Abstract

Much of the academic and policy literature on performance related pay focuses on its role as an incentive system. Its role as means for *renegotiating* performance norms has been largely neglected. The introduction of performance related pay, based mostly on appraisals by line managers, in the British public services in the 1990s can be considered as a large-scale social experiment in the change from a seniority - to a performance-based payment system. When reviewing academic research and management inside information on the schemes, a recent government report concluded that they had failed to motivate staff and their operation had been divisive. Nevertheless, other information suggests that productivity rose. This article seeks to resolve the paradox by showing that performance pay was the instrument of a major renegotiation of productivity norms, and that this rather than motivation was the key story. It concludes that when analysing incentive systems, more attention needs to be given to contract theory, and in particular to the articulation of different levels of principal-agent relationships within organisations. The key to the rise in productivity in the British public services lay in how the appraisal activities by line managers were articulated with incentives and goal setting for the different levels of organisational performance in order to secure the passage to different performance norms.

This paper was produced as part of the Centre's Labour Markets Programme

Acknowledgements

Anonymous referees and seminar participants at CEP, SASE and Grenoble.

David Marsden is Professor of Industrial Relations, London School of Economics and Political Science (appointed Lecturer at LSE in 1980), and Member of the Centre for Economic Performance.

Published by
Centre for Economic Performance
London School of Economics and Political Science
Houghton Street
London WC2A 2AE

ISBN 0 7530 1639 7

Individual copy price: £5

Renegotiating Performance: the Role of Performance Pay in Renegotiating the Effort Bargain

David Marsden

July 2003

The Centre for Economic Performance is financed by the Economic and Social Research Council

1. Introduction and Outline of the Argument

In the public policy debate it has been common to associate the introduction of performance related pay (PRP) with the aim of improving incentives and motivation among public employees (Brown and Heywood, 2002). This has been a key element in government and top management thinking in the British public services, echoed in two recent government reports (Bichard, 1999; Makinson, 2000), and is strongly echoed in the work of the OECD's public management reform programme (Maguire, 1993; OECD, 2001; 2002). It is also a recurrent theme in much of the Personnel Economics and Human Resource Management literature (e.g. Lazear, 1998; Milkovich and Wigdor, 1991; Mitchell et al., 1990; Armstrong and Murlis, 1994). From the late 1980s, the British public services embarked upon one of the most systematic and sustained policies of extending and developing performance related pay of any OECD country, mostly replacing annual seniority-related pay increments with performance-related increments based on goal setting and appraisals by line-managers, sometimes called 'appraisal-related pay'. The perseverance of top public management and of successive governments is hard to understand if employee motivation is the main story. The Makinson report, which drew on both academic research findings and inside management information, concluded that performance pay had not motivated public employees in Britain, and its operation had been divisive. Given that the policy has been sustained by three successive prime ministers of quite different political persuasion, two Conservative and one Labour, and similarly among top managers, it is hard to believe its continued use can be explained by political dogma. We need to look elsewhere for an explanation.

That alternative explanation lies in the role of performance pay, and of performance management more widely, in providing a framework for the *renegotiation* of performance standards among public employees. The idea is most simply expressed in terms of contract theory. A worker and a firm agree the terms of their exchange when the worker is hired. A key feature of the employment contract is that it is open-ended in terms of both its duration and its content. Workers agree to give the employer's agent - management, some flexibility to adapt that content to changing demands, but only *within certain limits* (Coase, 1937). From time to time, it becomes necessary to revise these limits. This becomes an occasion for renegotiation. This time, however, each party has made investments in the relationship and is vulnerable to pressure tactics from the other. Much of the contract literature has emphasised pay because of changes in the market valuation of employee output (Malcomson, 1997).

Less visible, but just as important for management, is the ability to revise job boundaries, and to redefine the kinds of performance it requires from employees. This need for renegotiation is recognised by labour law. For example, under English law it has to be by mutual agreement, whereas US law gives the employer more scope to act unilaterally, 'at will', although in practice, many US employers commonly act by mutual agreement (Malcomson, 1997).

In their advocacy of corporatism as a basis for orderly renegotiation, Teulings and Hartog (1998) focus mainly on pay. However, pay rules are codified and can be administered some distance from the place of work. In contrast, the very flexibility of job contents that makes the employment relationship so useful to employers means that a large part of any job remains uncodified. Job classifications introduce a degree of order, but in Williamson's (1975) terms, much of the job content remains 'idiosyncratic' and uncodified, accessible to higher management only through the eyes of first-line managers. To change performance norms, therefore, top management must engage in a much deeper process throughout their organisations, and place a major responsibility on line managers for their renegotiation with the groups of employees they supervise.

The argument I wish to develop in this paper is that performance pay, and more widely, performance management, played a key role in this process of renegotiating job-level performance in the British public services, and there is every reason to believe a similar process occurs elsewhere with performance pay. The incentive and goal setting features of performance pay played a key part, but motivation was only their secondary function. Their primary function was, through the appraisal and goal-setting process, to enable management to redefine performance norms in their organisation, and then to operate them effectively, and with the explicit or tacit agreement of as many employees as possible. This argument resolves a paradox that has run through the research on performance pay in the British public services: the various schemes appear to have reduced motivation, as the Makinson report observed, but to have benefited productivity, hence the perseverance of top management and successive governments. There would of course be no paradox if the employees had narrowly defined jobs and tight supervision, but this is not the case for much public service employment. Because so little has been written about renegotiation in the context of performance pay, it is useful to consider an example adapted from the author's fieldwork, to give a feel for how it can work (Box 1).

Box 1. An example of the use of performance pay to induce more flexible working

In many organisations, it has long been common to regulate the supply of hours from current employees by the system of paying overtime and weekend hours at premium rates. This gives the employer flexibility, but it also protects employees against unreasonable variation in their workloads. It is an example of one of the limits within which the right to manage operates. Overtime rules are clear and unambiguous, and can be easily enforced even in low-trust work environments. Despite their simplicity, it is increasingly common for them to conflict with modern patterns of team working, especially when different groups of employees have different pay and working time preferences. This was a problem in one of the hospitals in this study which wanted to introduce more flexible, patient-centred, care teams. The administrative problems of different pay rates for different hours, and the desire of some employees for long hours in order to get higher pay, made it difficult to operate such teams and to ensure the availability of the desired mix of skills.

One solution for management is to replace the overtime and unsocial hours payments with performance pay in which the willingness to work flexible hours is one of the criteria of good performance. Figure 1, adapted from the standard analysis of overtime working (e.g. Hamermesh and Rees, 1993), compares the two pay systems. The hourly rate of pay for normal hours and overtime hours is shown by the line AED, the slope increasing after 40 hours to reflect the enhanced overtime rate of pay. With performance pay, the hours shown represent average hours over a certain period, say one month, and the kinked wage schedule is replaced by two pay schedules: one including and the other excluding performance pay, respectively AD and AB.

Line managers can now back up their requests for extra hours with the offer of good appraisals for cooperative working, and sanction lack of cooperation with bad appraisals and no performance pay. The indifference curves show that the *median* employee is better off with higher average (flexible) hours and performance pay. So one could say, that the introduction of performance pay has been used to negotiate the new, more patient-oriented, working patterns that management wanted. In terms of the output that management values, productivity has risen. Depending on the savings from overtime payments and the efficiency gains, the employer might be better off. The median employee may accept the new norm, finding it financially worthwhile without finding it motivating. The greater the variation of employee preferences around the median, the more likely is discontent and loss of motivation.

How successfully performance pay and appraisal achieves the new working patterns depends heavily on how effectively it is operated by line managers. Although there is evidence that line manager appraisals can reflect actual performance (Boswell and Boudreau, 2000), it is not a foregone conclusion. In the famous case of one British car manufacturer in the 1970s, the introduction of 'Measured Day Work' became known on the shop floor as 'Leisure Day Work' because line managers lost control of performance, and productivity crashed. In the current example, the flexibility of the new system contains an element of vagueness: individual employees are not expected to provide the extra hours on every occasion, but to show goodwill when they are asked. In appraisals, line managers have to judge whether or not goodwill has been shown. Their judgement can be contested by their staff so that leniency gives them a quiet life. Hence the degree of support line managers receive from higher management will determine whether the performance pay and appraisal encourages and reflects actual changes in performance or whether it is just a fiction.

The analysis in this article advances in three stages. First, it reworks the data collected by Marsden and Richardson (1992; 1994) and Marsden and French (1998) in a series of attitude surveys across a range of public services on employee and line-manager judgements as to the effects of performance pay. These were also among the evidence considered by Makinson (2000). They show that only a rather small minority appreciated the incentive effects of their performance pay schemes, and that many found them divisive. However, they also showed that a substantial minority of line managers thought performance pay had raised productivity. Using individual employees' performance appraisal scores as measures of their performance, it is shown that the performance pay and appraisal systems were mildly effective as an incentive, reflecting no doubt the minority who responded to them favourably, but they were much more effective as a means of directing employee performance towards the goals management wanted. Employees who reported that the appraisal process had been well-conducted, on this measure, had performed better. The measure of appraisal quality was subjected to a number of methodological checks for its independence with regard to employees' appraisal scores. Thus it is established that the performance pay schemes could be *instrumental* in the renegotiation process even though their incentive effects were weak.

The second stage of the analysis sets employee goal setting and appraisal in the wider context of performance management for the whole organisation. This is essential for two

reasons. As social scientists, we need to know whether performance appraisals represent actual performance. Top management is in the same position, and it needs to ensure that the judgement exercised by line managers in appraisals tracks the kind of performance it wants the organisation to achieve. Both questions can be answered simultaneously by considering performance management as a hierarchy of principal-agent relationships. Line managers are under constant pressure from those they supervise to be generous with appraisals and performance pay. The tendency for scores to drift upwards is well-known and well-documented (Milkovich and Wigdor, 1991). Thus top management need to 'monitor the monitors', and the different levels of performance management have to be articulated if performance pay and appraisal are to be used successfully to raise performance. This section therefore explores the *procedures* top management uses to control such tendencies, and more important still, it examines the *behaviour* of performance indicators for different organisational levels and their articulation. For the latter, the analysis uses eight years' time series archival data for ten administrative units within the tax service on employee performance appraisals, and office targets and outcomes, plus overall performance data for the tax service as a whole to examine how the different levels articulate. This is done by considering two alternative readings of the data: the management *leniency hypothesis*, and the *productivity hypothesis*. According to the first, variations in performance appraisals between offices are the consequence of local management leniency, and of the failure of office-level targets to bind on performance. According to the second, the targets and different levels at which performance is measured are mutually supportive. This judgement is based on the behaviour of the targets and their relationship with outcomes across offices over the eight years between 1993 and 2000. It is shown that the *productivity hypothesis* gives the better account of the two.

The third stage of the argument explains how the *productivity hypothesis* is related to the evidence of demotivation and divisiveness. It is argued that appraisal and goal setting have two faces: one in which targets are agreed, and one in which they are accepted under duress. This fits with the renegotiation argument because many incumbent employees did not want to change. For example, some may have found the previous performance norms fitted better around their non-work and domestic commitments. The two hospitals in the study dealt with this problem by making transition to the new pay scheme voluntary for current employees, but compulsory for new hires and those who are promoted.

The overall strategy of the argument is to infer renegotiation from the data rather than rely upon interpreting management communications to staff about the purpose of

performance management. In fact, in the tax service, the scheme espoused what it called a 'contractual approach' of agreeing targets, appraising against them, and linking them to wider organisation objective of continuous improvement (Inland Revenue, 1994, 1995). However, it is clear from the survey replies that one cannot assume the scheme worked this way in practice. This can only be judged from the data reviewed here.

