

The use of sanctions and rewards in the public sector.

Literature review accompanying the
report prepared for the NAO by Deloitte

10 September 2008

Important Notice from Deloitte

This report has been prepared by Deloitte & Touche LLP ("Deloitte") for the NAO in accordance with the contract with them dated 18 March 2008 and on the basis of the scope and limitations set out below.

The report has been prepared solely for the purposes of the NAO, to assist them with their understanding of the use and effectiveness of sanction/reward mechanisms in the UK public sector. The distribution of this document to parties other than the NAO is subject to the restrictions on use specified in the contract. No other party is entitled to rely on this document for any purpose whatsoever and we accept no responsibility or liability to any other party in respect of the contents of the report. Deloitte accepts no responsibility for any reliance that may be placed on this document should it be used by any party other than the NAO or for any purpose that is not in accordance with the terms of the contract with the NAO.

The scope of our work has been limited by the time, information and explanations made available to us. We have relied upon the documents and data provided by the NAO and on third-party data sources which have been clearly referenced in the appropriate sections of this report. We have no responsibility for the accuracy or completeness of this information which we have reviewed but not validated.

Accordingly, no representation or warranty, express or implied, is given and no responsibility or liability is or will be accepted by or on behalf of Deloitte or by any of its partners, employees or agents or any other person as to the accuracy, completeness or correctness of the information contained in this document or any oral information made available and any such liability is expressly disclaimed.

Deloitte & Touche LLP is authorised and regulated by the Financial Services Authority.

Deloitte & Touche LLP is a limited liability partnership registered in England and Wales with registered number OC303675 and its registered office at Stonecutter Court, 1 Stonecutter Street, London EC4A 4TR, United Kingdom. Deloitte & Touche LLP is the United Kingdom member firm of Deloitte Touche Tohmatsu ('DTT'), a Swiss Verein whose member firms are separate and independent legal entities. Neither DTT nor any of its member firms has any liability for each other's acts or omissions. Services are provided by member firms or their subsidiaries and not by DTT.

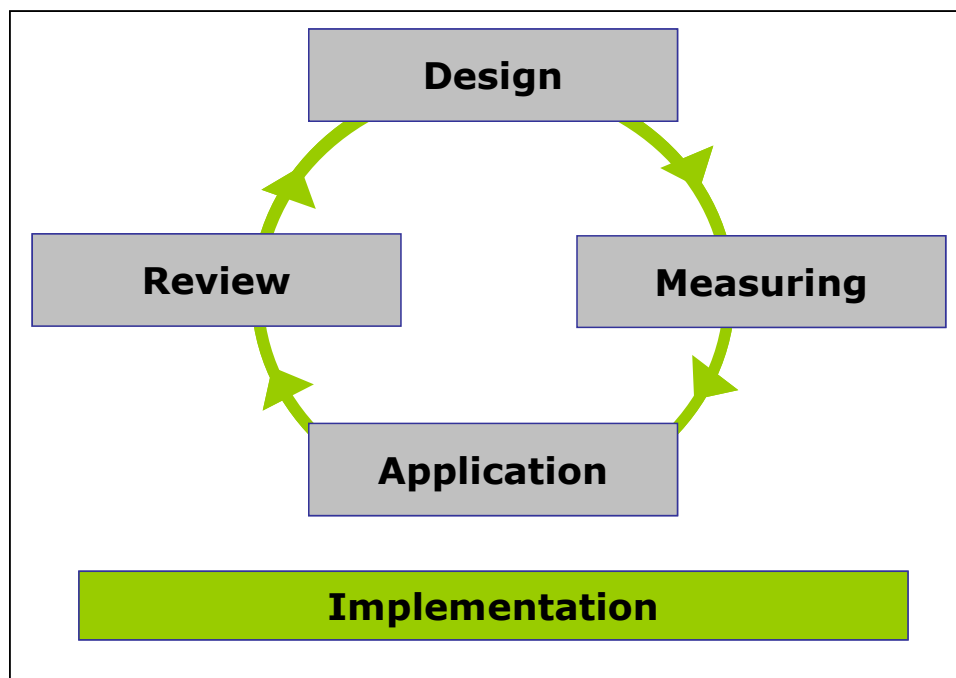
Contents

- Important Notice from Deloitte 2
- 1 Introduction 4
- 2 Economic theory of incentives 6
 - 2.1 General Insights6
 - 2.2 Public Sector Insights7
- 3 Design features which lead to effective sanction/reward mechanisms 11
 - 3.1 The sanction/reward mechanism should be developed together with the agents and all other stakeholders 11
 - 3.2 Incentivise the right performance measure 12
 - 3.3 The agents should be able to influence the outcome which is being incentivised 18
 - 3.4 Motivation of agents 20
 - 3.5 Teams and incentives 22
 - 3.6 The sanctions or rewards must be of sufficient value to incentivise behaviour 24
- 4 Measurement issues and requirements 27
 - 4.1 Systems are in place to collect good quality performance data 27
- 5 Application issues and requirements 29
 - 5.1 There must be consistency in the application of the reward/sanction mechanism to create credibility²⁹
- 6 Issues and requirements during the review phase 30
 - 6.1 The sanction/reward mechanism must be regularly reassessed 30
 - 6.2 Assessment of effectiveness should be fit-for-purpose 30
- 7 Implementation 35
- Appendix: Bibliography 36

