 3686, August 2005 <http://heapol.oxfordjournals.org/cgi/content/abstract/22/3/128>

Shearer, Bruce (2004) "Piece Rates, Fixed Wages and Incentives: Evidence from a Field Experiment" The Review Of Economic Studies Vol.71, p.513-534

Shi, Lan (Jan, 30, 2006) "Does Oversight Reduce Policing? Evidence from the Cincinnati" Department of Economics, University of Washington [http://faculty.washington.edu/lanshi/Research/policing\\_jan\\_06\\_lan\\_shi.pdf](http://faculty.washington.edu/lanshi/Research/policing_jan_06_lan_shi.pdf)

Stoddard, Christiana and Kuhn, Peter (2006) "Incentives and Effort in the Public Sector: Have U.S. Education Reforms Increased Teachers' Work Hours?" NBER Working Paper No. 11970, Issued in January 2006 <http://www.nber.org/papers/W11970>

Sutton, Matt, and Gary McLean (2006) "Determinants of primary medical care quality measured under the new UK contract: cross sectional study" British Medical Journal 332: 389-390 <http://www.bmj.com/cgi/content/abstract/332/7538/389>

Vollaard, Ben (2003) "Performance contracts for police forces," CPB Documents 31, CPB Netherlands Bureau for Economic Policy Analysis. <http://ideas.repec.org/p/cpb/docmnt/31.html>

Vyrastekova, Jana; Onderstal, Sander and Koning, Pierre (2006) "Team Incentives in Public Organisations: An experimental study" CPB Discussion Paper, Netherlands <http://ideas.repec.org/p/cpb/discus/60.html>