2. The Evidence on Motivation and Divisiveness of Performance Pay

Some details of the disenchantment observed by Makinson (2000) are captured in Table 1, the employee replies to attitude surveys by Marsden and Richardson (1992, 1994) and Marsden and French (1998). These relate to six areas of public service work: the Inland Revenue in 1991 and 1996 (taxation); the Employment Service (job placement and benefit payments); two National Health Service trust hospitals; and head teachers in primary and secondary schools (elementary and high schools). These were chosen to represent a cross-section of public organisations using performance pay at the time. Methodological details are given in the appendix, and the 1991 and 1998 publications can be obtained online from www.cep.lse.ac.uk). In brief, postal questionnaires asked about employee and line manager personal experiences with the operation of their performance pay and appraisal scheme in their service, their views as to whether it provided them with an incentive to perform in specific ways, whether their jobs gave them scope to do so, their judgements as to how management operated their scheme, and some biographical data. Many of the motivational questions were modelled on expectancy theory. In some cases, management gave their support and it was possible to survey a sample of all employees covered by the scheme in their organisation. In others, management refused access for the survey work, although they did provide other information, and we had to rely on the unions to provide us with a sample frame based on their membership lists. They all had high membership rates.[1] In the organisations where management cooperated, we were able to check whether union membership affected the replies, and it appeared to have no great influence.

The findings are broadly consistent with the results of other attitudinal surveys that applied the same methodology as that used by Marsden and Richardson (1992), notably,

Thompson (1993), Kessler and Purcell (1993), Heery (1998), IRS (1999), and in the private sector, Carroll (1993). Despite broad support for the principle of linking pay to performance, only a small percentage of employees thought their existing performance pay schemes provided them with an incentive to work beyond job requirements or to take more initiative. Of even more concern to top public management, was the evidence that the performance pay schemes in place were seen by staff to be divisive and to undermine cooperation among staff, and a worrying percentage of line managers reported that the schemes had made staff less willing to cooperate with management.

This cannot be explained by a naïve design of the schemes, summarised in the methods appendix (Table A1). With the possible exception of the scheme in force in the tax service in 1991, which was one of the first in operation, all of the schemes obeyed the canons of good HR practice (as set out for example by ACAS, 1990 and Armstrong and Murlis, 1994) and had been developed with substantial inputs from private sector expertise. They were seriously thought-out schemes. With the knowledge that ratings often drift upwards, and that their application can be discriminatory, all the schemes incorporated substantial review mechanisms, and shared information with the relevant unions on the distribution of ratings across different categories of staff and workplaces. Reflecting the degree of task complexity in many public service jobs, all the individual schemes involved performance appraisals by line-managers based on a mixture of judgement and recorded data. Written records were kept of appraisals. Nor was the financial incentive negligible. Up to the top of the pay scale for a person's grade, they replaced annual salary increments, and were consolidated into basic pay, and several years' of good performance could lead to substantially faster pay progression. For those who would previously have 'topped out' at the maximum for their grade, PRP brought the opportunity of non-consolidated annual bonuses in some organisations, and of further progression in others.

---

[1]. It was 90% in the Inland Revenue, 60% in the Employment Service middle management grades studied, and around 90% among head teachers. Public hospitals are also highly unionised.

2. The Instrumentality of Performance Management in Renegotiating Performance Norms

All the attitudinal surveys cited by Makinson investigated schemes that covered practically all the employees in certain grades in their organisations so the scope for comparing appraisals and their outcomes for covered and non-covered employees was severely limited. Therefore, the strategy adopted here is to explore whether key features of the performance pay and appraisal schemes were functioning in a way that would promote changed performance. This section looks at two main routes through which such effects could take place: through employee perceptions of incentive and divisiveness, and through direct effects of effective appraisals on performance. These channels are shown in Figure 2.

2.1 Impact of appraisal on perceived incentive and divisiveness

The choice of variables in this analysis is based on the three dominant theoretical approaches to the study of performance pay: agency, expectancy, and goal setting theory. To varying degrees, they all stress the importance of employee choice over effort levels, and hence of motivation in determining their willingness to perform. This choice is influenced by the rewards that flow from good performance, and by the manner and effectiveness with which performance goals are set.

Agency theory stresses the role of performance and output incentives as a means of encouraging employees to work hard (and not to 'shirk') when management find it costly to monitor their effort closely. Management can act in two ways: it can tie pay to output in some way so as to induce employees to choose a higher level of effort (Lazear, 1995, Ch. 2). It can also invest in better systems of work design and performance measurement to improve the correlation between effort and measures of performance (Milgrom and Roberts, 1992: 226). Agency theory also warns against the dysfunctions of inappropriate incentives, for example, that individual incentives may encourage employees to boost their own performance at the expense of cooperation with their colleagues (Drago and Garvey, 1998).[2] This trade-off suggests we need to consider both incentive and divisiveness effects of individual PRP schemes.

[2]. Strictly speaking, their evidence relates to promotion.

Expectancy theory, associated for example with Vroom (1964), Porter and Lawler (1968), Lawler (1971, Ch. 6), and Furnham (1997), like agency theory, treats employees as having a degree of choice and places a strong emphasis on the motivational effects of incentives, and the problems posed by poorly defined targets. Simplifying somewhat, it identifies a potentially virtuous circle. Employees will respond to the incentive or reward on offer if they value it (its *valence*), if they believe good performance will be instrumental in bringing the desired reward (*instrumentality*), and if they expect their efforts will achieve the desired performance (*expectancy*). The circle of Valence-Instrumentality-Expectancy can be broken at a number of points. Employees may feel they lack scope to increase their effort, or that their effort will make little difference to their performance, such as might arise if they are given inappropriate work targets by management. This undermines expectancy. They may believe that management lacks the competence or the good faith to evaluate and reward their performance fairly, which undermines instrumentality, and may cause employees to view the schemes as unfair and divisive.

Goal setting theory, although placing less emphasis on rewards, stresses the motivating power of defining appropriate work goals and engaging employee commitment to them (Locke and Latham, 1990; Latham and Lee, 1986; Brown and Latham, 2000). Of special relevance in the current context, is the emphasis on dialogue between line-managers and employees to exchange information about realistic goals, and on agreeing goals so that employees adopt them as their own. Thus, although the three approaches differ in emphasis, they point to the same key variables for the analysis of performance pay systems: reward and motivation on the one hand, and goal definition and evaluation on the other. Being concerned with negotiation at the level of individual employees in this case, contract theory comes closest to goal setting theory, placing the primary emphasis on the goal setting and appraisal functions of PRP as a means of communicating the new performance norms to employees, and securing their acceptance.

A simple, informal model of the key relationships, and their signs, may be summarised as follows:

Perceived incentive (+) = f(effective appraisal (+), clear target setting (+), scope for employees to boost performance (+), financial incentive (+), (control variables))

Perceived divisiveness (+) = f(effective appraisal (-), clear targets (-), scope for employees to boost performance (-), (control variables))

The questions relating to 'perceived incentive' in Table 1 were chosen to represent aspects of these incentive theories. The first two questions capture the perceived disutility or cost to the employee of effort required to gain the reward: willingness to work beyond job requirements, and to take more initiative in order to get PRP. The one entails more effort; the other, more risk of failure. The third question captures the element of perceived reward for good work as opposed to 'shirking'. This measure of perceived incentive is close to that of valence of rewards in expectancy theory: are the rewards sufficiently valued to warrant the extra effort?

The downside, 'perceived divisiveness', is explored by three questions chosen to capture the disutility of poorer work relations, and also that of diminished cooperation that may jeopardise the achievement of work targets. If staff are less willing to help their colleagues, the risk of failure to achieve targets is individualised, and the safety net of helping hands is removed. Likewise, should the pay system cause jealousies. Reduced willingness to cooperate with management captures the vertical as opposed to the horizontal aspects of cooperation among work colleagues. The indices of perceived incentive and divisiveness were computed simultaneously using factor analysis based on these questions.

The key independent variable, the quality of the appraisal process ('effective appraisal'), plays a key part in both agency and expectancy theory. This is built up from three questions: does an employee know what she needs to do to get a good appraisal; is she able to do it; and does she understand her last appraisal rating. These questions were validated against a more concrete, descriptive, set of questions about the appraisal process used in one of the hospitals in the study, and which were very unlikely to be coloured by whether or not the employee got a good rating (see methods appendix). For clarity of target setting in PRP just one question could be matched across the organisations: did PRP lead managers to set targets more clearly. This was supplemented by a question to line managers in the same office on the scope employees have to raise their performance. The results in Table 2 show that having an effective appraisal increased employees' perceived incentive and reduced perceived divisiveness. The same is true of improved target setting. As anticipated, lack of scope to improve performance increases perceived divisiveness, although the effect on incentive is barely significant.

Financial incentive could not be measured directly because good appraisals trigger performance pay, and later appraisal scores are used as a measure of employee performance. However, its presence can be measured indirectly in two ways. On the one hand, those on the top of the pay scale for their grade get one-off bonuses instead of an increase in their basic salary. One would expect such employees to feel less incentive than the others. On the other,

those who were both of long service and on their grade maximum would remember the former pay system of about 3-4 years before, with its ceilings on pay whereas those more recently recruited would not. Thus, an additional measure of the presence of financial incentive from PRP can be found by interacting employees' being on their grade maximum with their length of service. The results show that being on the top of one's pay scale diminishes perceived incentive, whereas the positive interaction with length of service indicates that longer serving employees are conscious of the improvement compared with the previous age-incremental pay system.

Affective commitment, as measured by Meyer and Allen (1997), provides an indirect proxy for 'shirking' behaviour, which is otherwise difficult to explore in a questionnaire survey to the individuals concerned. Individual shirking is bad for the employer and usually bad also for one's work colleagues as it usually disrupts their work and adds to their workload. In contrast, commitment, and especially affective commitment, implies a degree of emotional identification with one's workplace, and one's work colleagues. It was included because it was thought that commitment might be strong among public employees, many of whom have quite long service. In the regression, commitment enhanced the perceived incentive of PRP and reduced its perceived divisiveness.

The analysis in Table 2 also includes a number of organisational and demographic controls. Organisation dummies are used to control for fixed effects arising from differences in the schemes in operation in each organisation, the most notable being variations in the share of employees getting performance pay owing to differences in the design of their schemes. Occupational controls were used, comparing each occupational group to managers, the one occupation that could be clearly identified across all the organisations. 'Occupation' captures many possible effects, but one notable one is that the clerical and service occupations generally have less control over the detail of their work than do managers, and professionals, and hence less scope to respond to performance pay incentives. On the other hand, the simpler nature of their tasks may make their performance easier to evaluate. Both effects seem to be present in Table 2: the lower down the hierarchy, the stronger the perceived incentive of PRP, but also the stronger the perceived divisiveness. For divisiveness, professionals are the exception, possibly because they have long been accustomed to exercise considerable discretion in their work and so resent the extra management control that comes with performance management. Length of service and gender are introduced as additional demographic controls. Long service employees may be generally more resistant to change having invested more in the former pay systems, and this

appears to be the case in Table 2, but the coefficients are small. One might expect men to be more responsive to individual performance rewards than women, but in this sample, the effects of gender appear to be weak or non-significant.

Finally, the coefficient for the group PRP scheme hospital deserves comment. It shows that the group scheme was considerably less divisive than the individual PRP schemes used in the other organisations.[3] This supports the argument of Drago and Garvey (1998) that strong individual incentives may diminish helping behaviour among colleagues if this gets in the way of individual targets.

Thus, a first conclusion is that the performance pay and appraisal schemes were actively influencing employee motivation, and that they did so in the manner the main theories predict.

2.2 The impact of perceived incentive and divisiveness on appraised performance

To be instrumental in renegotiating performance, the performance pay and appraisal schemes need to reach beyond motivation to influence the performance of individual staff. Because of the need for line manager judgement, the researcher, like top management, is dependent on appraisal scores for a measure of individual performance. This section seeks to show that appraised performance was better when motivation was better and when the process of appraisal was conducted well. The section after, on organisational performance, completes the argument by linking appraised to actual performance.

Employees reported their latest appraisal score in the attitude surveys. It is likely that they remembered these accurately because they affected their pay directly. To judge by the close match between the distributions of appraisal scores in the sample surveys and in the archival data, respondents also reported them accurately, and there was no significant response bias by appraisal scores. Because performance was graded differently across the schemes, outcomes were classified into 'superior' and 'acceptable', the latter including both satisfactory and the very small number of unsatisfactory ratings. The performance variable is therefore a binary one, and a logit regression was used (Table 3). The analyses regress employee performance first on perceived incentive and divisiveness plus the same batch of controls used earlier, then on appraisal effectiveness, and finally on all three combined in order to measure the interactions. Mostly, the latter are not significant.

The logistic regression in Table 3 examines variants of the following relationship:

Appraised performance level (+) = f(perceived incentive (+), perceived divisiveness(-), appraisal quality (+), interactions, control variables)

Table 3 shows quite clearly that incentive and divisiveness do affect individual performance. The effect of the first is positive and of the second is negative, and both are strongly significant (equations 1 and 3). As an approximate guide, given the crude nature of the Likert scales, one can say that a one standard deviation increase in perceived incentive would raise the probability of 'superior performance' by about 0.6 and a similar increase in perceived divisiveness would reduce it by about 0.4.[4] The strong coefficient for effective appraisal (equations 2 and 3) deserves comment: it implies that a standard deviation increase in effectiveness of appraisal would lead roughly to a 0.7 increase in the probability of superior performance (equation 2). The robustness of this coefficient in both equations 2 and 3 indicates that there is also a *direct* effect of appraisal on performance. In other words, effective appraisal can raise performance directly, independently of its effects mediated through motivation, as shown in Figure 2.

It is possible that the performance appraisal scores of individual employees might colour their reporting of the quality of their appraisal process and the measures of perceived incentive and divisivenes. Although a recent study found that appraisal scores had little influence on perceptions of the appraisal process, this may depend on how it is operated in different organisations (Boswell and Boudreau, 2000). This can be checked further in two ways. The first test, discussed in the previous section, was to use the richer descriptive data collected on the appraisal process in the two hospitals and show they also correlated well with the measures of effective appraisal. Second, a two-stage least squares regression was run. This sought to predict, respectively, perceived incentive and perceived divisiveness from the effective appraisal variable shown in Table 2, and then, using the predicted values of incentive and divisiveness, to predict performance appraisal scores. These had the correct

---

[3]. The individual trust hospital was not included in the regression because the pay system did not operate scale maxima and so absence of that variable excluded data from that hospital.

[4] The standard deviation of both perceived incentive and perceived divisiveness is 1.0. The logistic regressions estimates the change in the log of the odds of achieving superior performance associated with a unit change in a given independent variable, that is log(p/(1-p)), where p is the probability of the event, i.e. achieving superior performance. With a

signs and were highly significant, and so confirm that even though there may be some perceptual bias caused by the employee's appraisal score, it was not such as to undermine the model proposed here (details in the methods appendix).

Thus, the attitudinal survey data so far confirm that employees who experienced well-operated objective setting and appraisal with their managers are likely to find their scheme motivating, whereas those who experienced the opposite find their PRP schemes demotivating and damaging to work relations, and that this had the predicted effect on individual appraised performance. Equally important was the strong direct influence of well-conducted appraisal on appraised performance.

3. Individual and Organisational Performance: Management Leniency or Productivity?

So far the discussion has focused on appraised performance. Although this presents a picture in which it is plausible that appraised performance tracks actual performance closely, this cannot be taken for granted. The next step is to see how employee appraisals tie in with other measures of organisational performance.

It would be easy to test for a link between individual and organisational performance and if one could simply aggregate some physical or financial measure of output from the individual to the organisational level[5]. However, this is not possible when organisations rely heavily on judgemental performance appraisals to measure employee performance. Public service organisations are not unique in this respect. Any organisation is likely to do this when its employees engage in multiple tasks whose output can be measured with varying degrees of difficulty (Landy and Farr, 1983; Holmstrom and Milgrom, 1991). To show there is a link between judgemental appraisals and organisational performance measures, one needs to assess whether the schemes in the British public services had the *procedures* to enable

---

standard deviation of 1 for both motivation variables, p = e exp(b)/(1+ e exp(b)) where b is the regression coefficient.

[5]. The issue is not whether line managers are able to appraise actual performance accurately: there is other evidence that they can. One recent study, which unusually combined line-manager appraisals and employee productivity measures, and sought to evaluate the extent of rater bias, found that despite its presence, judgemental appraisals correlated strongly with measures of actual job performance (Elvira and Town, 2001). Rather, it concerns how top management can ensure that they do.

management to evaluate productive performance, and that *behaviourally*, this was what they actually did. Did poorly performing offices improve during the operation of performance management, and could this be related to the operation of their targets? To assist the presentation of the evidence, this section contrasts two alternative hypotheses: that appraised performance scores and performance management are mostly governed by local management indulgency, the *'leniency hypothesis'*; and that they represent processes enhancing individual employee productivity, the *'productivity hypothesis'*. A summary of the key tests is given in Figure 3.

3.1 Procedural aspects

Taking the procedural question first, did the organisations have performance appraisal procedures that were able to measure underlying productivity, as the *productivity hypothesis* would require? Job classifications are a key set of procedures to support judgemental appraisal systems. As Betters (1931) showed, these enable management to define and benchmark performance standards for employees in similarly classified jobs. In principle, these establish contours of job comparability within an organisation, and thus enable managers to compare performance levels between employees in the same jobs, and outcomes for the same employee over time. This is facilitated by the degree to which many white-collar jobs involve the application of routines that govern day-to-day working. For example, tax officials follow certain routines in checking tax returns, and health professionals apply certain standard diagnostic routines. In the absence of any absolute measure of performance, the benchmarking supported by job classifications enables management to compare performance of employees in similar jobs both over time and between employees. Job classifications are widely used within the public services, so this part of the infrastructure for performance management is in place.

The procedures for the administration of performance pay can also help to keep line managers focused on performance that is useful to the organisation, and to ensure that the judgemental element of their ratings is grounded in observed performance, and can be checked by others. In this respect, the various schemes had been set up with a substantial input from outside and from the private sector, and they conformed by and large to the canons of good HR practice of the time, as set out for example by ACAS (1990) and Armstrong and Murlis (1994). Appraisals were written, and line managers encouraged to give specific objectives, and the schemes had provisions enabling higher management to

monitor sources of potential bias. All appraisals were vetted by the next higher level of management, and information on the distribution of appraisal scores was made available to the unions. Job classification again facilitates standardisation of appraisals across different parts of the same organisation. In addition, a survey of tax service staff confirmed that the great majority had regular contact with their line managers, so the latter should be well-informed about their work, and that line-managers spent several hours on each appraisal (Inland Revenue Department Whitley Council, 1991). Moreover, because most line-managers have been promoted 'from the ranks', they will know about the detail of the work to be undertaken. There is therefore good reason to believe that line-managers are in a position to make an informed judgement about the performance of their staff in these organisations, and that there is a degree of internal consistency.

These procedures were intended to protect the organisations against drift in appraisal scores and in the pay bill that would be associated with line-manager leniency, and so their presence favours the productivity hypothesis. However, having good procedures is a necessary and not a sufficient condition for management to gauge employee productivity effectively. To deal with this question, we need to study the behaviour of the appraisal scores and other organisational performance indicators over time.

3.2 Behavioural aspects

Analysis of the behavioural aspects is based on indicators of organisational performance assembled from public accounts. For this study, they are most complete over time and across internal administrative units for the tax service, although the more fragmentary data for the Employment Service and hospitals paint a consistent picture.[6] For the tax service, it has been possible to piece together an annual time series for 1993-2000 relating to the distribution of employee appraisal scores for each of its ten regional offices, each office's annual performance targets and outcomes, and measures of performance for the organisation as a whole.[7]

The measures of performance for these administrative units stem from the new performance management system introduced in the tax service in 1993. Then, it greatly

[6]. For the pooled analysis across all the organisations, it is possible to identify 36 administrative units, but not consistently over time. The results, not shown here, are consistent with those for the Inland Revenue.

extended its range of organisational performance indicators covering two types of measure in particular: quantity and quality. The first related to the percentage of different kinds of tax work cleared within so many days of its reception ('clearance'), and the second, to measures of quality of service, including the percentage of work that is 'right first time', and time taken to respond to tax payer queries, answer the telephone and so on ('quality') (see methods appendix for details). Each office is given its target at the start of the annual cycle, and its performance against that is measured at the end. The tax service's Internal Auditor verifies the published outcomes in its annual report and accounts which are laid before Parliament. For this analysis, the annual mean value was computed across the 'clearance' and the 'quality' targets for each office, and likewise for their attainment. The clearance and quality 'gaps' are measured by subtracting target from achieved performance so that a shortfall has a negative sign, and an overshoot, a positive one. Overall organisational performance is measured by real tax revenues per employee: a measure of organisational workload. When economic growth picks up, not only do tax revenues rise from current tax payers, but employment rises, thus increasing the number of taxpayers and of tax transactions. Rising business activity will also increase the number and complexity of tax transactions. The measure used here is consistent with another one published by the tax service: the audited cost of collection as a percentage of tax yield. This fell every year from 1993 to 2000, from 2.14% to 1.11% (Board of Inland Revenue, 2001).

These archival data are used to analyse office performance with three main questions in mind. Did poorly performing offices improve over time as the *productivity hypothesis* would require? Were targets used in such a way as to promote this process? Did the proportion of good employee appraisals correspond to lenient management handing them out to buy peace, or did they appear to be used to mobilise greater effort? Finally, did the increased demand for taxation services brought about by economic growth translate into increased pressure for recruitment and increased delays, or did it translate into higher productivity?

According to the *productivity hypothesis*, top management will use performance indicators to raise the performance of the poorest performing offices, and spread good practice from the best ones. This is indeed what the government's Audit Commission (1999) recommends, and has been practiced by the Inland Revenue for many years (NAO, 1989). In contrast, the *leniency hypothesis* would predict that performance problems in poorly managed

---

[7]. These data were mostly not published, but were shared between management the unions as

offices would be left to fester, and there would be persistence both at the top and the bottom ends of the performance league table of Inland Revenue offices. The best of the easily available measures of office performance are the targets set to each office by top management. If the *productivity hypothesis* is true, then these will be informed by the best information available to top management at the time. If that hypothesis is true, then the best-run offices will have the most demanding targets in any year. Because good practice is shared, they will lose their position at the top, but they will do so gradually. It takes time to effect the organisational changes needed to help other offices to improve. There will therefore be a gradual regression towards the mean. If the *leniency hypothesis* is true, then there will either be persistence at the top and at the bottom of the league table, as poorly performing offices are left to their own devices, or there will be unsystematic fluctuations as the targets reflect the vagaries of bargaining relationships between office and top management. The first two columns of Table 4 show there is indeed gradual improvement by the weakest performing offices, and a gradual loss of lead by the best performing ones over one or two years, as correlations with the initial year decline in each successive year.

A second indicator of whether office targets are being used actively by top management relates to how they are revised each year. As with the levels of performance indicators, year-to-year adjustment of targets in the light of our achieved outcomes shows that top management are using targets to steer office performance. If there is an overshoot, then top management will revise the target upwards, and if there are good reasons for a shortfall on targets, it can revise them downwards, the aim always being, in line with goal-setting theory, to keep targets achievable but stretching. If top management knows there is a local problem, it makes more sense to allow time for it to be addressed, and maintain the integrity of the targets rather than adjust them fully and immediately. Thus adjustment in this sense is consistent with short-term persistence of over- or under-shooting, but not extending into the medium or longer term. If the *leniency hypothesis* were true, then one would expect no such active management of targets as offices would be allowed to undershoot or overshoot persistently. The middle columns of Table 4 show that the 'clearance' and 'quality' gaps of individual offices are gradually eliminated.

The percentage of staff in an office achieving superior performance can be analysed in a similar fashion. The *leniency hypothesis* predicts that poorly managed offices will be persistently over-generous with appraisals as they try to buy peace as they struggle with poor

---

part of the joint monitoring of the performance pay scheme.

organisation and bad employee relations. In similar vein to the preceding indicators, it might also be consistent with lots of year-to-year variation as managers spend their time 'fire-fighting'. In contrast, the *productivity hypothesis* predicts again a limited degree of year-to-year persistence because the problems that local management seek to address in pushing their staff to superior performance on that scale are likely to be quite serious. There were two mechanisms by which management could use their scheme to this end. They could indicate to their staff that the additional performance they were seeking would be recognised when it came to their appraisals, and they could use the special provisions for posts that have above average demands, called 'extra loaded' posts, and which also count towards achieving superior performance. The last column of Table 4 confirms the pattern predicted by the *productivity hypothesis*.

The relationships among the performance indicators paint a similar picture in Table 5 which shows correlations for selected indicators pooled over time and between offices. Top management's active use of targets is confirmed by the correlation between changes in clearance and quality targets and the previous year's achievements. They adjust the targets for the coming year in the light of the additional information from outcomes in the year just ended. At first reading, the negative correlation between the percentage of employees achieving superior performance in an office and the degree of under- or over-shoot on targets seems to point towards the *leniency hypothesis*: management are over-generous with good appraisals and so fail to achieve their office targets. However, if this were true, then there would be a high degree of persistence in bad and good office performance, contrary to what was shown in Table 4. The alternative explanation, from the *productivity hypothesis*, is that as management know they are under pressure to achieve their targets, they use performance management to mobilise the extra effort and quality needed to reach them. The anticipated shortfall in this case is an indicator of the pressure on the office.

Finally, there is the background of rising workload and productivity in the Inland Revenue throughout the period of performance management (Table 5 and Chart 1). As argued earlier, rising tax revenues associated with economic growth are a proxy for increasing workload brought by economic growth. Under the *leniency hypothesis*, management would be faced with pressures from staff and unions to increase recruitment, and there would be increased delays in dealing citizens' tax affairs, and in the collection of tax revenues as tax staff held their level of effort constant. The data in Chart 1 show this did not happen. Hiring was held down and employment fell from 1993-1998. Clearance targets were adjusted downwards only slightly and quality targets were actually raised slightly.

Moreover, the shortfall on targets was kept under tight control. This was reflected in the correlation between the percentage of employees with superior performance and overall productivity. Nor was this achieved by accident. With the knowledge from the macroeconomic forecasts used to predict government tax revenues, top management could easily anticipate that workload would increase. Given its control over targets and their implementation, it seems quite clear that it was able to use them to translate the increased workload into increased efficiency, a point stressed in successive issues of its Report and Accounts. This argues clearly in favour of the *productivity hypothesis* and against the *leniency hypothesis*.

How did management succeed in persuading employees to provide the extra performance? The tax service, like much of the public service is highly unionised, and due process applies to all dismissals, so management cannot just bludgeon employees into compliance with threats of dismissal. The answer lies in the decline in basic salaries in the tax service, excluding performance pay, compared with average white-collar pay in the economy ('relative basic salary' in Chart 1). This shows the penalty facing employees who refused to aim for satisfactory performance under the new system - echoing the options facing employees shown in Figure 1.

The performance information on the Employment Service and the two NHS trust hospitals is far less complete. Nevertheless, the fragments available suggest a similar picture to that of the tax service. In the ES, job placements per employee rose sharply between 1993 and 1998, and were rising steeply in the early years of its PRP scheme (Employment Service, 1997). Using the indicators from the NHS Performance Guide (NHS Executive, 1997), it is possible to show that across a wide range of performance indicators, performance was improving greatly at the trust using individual performance pay, and to a lesser extent, as the one using the trust-wide bonus. This is consistent with the reports to the sample survey by a large minority of line-managers that performance pay had caused staff in their service to work harder.

4. Productivity and De-Motivation: the 'Two Faces of Appraisal'

If the *productivity thesis* is correct, and employees did shift to new performance patterns, then it is natural to ask why so many of them replied that their schemes lacked incentive and were

divisive in their operation. One clue lies in the direct impact of appraisal quality on employee performance observed in Table 3. One might go further to suggest that there are 'two faces' of appraisal. The appraisal interviews can give incentives by clarifying work goals and giving recognition when they are achieved, but they can also be a vehicle for management to pressurise employees into giving higher levels of performance for fear of losing pay or even losing their jobs. The sample data for the tax service provide some evidence on this. Respondents were asked whether staff felt pressured to accept management's choice of objectives, as opposed to agreeing them, the latter being the express philosophy of the service's performance management scheme (Inland Revenue, 1994). They were asked whether they thought everyone was in effect given the same targets – despite the philosophy that targets should be adapted to what individual employees can contribute. They were also asked about the negotiation of objectives: whether those who were awarded superior appraisals did so because they were cleverer at negotiating their objectives; and whether, when agreeing their objectives, they were more concerned to avoid the risk of a bad appraisal than to aim for a superior performance rating. They were also asked about how they thought management operated the scheme, fairly or otherwise, captured by whether or not they thought management applied a quota on good appraisals, and whether they used the scheme to reward their favourites. Table 6 shows how these assessments are reflected in judgements of effective appraisal, perceived incentive, divisiveness, and achieving superior performance.

One group of employees that was particularly sensitive to feelings of duress was part-timers, who are particularly numerous in the public sector. Given that many of them take on such work in order to balance work and domestic responsibilities, this group is especially likely to be reluctant to agree to a new trade-off between effort and reward and hence to renegotiate reluctantly. One notable feature of these results is that part-timers were twice as likely as full-timers to report pressure to agree targets, and they were also more likely to express cynical views about the other questions on appraisal in Table 6. In other words, for many of these employees, what they opposed was the renegotiation of the old performance norms that management sought through the appraisal process.

The picture to emerge is that those who feel they accept work targets under duress and that they are appraised against targets they did not agree are more likely to find the scheme divisive, are less likely to report favourably on the appraisal process, and on the whole, are less likely to achieve superior performance.

5. Appraisal and the Re-Negotiation of Performance

The final question is what causes these feelings of performing under duress? Is it just that some line managers are bad at appraisal and goal setting, and so do it in a threatening way, in which case, better design and more training might be the answer. Or is it related to the degree of pressure from the employer to raise performance, as part of a renegotiation of performance levels, especially when faced with diverse employee preferences? To answer this question about duress more fully, it is helpful to consider the respective positions of individual and collective bargaining (Table 7).

One can examine the intensity of re-negotiation at the individual and the collective levels. For the first, one can take the degree to which application of the new scheme is compulsory for all employees, so that all employees have to agree work objectives and accept monitoring of their achievement. Thus, in the Inland Revenue and the Employment Service, the schemes were universal and compulsory. In contrast, in the two hospitals, incumbent employees were offered a choice between higher basic pay under the new local performance pay scheme, and remaining on the old nationally negotiated time-based pay scales without performance pay. By doing this, the hospital management avoided conflict with some groups of employees which had built up considerable premium payments under the old pay system, for example, for weekend working. School head teachers came in an intermediate position because the implementation of performance pay for them depended on the initiative of their school's governors whom they could often influence. Finally, the scheme in force at the Inland Revenue in 1991 was very much a hybrid between the old seniority-incremental system and the new performance management system. In the words of the union negotiators, it was 'bolted on' to the old pay and appraisal system. Thus performance pay meant accelerated movement up the old incremental scale. The old performance appraisal system was based primarily on performance against a uniform set of criteria, with little reference to targets. The scheme had carrots but no sticks.

Collective bargaining has played a somewhat smaller role because it cannot do much more than set up a framework and establish incentives. The levering up of performance levels and the detailed reorientation of performance has to be done at the individual level between line managers and their staff. Nevertheless, the two collective agreements that ushered in performance pay at the Inland Revenue were conflictual. The 1988 agreement was obtained with a management threat that if performance pay were not included, there

would be no national agreement, and the 1993 agreement was preceded by a bitter strike despite early working parties on pay reform. The hospitals had the least conflictual introduction of performance pay as it came with new provisions for local bargaining.

Thus, *prima facie*, it would seem that the pressure from management as expressed through the extent and intensity of individual negotiation partially accounts for the different levels of perceived divisiveness in the various organisations in this study (Table 7).

6. Conclusion

Before drawing conclusions, some possible objections need to be set aside. The first is that performance pay eliminated widespread 'shirking' among public servants, and so naturally productivity would rise along with employee resentment. This is not consistent with the levels of organisational commitment found, whereby the great majority of respondents (67%) felt a strong sense of commitment to their place of work.[8] There may have been a small minority of 'shirkers' but it does not seem large enough to explain the widespread disenchantment noted by Makinson (2000).

The second concerns Lazear's (1998) finding that improving incentives attracted more productive employees, so that it would be possible that incumbent staff felt alienated while productivity rose as a result of the new recruits. This is ruled out by the low levels of recruitment in the public services during the 1990s, and notably in the Inland Revenue where employment was falling.

The third concerns the technical and other organisational changes that were taking place at the same time. Might performance management have acted simply as a 'lightning conductor' for the resulting discontent. This possibility cannot be entirely excluded because such changes were taking place. However, performance management was not just accidentally 'caught in the crossfire'. To work the new organisational patterns and to integrate the new technology into working practices, management had to renegotiate performance norms. In the 'care team' overtime example given at the start, based on one of the hospitals in the study, performance pay was the vehicle for introducing the organisational change, and the means of making it operative. Indeed, management saw it quite explicitly in

[8]. The correlation between responses to this question and the scaled measure of commitment was 0.736 significant at the 1% level.

this way. In the Inland Revenue's documents to explain performance management to its employees, it stressed the need to modernise working methods (Inland Revenue, 1995).

Hence there is a real paradox. Performance pay played an important part in the rise in productivity, but it did not do so in the way expected, by improving incentives and raising motivation. Instead, it achieved this result because it was the framework within which management was able to renegotiate performance norms in major areas of the public services. Performance appraisals, and particularly the fixing of objectives at the start of the cycle, are critical. Seen in this light, successive governments' policies to extend performance pay do not need to be explained by attachment to political dogma. Rather, they, and their top management teams were carrying out a successful renegotiation of performance.

The wider significance of this lies in two areas. First, it has been a major social experiment in change of reward systems from seniority-based to performance-based annual pay increments affecting many thousands of employees across a wide range of service activities including administration, service to job seekers, health and education services. In his JEL review of work on incentives, Prendergast (1999) commented on the need to extend the study of incentives beyond CEOs, sales and sports personnel. In comparison with the public employees in the present study, these categories generally have simpler metrics for their output, and so have simpler principal-agent relationships. Arguably, this gives a false view of the work situation in which incentives are applied to many other kinds of employees. In the activities covered here, top management have to rely heavily in the judgements made by line managers. Although supported by objective data, their appraisals always contain an irreducible element of judgement, and hence, top management has to delegate much of performance management to them. There are therefore at least two levels of principal-agent relationship: between line managers and those they supervise, and between top management and line management. The success, particularly of the Inland Revenue performance management system, lay in the articulation of performance measures for both relationships. The regional office targets kept up the pressure on local management teams, and these kept up the pressure on line managers to appraise performance realistically and to avoid leniency. Hence, the 'Measured Day Work' to 'Leisure Day Work' disaster of the British car industry was avoided.

The second major conclusion concerns the importance of contract theory as a supplement to that of incentives and motivation. The moral hazard paradigm of effort-minimising employees exploiting management's inability to monitor their effort is the wrong model in much of the public sector. The high levels of commitment suggest this. More

important however is that people have a view of what they agreed with their employer when they were hired, of what is their employment contract, and how this evolved subsequently. Employees who resented performance pay and found it divisive were more concerned about the change to their previously existing employment conditions and performance expectations. A more appropriate paradigm is to think of performance pay and other incentive schemes as being introduced in the context of already established employment relationships so that an element of renegotiation is always present. The higher proportion of long-term employees in the public sector may make this more of an issue there, but as work by the OECD (1993, 1997) has shown, long-term employment is a feature of all OECD economies and extends also to the private sector.

Tables and Charts

Table 1: Replies to employee attitude surveys in selected public service organisations

	Civil Service			NHS trust hospitals		Schools	
Question: % in each cell replying 'agree' or 'agree strongly'	Inland Revenue 1991	Inland Revenue 1996	Employment Service	Individual PRP trust	Group PRP trust	Primary (NAHT)	Secondary (SHA)
Pay and work orientations							
PP a good principle	57	58	72	62	52	29	42
Motivation: perceived incentive							
PP gives me an incentive to work beyond job requirements	21	18	12	32	22	8	10
PP gives me an incentive to show more initiative in my job	27	20	20	36	19	9	11
PP means good work is rewarded at last	41	19	24	47	34	38	40
Motivation: perceived divisiveness							
PP causes jealousies	62	86	78	61	51	58	70
PP makes staff less willing to assist colleagues	28	63	52	22	19	51	54
PP has made **me** less willing to cooperate with management	10	30	26	19	14	7	4
Relations with management: non-manager replies:							
Management use PP to reward their favourites	35	57	41	41	27	Na	na
There is a quota on good assessments*	74	78	74	57	36	48	45
Line manager views:							
PP has reduced staff willingness to cooperate with management	20	45	39	30	27	Na	na
PP has increased the quantity of work done	22	42	28	52	34	Na	na
N (total replies)	2,420	1,180	290	680	900	1,050	860
Response rate (%)	61	30	33	28	21	51	21

Note: based on five-point Likert scales: 'strongly disagree', 'disagree', 'no view', 'agree' and 'agree strongly'. NAHT: National Association of Head Teachers (mainly primary schools); SHA: Secondary Heads Association (mainly secondary schools). For an explanation of the nature of the surveys, see the methods appendix.

Table 2: Determinants of perceived incentive and divisiveness (individual employees)
(OLS regression: Dependent variables: perceived incentive and divisiveness)

Dependent variable ?	**Perceived incentive**			**Perceived divisiveness**		
Independent variables	Unstandardized Coefficients		Standardized Coefficients	Unstandardized Coefficients		Standardized Coefficients
	B	SE	Beta	B	SE	Beta
Operation of PRP schemes						
Effective appraisal	.195**	.020	.175**	-.213**	.019	-.194**
Mgrs set targets more clearly	.263**	.018	.250**	-.042**	.017	-.041**
No scope to raise performance§	.124+	.085	.030+	.221**	.080	.055**
Financial incentive						
Max on payscale	-.204**	.064	-.098**	.001	.060	.000
Interaction: length of service*pay_max	.011*	.005	.099*	-.001	.005	-.014
Commitment						
Affective commitment	.173**	.020	.153**	-.183**	.019	-.165**
Goal commitment	.153**	.022	.131**	.030	.021	.026
Organisational controls						
Inland Revenue 96 dummy	-.022	.052	-.010	.577**	.049	.252**
Employment Service dummy	-.189+	.120	-.029+	.396**	.113	.062**
Group trust dummy	-.085	.116	-.024	-.706**	.110	-.202**
Occupational and demographic controls						
Professional	-.159	.153	-.034	.421**	.144	.091**
Technician dummy	.165*	.079	.060*	.185**	.074	.068**
Clerical dummy	.311**	.074	.140**	.262**	.070	.120**
Service isco dummy	.475**	.193	.057**	.357*	.182	.043*
Craft dummy	.357	.703	.009	1.020+	.663	.026+
Length of Service in Org	-.016**	.004	-.130**	.014**	.004	.115**
Male dummy	-.080*	.040	-.037*	.050	.037	.023
(Constant)	-.978**	.282		-1.039**	.266	
Adjusted r2			0.203			0.264
Sig			0.000			0.000
N			2752			2752

Significance levels: ** 2%; * 5%; + 15%.
Notes: § Based on line manager judgement that staff in their office have no scope to improve their performance. Non-managers in workplaces with sample observations >19 employees. Note that analysis excludes line managers in order to use their judgements of whether employees in their workplace had too little control over their jobs to raise or change their performance.
Equations shown exclude head teachers, but their inclusion does not alter the main results, except for the occupational control variables.

Table 3: Effects of perceived incentive and divisiveness on employee performance
Logit regression: Dependent variable: probability of achieving 'superior' performance

	Eqn 1		Eqn 2	
	B	S.E.	B	S.E.
Incentives and commitment				
Perceived incentive	.372**	.040	.273*	.135
Percieved divisiveness	-.273**	0.45	-.275*	.136
Operation of appraisal and target setting				
Effective appraisal			.996**	.065
Targets set more clearly			-.149**	.050
No scope to raise performance§			-.985**	.219
Commitment				
Affective commitment			-.075#	.055
Goal commitment			-.282**	.056
Interactions				
Incentive*appraisal quality			.051	.057
Divisiveness*appraisal quality			-.143**	.056
Incentive*targets			.015	.044
Divisiveness*targets			.062#	.048
Incentive*divisiveness			.000	.048
Occupational and demographic controls				
Professionals	-1.878**	.285	-1.638**	.325
Technicians	.302+	.182	.299#	.197
Clerical	.198	.169	.306+	.184
Service employees	-6.968+	3.871	-6.596#	4.262
Craft	-1.1181	.935	-1.922#	1.508
Length of service	.021**	.005	.026**	.006
Male (dummy)	.009	.092	-.155#	.102
Organisational controls				
Inland Revenue 96	-.259**	.101	.792**	.131
Employment Service	-2.547**	.569	-2.242**	.594
NHS trust hospitals	.509**	.215	.826**	.249
Schools (not included)	-	-	-	-
Constant	-.883**	.198	2.038**	.722
R2 (Cox & Snell)	.125		0.226	
R2 (Nagelkerke)	.171		0.308	
% correctly predicted	65.6		72.0	
N	2991		2819	

Note: superior performance includes 'exceed' and 'succeed at extra-loaded' jobs.
§ Based on line manager judgement that staff in their office have no scope to improve their performance.
** 2%; * 5%; + 10%; # 20%.
Equations shown exclude head teachers, but their inclusion does not alter the main results, except for the occupational control variables.

Table 4: Year-to-year rank correlations by regional office for performance targets

Mean rank correlation with base year (Year 0).	Target clearance	Target quality	Clearance gap	Quality gap	% of employees with superior performance
Year 1	0.462	0.584	0.255	0.469	0.789
Year 2	0.267	0.264	-0.099	0.365	0.507
Year 3	0.194	0.087	-0.207	0.224	0.129
Year 4	0.251	-0.007	0.031	0.123	-0.159

Note: Inland Revenue, Spearman rank correlations among 10 regional offices.

Table 5: Correlations between key measures of IR performance

	Clearance gap	Quality gap	% Change in clearance target at end of year	% Change in quality target at end of year	% of employees with superior performance	Productivity
Clearance gap (shortfall or excess on targets)	1.000	.264**	.641**	.338**	-.472**	-.451**
Quality gap (shortfall or excess on targets)		1.000	.102	.277**	-.373**	-.187+
Change in clearance target at end of year %			1.000	.451**	-.303**	-.181#
Change in quality target at end of year %				1.000	-.236*	-.336**
% of employees with superior performance					1.000	.612**
Productivity: tax revenues/salary bill						1.000
Relative basic salaries						
N	80	80	80	80	70	70 (8)

Correlations across 10 individual offices and over 8 years.
N=80 for clearance and quality data and N = 70 for average employee performance measures; Note also that the measures of output (tax revenue/salary bill, and relative basic salary have a single value for all offices in any given year). Note that the productivity and relative salary data are available for the whole organization only, giving eight different values.
Significance: ** at 2% level (2-tailed), * at 5%, + at 10%, and at # 20%.

Table 6: Management pressure within appraisals? Inland Revenue 1996

Dependent variable ?	**Effective appraisal**	**Perceived incentive**	**Perceived divisiveness**	**Superior performance**
OLS regressions: standardised beta coefficients				
Staff pressured to accept management's performance objectives.	-.115**	-.154**	.191**	.070+
Everyone is given the same targets.	-.047#	-.109**	.022	-.052#
Those getting good appraisals are cleverest at negotiating their performance agreements	-.114**	-.063+	.124**	-.075*
Agree my objectives to avoid a bad appraisal	-.211**	-.121**	.104**	-.270**
Managers use PRP to reward their favourites	-.122**	-.050#	.354**	-.070+
There is a quota on good appraisals	-.059#	-.117**	.095**	-.028
N	770	758	758	741
Adjusted R^2	.158	.142	.348	.095
Sig	**	**	**	**

Note: the table includes the same occupational and demographic control variables as in Table 2.
Significance: ** 2%, * 5%, + 10%, # 20%.

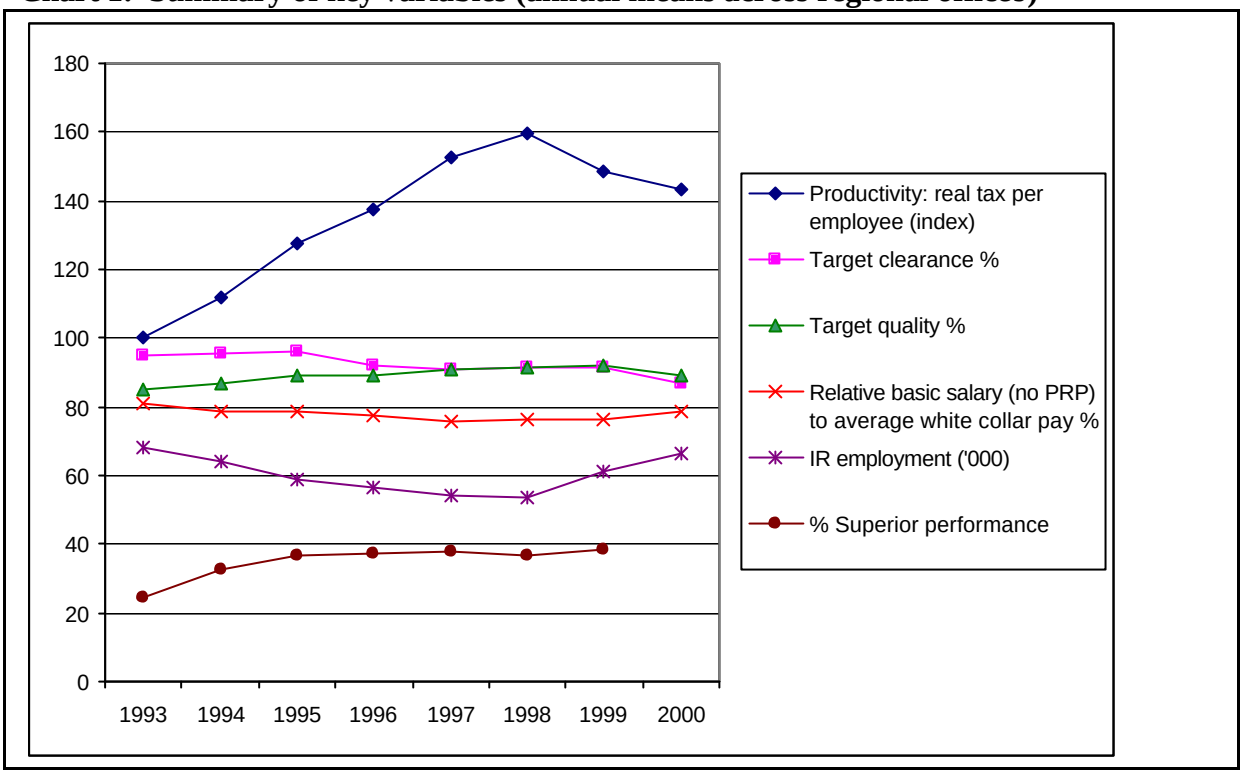
Table 7: Intensity of re-negotiation and perceived divisiveness

	Divisiveness: Standardised Mean	Standard error	Role of individual agreement	Role of collective agreement
Inland Revenue 1996	0.472	.035	Compulsory for all	1993 pay agreement after strike
Employment Service	0.252	.061	Compulsory for all	Series of agreements for different staff grades 1994-95
Schools: Head teachers	0.142	.060	Compulsory if adopted by school governors	Implemented by government after pay review as one criterion for pay awards by school governors
Hospital with individual PRP	-0.041	.066	Voluntary for current staff	Implemented by local mgt; subsequent agreement with unions
Inland Revenue 1991	-0.158	.067	Compulsory but no losers	1988 pay agreement
Hospital with trust-wide bonus	-0.486	.067	Voluntary for current staff	Implemented by local mgt; subsequent agreement with unions

Note: mean perceived divisiveness for all organisations combined is 0, with a standard deviation of 1, and a mean for each organisation of between 0.9 and 1.

The standardised means are derived using the organisational dummies and constant term based on the equations in Table 2, but excluding the question on line manager judgements of scope to raise performance which could not be asked of head teachers, and excluding that on whether someone was on the maximum for their grade because there was no limit on performance awards at the top of the grade for head teachers or for staff in the individual PRP hospital. This makes no difference to the rank order of divisiveness by organisation, nor does using the raw mean calculated directly from the sample.

Chart 1: Summary of key variables (annual means across regional offices)



Source: Inland Revenue Report and Accounts, (Board of Inland Revenue, various years), Economic Trends, Civil Service Statistics.

Figure 1: Overtime and performance pay compared

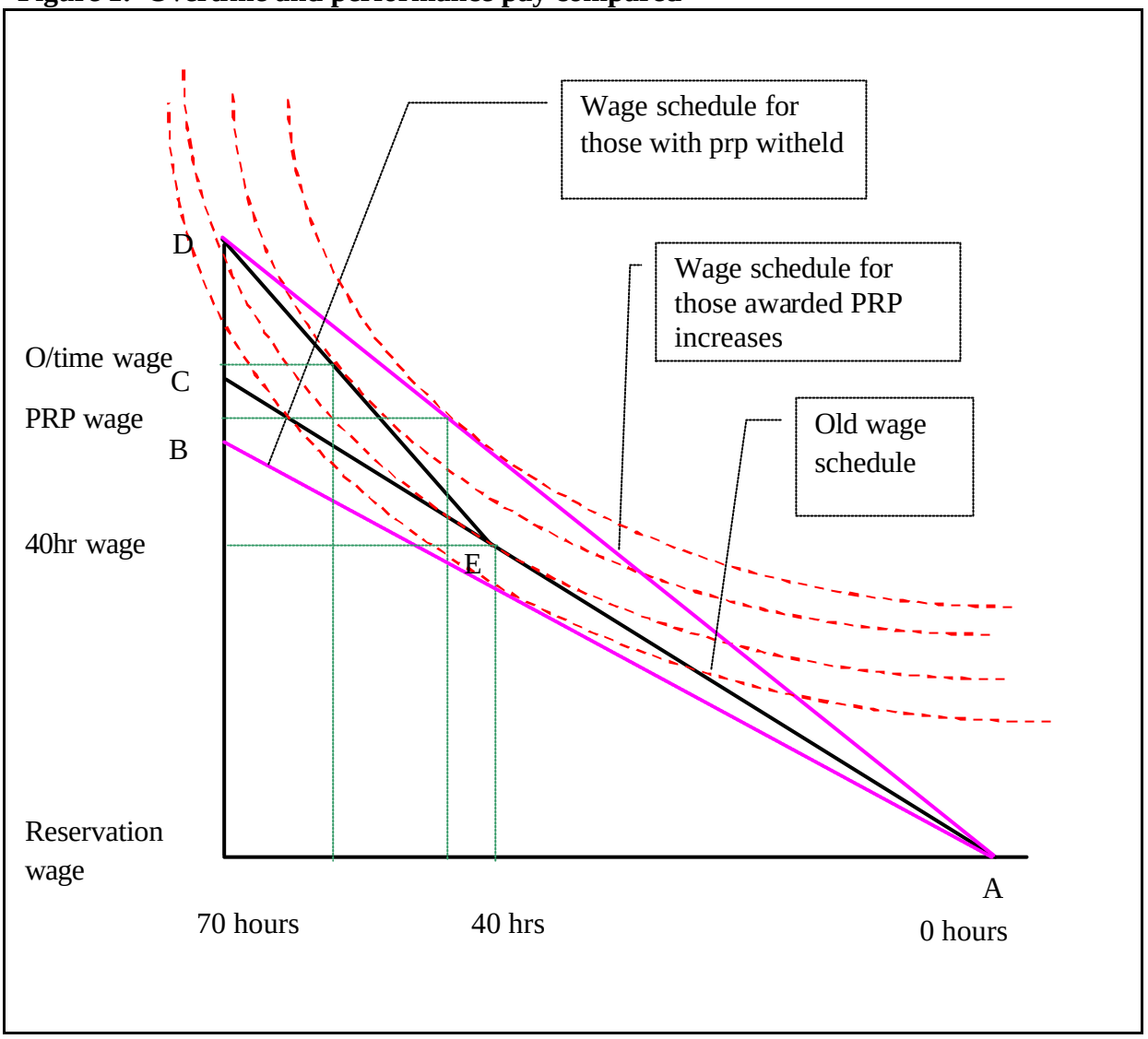


Figure 2: Different links between appraisal and performance

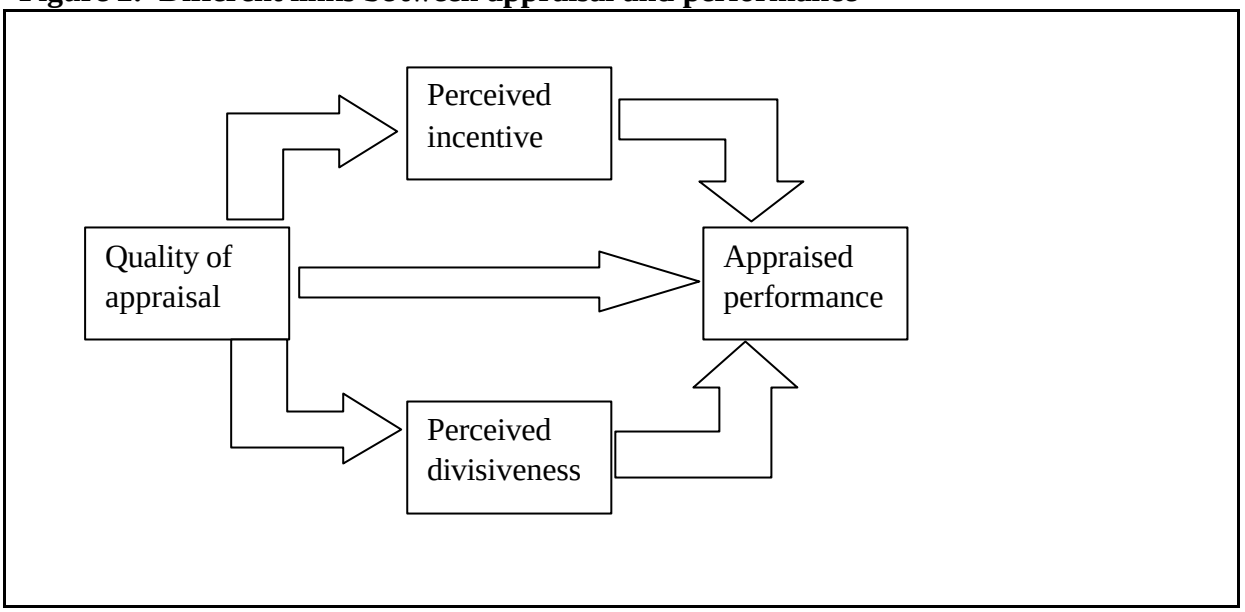


Figure 3: Performance indicators under the 'leniency' and the 'productivity' hypotheses.

	Leniency	**Productivity**
Procedural indicators		
Job classification for performance measurement	Not relevant	Essential
Appraisal system checks & balances	Not relevant	Essential
Behavioural indictors		
Inter-office performance differences	Persistently poor & good performers	Gradual improvement of poor performing offices, & regression towards mean of top performers
Achievement and revision of office targets	Targets not revised because no one takes them seriously	Targets adjusted in the light of past performance to keep them stretching
Awarding good appraisals and performance pay	Weak management hand out PRP to buy peace >> correlates with missed targets & persistently generous offices.	Active management use PRP to mobilise staff for office targets. Correlates with pressure on targets but no persistently generous offices
Increased demand for tax office services (econ growth)	Increased pressure for more staff & increased delays in work: productivity static.	Tight control of staffing, & performance targets maintained: productivity rises.

Methods Appendix

A1 Details of Schemes and Sources

Summary details of performance pay schemes studied

Table A1: Summary details of performance pay schemes studied

Organisation	Type of scheme	Treatment of employees at the top of their respective pay span	Per cent of employees on their pay span maximum
Inland Revenue 1991	Employees move up existing seniority pay scale faster on receipt of good appraisal by line manager. Appraisal against standardised criteria.	Smaller % merit increases for higher level grades and limit of 3 increments above span max for merit payments.	69.1%
Inland Revenue 1996	No seniority scales. Appraised as 'Succeeding' at agreed targets brings pay increase, and 'Exceeding' brings additional increase, as does 'Succeeding' at jobs classified 'extra loaded'. No inflation increase in some years.	Smaller % merit payments as staff progress up the pay span for their grade, and restrictions on overlapping with grade above	50.9%
Employment Service	No seniority scales. Pay increase depends on achieving appraised performance objectives & is based on a share of a negotiated pot.	Performance pay above the maximum for the grade is non-consolidated	58.9%
NHS hospital – individual PRP	No seniority scale. Pay increase dependent on appraised individual performance.	No scale max but bonus for above average performance is non-consolidated	Not applicable
NHS hospital – trust-wide bonus	No seniority scale. Pay increase depends on trust-wide bonus, poor performers only excluded.	Bonus at the grade maximum becomes entirely non-consolidated	29.6% (trust contracts); 80.1% (Whitley contracts)
School head teachers	Additional movement up pay spine for appraised excellent performance by school governors. No seniority increments	No limit on additional spine points that may be awarded	Not applicable

All of the schemes had been in operation for about three years before they were surveyed so that many initial teething problems should have been overcome. Union involvement has been secondary. Management made the initial decision on the design and implementation, and only subsequently were the unions involved.

a) Details of the employee attitude survey

Attitudinal data were collected by postal questionnaire sent to individual employees in each organisation. Mostly these were completed in the employee's own time. In three

organisations, management cooperated with the study, enabling lists of employees to be used for drawing the sample, and the internal mail for distributing and receiving back questionnaires. Lacking management support for the 1996-97 surveys of the civil service departments and for schools, union membership lists were used, but membership density is very high. It was 90% in the Inland Revenue, 60% in the Employment Service middle management grades studied, and around 90% among head teachers. Public hospitals are also highly unionised. Questionnaires were sent to the grades of staff organised by the Inland Revenue Staffs Federation, later merged into the PCS, for the tax service, and included staff from skilled secretary up to tax inspectors. Other unions organised more junior and more senior staff, respectively, the CPSA and the First Division Association. In the Employment Service, the sample included middle-management grades organised by the PCS. In the hospitals, all staff were included except medical doctors who were not subject to the schemes. Among head teachers, a sample was drawn from all those on the unions' membership lists.

Most of the attitudinal questions used 5-point Likert scales, ranging from disagree strongly to agree strongly. Questions were piloted with groups of employees or where management cooperation was lacking, with groups of union members. Preliminary results were presented to the organisations and interpretations discussed with management and unions in feedback seminars.

The questionnaires were divided into sections. Each dealt with a specific aspect: general attitudes to pay and performance; employee judgements of whether or not it gave them an incentive, their personal experience with their most recent performance appraisal; and line-managers' views of the effects of the scheme on staff. The full text of the questionnaires can be found in Marsden and Richardson (1992) and Marsden and French (1998).

The average survey response rate was 43%, ranging between about 60% in the tax service in 1991 to about 20% in one of the hospitals (see Table 1 main text). The questionnaires were long, mostly over 100 questions, imposing a considerable time burden on respondents.

Response patterns were compared with such demographic and other breakdowns as were available. Response rates were higher among the more managerial occupations, but all occupational levels were well represented in the sample. Response by gender and by age or length of service, and where asked, by ethnic background, and full- and part-time showed no great divergence from the organisations' employment figures. There was also a good response from across the regional offices of the tax and the employment services. Response

patterns were compared with appraisal markings and found to be very similar across performance ratings.

b) The archival data

The development of measures of organisational performance has gone hand in hand with the development of performance pay and the spread of new public management methods. The best data related to performance pay exist for the tax service, and there the range and complexity of performance indicators reported in the Annual Report and Accounts has improved over time. Similar types of performance data are available for the other organisations, but they are less complete and less consistent over time.

The administrative unit measures for the tax service consist of the distribution of appraisal scores by regional office for each of the ten regional offices for the eight years from 1993-2001. These were compiled by management and made available to the union as part of the consultation process. Similar data were available for the late 1980s and early 1990s, but there is a break in the series owing to renegotiation of the performance management scheme in 1992-93, and a radical restructuring of the regional offices, so that the time series cannot be linked between the two periods. In any case, the published data on targets for regional offices (administrative units) can only really be used from 1993. From the text of the annual reports, it appears that the tax service used these indicators as part of its internal management process, and there are several cases in which new indicators especially of quality of service were published provisionally before being fully used to monitor performance. The Annual Reports are published before Parliament, and scrutinised by the influential National Audit Office, which is responsible for monitoring the quality of public spending, and which has played a big part in diffusing new ideas on improved management across the public sector.

The tax service indicators are listed below, giving five quality and six clearance targets for the time period. Some other indicators were not available for all years, and so could not be included in the time series. As there was no obvious way to weight these, unweighted means were computed for the quality and clearance targets and their reported outcomes.

Quality of Service

- Calculating tax correctly in every respect first time
    - Schedule D (self employed)

    - Schedule E (employees, excluding wholly by computer)
- Telephone calls answered within 30 seconds
- Personal callers seen within 15 minutes
- Repayment claims dealt with in < 28 days

Clearance

- Correspondence dealt with in <28 days
- Clerical work (schedule E): % of taxpayers cleared by April
- Clearance: large cases (from previous years)
- Clearance: other cases (from previous years)
- Schedule D Self Assessment: returns processed by Sept 30th
- Collection: average monthly clearance (PAYE Band 1 assessed taxes)

The outcomes reported were checked by the tax service's internal auditor.

Indicators for the performance of the Employment Service were also published in its annual report and accounts to Parliament, and those for the NHS trust hospitals were published in the NHS Performance Guide.

A.2 Measures of Key Variables from the Survey Data

a) Measures of employee performance: validity and reliability

The performance appraisal systems used, especially after the first of the tax service studies, drew heavily on the experience of outside consultants. The systems used in the two hospitals were the Lloyd Masters system and Mediquate systems that are quite widely used in the health sector. The scheme in the tax service that was in operation in 1996 had a substantial input from private consultants, and incorporated many 'best practice' ideas from the private sector and from the HR profession generally. Indeed, even the scheme in operation at the time of the 1991 survey met many of the criteria for good appraisal set out by the government's Arbitration, Conciliation and Advisory Service, (ACAS, 1990). Through the 1980s and 1990s, the public sector made extensive use of private sector consultancy organisations. The schemes contained a number of checks and balances, notably, except for

head teachers, all line-manager appraisals were vetted by a higher level manager. The overall distribution of appraisal scores was also made available to the unions, and was monitored by management to ensure the schemes were operated without bias and to protect them against an upward drift in performance ratings. Measures of internal performance were also checked by the Audit Office which has overall responsibility for monitoring the quality of public spending. All of these help to ensure the reliability of individual performance ratings. Their validity, whether they represent actual performance, is a more difficult question, but as was shown in the main body of the article, the 'leniency' hypothesis proved less plausible than the 'productivity' hypothesis.

b) Derivation of measures of perceived incentive and perceived divisiveness

The variables measuring perceived incentive and perceived divisiveness were based on the questions shown in Table A2.

Table A2: Derivation of measures of perceived incentive and divisiveness by factor analysis

	Factor 1: Perceived incentive	Factor 2: perceived divisiveness
PRP means good work is rewarded	**.558**	-.487
PRP gives me an incentive to work beyond job requirements	**.888**	-.111
PRP gives me an incentive to take more initiative in my job	**.902**	-3.963E-02
PP causes jealousies/resentment	-6.312E-02	**.770**
PRP Undermines teams	-.122	**.776**
PRP has reduced my wish to cooperate with management	-9.746E-02	**.645**

Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization.
a Rotated Component Matrix, rotation converged in 3 iterations.

c) Derivation of the commitment variable

Organisational commitment is measured by adapting the established scales based on Meyer and Allen (1997), and the factor weightings for each of the questions used are show in Table A 3. Because this was cross-sectional survey, it was not possible to test whether commitment had the correct antecedents, but it was possible to test its correlates. Meyer and Allen's (1997) survey indicates a number of antecedents, most notably an employee's length of service in an organisation. Because professionals have a dual commitment, to their occupation and to their employing organisation one would expect their organisational

commitment to be less than that of other employees in jobs requiring organisation specific skills. Gender roles might also cause women employees to display higher levels of commitment especially in service organisations because of the greater emphasis on interpersonal relations. A simple regression showed that the measure of commitment increased strongly with length of service; it was lower among professionals than other occupations; and it was lower among males than females. These three variables are useful because they are not affected by the operation of PRP and so barring any self-selection effects, can be construed to be independent.

Table A3: Factor analysis of commitment variables

	Affective commitment	Goal commitment
Working in the Org. means a great deal to me	**.805**	.263
I feel "part of the family" in my present office/hospital/school	**.750**	-1.374E-02
I would be very happy to spend the rest of my career with the org.	**.731**	8.319E-02
I do **not** feel 'emotionally attached' to the Inland Revenue	**.714**	.189
Whenever changes made in this org employees usually lose out in the end	**-.387**	-.103
I think that I could easily become as attached to another organisation as I am to the Inland Revenue	**-.506**	-2.050E-02
By working in the Organisation, I feel that I am contributing to an important public service	.313	**.634**
Don't award PRP to retain staff	5.185E-02	**-.869**

d) Tests of the relationship between appraisal effectiveness and performance

Further analysis of the relationship between the effectiveness of the appraisal process and performance is made possible by the more detailed descriptive questions relating to the conduct of appraisal in the two hospitals. These can be used to check the validity of the more general questions on appraisal asked of all the organisations, and to check whether there are significant feedback effects from employees' performance ratings onto their judgements of the appraisal process.

Two tests of the extent of such feedback effects were carried out.

i) Use of more descriptive questions on the appraisal process

First, additional information from the two hospitals provides a number of more detailed, concrete, questions about the conduct of the appraisal process. These had been asked because

the HR managers who agreed access for the research were keen to know how well their appraisal schemes were working. Generally, the more concrete the question, the less likely it is that replies will be coloured by other related events.

These more concrete questions were simplified by factor analysis into three components, consultation, supportiveness and clarity of the appraisal (Table A4), and the component scores then correlated with the measure of effective appraisal used in the article (Table A5).

Table A4: Factor analysis of appraisal quality (Trust-wide bonus hospital)

	1 Consultation	2 Supportive	3 Clarity
Throughout the last year, I had sufficient opportunity to discuss my performance with my line manager	**.895**	-.193	.214
In the last year, I have had sufficient opportunity to discuss and clarify my role with my line manager	**.870**	-.241	.215
In the last year, I have had sufficient opportunity to identify objectives and targets with my line manager	**.855**	-.223	.248
In the last year, I have had sufficient opportunity to discuss my personal development needs with my line manager	**.843**	-.258	.264
I found the discussion irrelevant	-.219	**.835**	-.151
I found the discussion superficial	-.218	**.808**	-.212
I found the discussion threatening	-.120	**.757**	-.104
I found the discussion useful	.318	**-.679**	.297
I am clear about my current objectives and targets	.227	-.142	**.870**
I am clear about my current job role	.151	-.145	**.849**
I am clear about my personal development needs	.209	-.237	**.637**
I understand my manager's rating of my performance	.425	-.222	**.577**

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

Table A5: Correlations of 'effective appraisal' with the factors based on detailed questions (Trust-wide bonus hospital)

	Factor 1, Consultation	Factor 2, Supportive	Factor 3, Clarity
Effective appraisal	.376**	-.244**	.747**
N	367	367	367

** Correlation is significant at the 0.01 level (2-tailed).

ii) Two-stage least squares test of the effect of effective appraisal on performance

A second test was to use two-stage least squares, to test the model shown in Tables 2 and 3 in the main text, to see whether effective appraisal led to perceptions of incentive or divisiveness which, in turn, determined performance scores. The two-stage least squares

enables one to cut out the possible feedback from getting a good award onto perceptions of incentive and divisiveness.

As shown in Table A6, the standardised beta coefficients are quite large given the units of measurement, have the correct sign, and are strongly significant, and so confirm the influence of appraisal on performance, passing through perceived incentive and divisiveness.

Table A6: Estimated effects of perceived incentive and divisiveness on individual employee performance (two-stage least squares)
Dependent variable: Superior performance attained

	Individual survey data	
Explanatory variable	Perceived incentive	Perceived divisiveness
Multiple R	0.1749	.2213
B	0.2298**	-0.3441**
Beta	0.4828	-0.7138
Sig	0.0000	0.0000
N	2752	2725

Notes:
Individual employee data: Instruments: effective appraisal; improved goal setting; being on the grade maximum, and lack of scope to improve.
Note: dummies were included for the Inland Revenue 1991 and 1996, but they made very little difference to the results.
Significance: **, 2%; *, 5%.

Thus, although neither of these tests is definitive on its own, they both suggest that feedback effects from performance awards onto judgements of effective appraisal, although no doubt present to some degree, are sufficiently weak for the two variables to be treated as independent.

References

ACAS (1990), *Appraisal Related Pay*, ACAS Advisory Booklet No. 14, HMSO: London.

Armstrong, M. and Murlis, H. (1994), *Reward Management: a Handbook of Remuneration Strategy and Practice*, 3rd edn, Kogan Page: London.

Audit Commission (1999), *A Measure Of Success: Setting and Monitoring Local Performance Targets*, Audit Commission for Local Authorities and the National Health Service in England and Wales, London.

Axelrod, R. (1984), *The Evolution of Cooperation*, Basic Books: New York.

Betters, P. (1931), *The Personnel Classification Board: its History, Activities and Organisation*, Brookings Institution: Washington.

Bichard, M. (Chair) (1999), *Performance Management: Civil Service Reform - a Report to the Meeting of Permanent Heads of Departments*, Sunningdale 30 September-1 October 1999, Cabinet Office, London. Ref: CABI-5195/9912/D4.

Board of Inland Revenue (Great Britain) (2001), *Inland Revenue Annual Report: Report of the Commissioners of Her Majesty's Inland Revenue for the Year Ending 31st March 2000: One Hundred and Forty-Second* Report, Cm. 5029, Stationery Office: London.

Boswell, W. and Boudreau, J. W. (2000), 'Employee Satisfaction with Performance Appraisals and Appraisers: the Role of Perceived Appraisal Use', Human Resource Development Quarterly, 11: 3, Fall, pp. 283-299.

Brown, M. and Heywood, J. S. (eds.) (2002), Paying for Performance: an International Comparison, M. E. Sharpe Inc. Armonk: NY.

Brown, T. C. and Latham, G. P. (2000), 'The Effects of Goal Setting and Self-Instruction Training on the Performance of Unionized Employees', Relations Industrielles, 55: 1, pp. 80-95.

Carroll, M. (1993), *Performance-Related Pay: A Comparative Study of the Inland Revenue and a High Street Bank*, MSc dissertation, Faculty of Technology, UMIST: Manchester, October.

Coase, R. H. (1937), 'The Nature of the Firm', Economica, November, pp. 386-405.

Drago, R. and Garvey, G. T. (1998), 'Incentives for Helping on the Job: Theory and Evidence,' Journal of Labor Economics, 16: 1, pp. 1-25.

Elvira, M. and Town, R. (2001), 'The Effects of Race and Worker Productivity on Performance Evaluations', Industrial Relations, 40: 1, pp. 571-590.

Employment Service (1997), *Annual Report and Accounts 1996-97*, HC 270, The Stationery Office: London.

Furnham, A. (1997), *The Psychology of Behaviour at Work: the Individual in the Organization*, Psychology Press: Hove and New York.

Hamermesh, D, and Rees, A. (1993), *The Economics of Work and Pay* (5th edn), Harper Collins: New York.

Heery, E. (1998), 'A Return to Contract? Performance Related Pay in a Public Service', Work, Employment & Society, 12: 1, March, pp. 73-95.

Holmstrom, B. and Milgrom, P. (1991), 'Multitask Principal-Agent Analysis: Incentive Contracts, Asset Ownership, and Job Design', The Journal of Law, Economics and Organization, Vol. 7.

Industrial Relations Services (IRS) (1999), *IRS Study for PCS on Performance Related Pay*, Public and Commercial Services Union (PCS), London.

Inland Revenue (1994), 'Second Quality Review of Performance Management', *Chairman's Newsletter*, 10.04.1994, Inland Revenue: London.

Inland Revenue (1995), *The Guide* (management's guide for employees to Performance Management), Inland Revenue: London.

Inland Revenue Department Whitley Council (1991), *Review of Staff Appraisal and Reporting System: Staff Attitude Survey Second Analysis*, Vols. 1-3, Inland Revenue Department Whitley Council, London, November, mimeo.

Kessler, I. and Purcell, J. (1993), *Discussion Paper on Staff Pay Survey at Amersham International*, Mimeo, Templeton College: Oxford.

Landy, F. and Farr, J. (1983), *The Measurement Of Work Performance: Methods, Theory, and Applications*, Academic Press: New York.

Latham, G. P. and Lee, T. W. (1986), 'Goal Setting', in E. A. Locke (ed.), *Generalizing from Laboratory to Field* Settings, Lexington Books: D. C. Heath, Lexington, Ma.

Lawler, E. E. III. (1971), *Pay and Organisational Effectiveness: A Psychological View*. McGraw-Hill: New York.

Lazear, E. (1995), *Personnel economics*. MIT Press: Cambridge Mass.

Lazear, E. P. (1998), *Personnel Economics for Managers*, Wiley: New York.

Locke, E. A. and Latham, G. P. (1990), *A Theory Of Goal Setting and Task Performance*, Prentice – Hall: Englewood Cliffs, NJ.

Maguire, M. (ed.) (1993), *Pay Flexibility in the Public Sector*, OECD: Paris.

Makinson, J. (Chair) (2000), *Incentives for Change: Rewarding Performance in National Government Networks*, Public Services Productivity Panel, HM Treasury: London.

Malcomson, J. (1997), Contracts, Hold-Up, and Labor Markets, Journal of Economic Literature, 35:4, December, pp. 1916-1957.

Marsden, D. W. and Richardson, R. (1992), 'Motivation and Performance Related Pay in the Public Sector: A Case Study of the Inland Revenue', Discussion Paper No. 75, Centre for Economic Performance, London School of Economics, London. Available under publications, discussion papers online from http://cep.lse.ac.uk

Marsden, D. W. and Richardson, R. (1994), 'Performing for Pay? The Effects of 'Merit Pay' On Motivation in a Public Service', British Journal of Industrial Relations, 32:2, pp. 243-262, June.

Marsden, D. W. and French, S. (1998), *What a Performance: Performance Related Pay in the Public Services*, Centre for Economic Performance Special Report, London School of Economics, London. Available under publications, special reports online from http://cep.lse.ac.uk

Meyer, J. P, and Allen, N. J. (1997), *Commitment in the Workplace: Theory, Research and Application*, Sage: Thousand Oaks, California.

Milgrom, P, and Roberts, J. (1992), *Economics, Organisation and Management*, Prentice-Hall: New Jersey.

Milkovich, G. T, and Wigdor, A. K. (eds.) (1991), *Pay for Performance: Evaluating Performance Appraisal and Merit Pay*, National Academy Press: Washington D.C.

Mitchell, D. J. B., Lewin, D. and Lawler, E. III, (1990), 'Alternative Pay Systems, Firm Performance, and Productivity', in A. S. Blinder (ed.), *Paying for Productivity*, Brookings Institution: Washington DC.

NAO, National Audit Office (1989), *Manpower Planning in the Civil Service*, Cm 398, HMSO: London.

NHS Executive, Department of Health (1997), *The NHS Performance Guide 1996-97*, NHS Executive, Department of Health: London.

OECD (1993), *Employment Outlook* 1993, OECD: Paris.

OECD (1997), *Employment Outlook* 1997, OECD: Paris

OECD-PUMA, Human Resources Management (HRM) Working Party (2002), *Summary Record 2002 Human Resources Management Working Party Meeting*, OECD, PUMA, Human Resources Management (HRM) Working Party, OECD Headquarters: Paris, 7-8 Oct 2002.

OECD-PUMA Äijälä, Kirsi (2001), *Public Sector: An Employer of Choice?* Report on The Competitive Public Employer Project OECD Public Management Service, OECD: Paris.

Porter, L. W. and Lawler, E. E. (1968), *Managerial Attitudes and* Performance, Dorsey Press: Homewood, Illinois.

Prendergast, C. (1999), 'The Provision of Incentives Within Firms', Journal of Economic Literature, 37, pp. 7-63.

Teulings, C. and Hartog, J. (1998), *Corporatism or Competition? Labour Contracts, Institutions and Wage Structures in International Comparison*, Cambridge University Press: Cambridge.

Thompson, M. (1993), *Pay and Performance: the Employee Experience*, IMS Report 218, Institute of Manpower Studies: Brighton.

Vroom, V. H. (1964), *Work and Motivation*, Wiley: New York.

Williamson, O. E. (1975), *Markets and Hierarchies: Analysis and Antitrust Implications*, Free Press: New York.

CENTRE FOR ECONOMIC PERFORMANCE
Recent Discussion Papers

577	S. Nickell	A Picture of European Unemployment: Success and Failure
576	A. de Coulon M. Piracha	Self-Selection and the Performance of Return Migrants: the Source Country Perspective
575	H. Steedman K. Wagner J. Foreman	The Impact on Firms of ICT Skill-Supply Strategies: An Anglo-German Comparison
574	S. Gibbons	The Costs of Urban Property Crime
573	R. Griffith S. Redding H. Simpson	Productivity Convergence and Foreign Ownership at the Establishment Level
572	S. Redding P. K. Schott	Distance, Skill Deepening and Development: Will Peripheral Countries Ever Get Rich?
571	B. Petrongolo C. A. Pissarides	Scale Effects in Markets with Search
570	M. Coles B. Petrongolo	A Test Between Unemployment Theories Using Matching Data
569	A. Bryson L. Cappellari C. Lucifora	Does Union Membership Really Reduce Job Satisfaction?
568	A. Bryson R. Gomez	Segmentation, Switching Costs and the Demand for Unionization in Britain
567	M. Gutiérrez-Domènech	Employment After Motherhood: A European Comparison
566	T. Kirchmaier	The Performance Effects of European Demergers
565	P. Lopez-Garcia	Labour Market Performance and Start-Up Costs: OECD Evidence

564	A. Manning	The Real Thin Theory: Monopsony in Modern Labour Markets
563	D. Quah	Digital Goods and the New Economy
562	H. Gospel P. Willman	High Performance Workplaces: the Role of Employee Involvement in a Modern Economy. Evidence on the EU Directive Establishing a General Framework for Informing and Consulting Employees
561	L. R. Ngai	Barriers and the Transition to Modern Growth
560	M. J. Conyon R. B. Freeman	Shared Modes of Compensation and Firm Performance: UK Evidence
559	R. B. Freeman R. Schettkat	Marketization of Production and the US-Europe Employment Gap
558	R. B. Freeman	The Labour Market in the New Information Economy
557	R. B. Freeman	Institutional Differences and Economic Performance Among OECD Countries
556	M. GutiJrrez-DomPnech	The Impact of the Labour Market on the Timing of Marriage and Births in Spain
555	H. Gospel J. Foreman	The Provision of Training in Britain: Case Studies of Inter-Firm Coordination
554	S. Machin	Factors of Convergence and Divergence in Union Membership
553	J. Blanden S. Machin	Cross-Generation Correlations of Union Status for Young People in Britain
552	D. Devroye R. B. Freeman	Does Inequality in Skills Explain Inequality of Earnings Across Advanced Countries?
551	M. Guadalupe	The Hidden Costs of Fixed Term Contracts: the Impact on Work Accidents

To order a discussion paper, please contact the Publications Unit
Tel 020 7955 7673 Fax 020 7955 7595 Email info@cep.lse.ac.uk
Web site http://cep.lse.ac.uk