1 Introduction

1. There is a body of empirical evidence that agents¹, including those in the public sector, do respond to incentives and that incentives can promote effort and performance (e.g. Prendergast, 1999, Lazear, 2000, Prentice et al, 2007). In this review, we will survey this evidence and understand how sanction/reward mechanisms, supporting a range of performance levers, can be effective in improving measured outputs.
2. However, although agents will often change their behaviour in response to sanction/reward mechanisms, the literature suggests that this is not always in desired ways. Sanction and reward mechanisms can have consequences quite different from those intended by their designers. Careful design and implementation, and the tailoring of the mechanism to fit the particular circumstances of an organisation, is crucial in ensuring correct incentives and an effective sanction/reward scheme. Through a survey of the academic economics literature, the UK policy literature, the international policy literature, and the management literature on the private sector experience, this review will discuss issues and requirements for an effective sanction/reward mechanism.
3. We will examine these issues and requirements in the context of the five different stages in a system of sanctions and rewards. The figure below shows these stages. The first stage is the **design** stage, where the type, parameters, value and subject of the sanction/reward mechanism is determined. The next stage is the **measuring** stage, where the data systems which will measure performance against the defined parameters are developed. This is followed by the **application** stage, where the sanction/reward mechanism is applied to the defined parameters. In the **review** stage, the effectiveness of the mechanism is evaluated. Finally, in the implementation stage, the sanction/reward mechanism is rolled out. During this phase, various aspects of the previous four phases may need to be revised if the sanction/reward mechanism is not functioning as expected.

Figure 1: The four stages of a sanction/reward mechanism



¹ By agent, we mean individuals, teams or organisations subject to a sanction/reward mechanism.

4. In this literature review, we will focus mainly on formal sanction/reward mechanisms. Unplanned or informal incentive mechanisms, arising out of performance levers such as legislation, targets, regulation, funding and inspection, can also incentivise the delivery of Public Service Agreement (“**PSA**”) outcomes. However, a comprehensive review of the considerable literature on all performance levers and associated informal incentive mechanisms is beyond the scope of this report, and so our focus is on the literature relevant to formal sanction/reward mechanisms in particular.
5. The remainder of the review is arranged as follows.
 - Section 2 reviews the basic economic theory of incentives, including the theory specific to incentives in the public sector;
 - Sections 3 - 7 examine issues and requirements in each stage of the life cycle of a sanction/reward mechanism;
 - Section 3 highlights the key issues and requirements in the design of sanction/reward mechanisms;
 - Section 4 discusses measurement issues and requirements;
 - Section 5 discusses issues and requirements during the application phase of sanction/reward mechanisms;
 - Section 6 examines the review phase of the sanction/reward mechanism life cycle, including approaches to reviewing the effectiveness of sanction/reward mechanisms from the literature;
 - Section 7 discusses issues and requirements in the implementation phase of a sanction/reward mechanism.

2 Economic theory of incentives

6. There is a well-established economics literature which considers the optimal design of incentives. This is usually applied to contracts between employers and employees, but can be relevant to incentives for teams and organisations as well. Indeed, the economic theory is framed in general terms as a problem whereby one party (the “principal”) needs another party (the “agent”) to act to deliver the first party’s objectives, but the principal cannot fully control (or even observe) the activity of the agent. In the public sector, this relationship may arise between elements in the delivery chain, such as government departments acting as principals and delivery bodies as the agents, or within organisations in the delivery chain, with managers as the principals and employees as agents.
7. Economic theory provides some insights into the optimal design of incentives for principals to apply to agents, under various assumptions. In the first section below we discuss the insights which are common to private and public sectors alike; the second section considers the key factors which are different in the public sector and what those imply for the use and design of incentives in the public sector.

2.1 General Insights

The Basic Case

8. In the simplest case, when the agent’s output is fully observable and the value of his output is a perfect indicator of his effort, the optimal contract is for the principal to select the optimal level of effort and then compensate the agent for the effort exerted. But when the agent’s output also depends on some random factor outside his or her control, the principal cannot infer the agent’s effort precisely. If the agent is risk averse, then the optimal incentive scheme consists of a fixed payment (a risk premium) and a marginal payment proportional to the output. The optimal balance between these incentives and insurance – known as the “intensity” of the scheme-- depends on four factors: the extent to which additional performance results from extra effort, the accuracy with which activities can be measured, the agent’s risk tolerance, and the agent’s responsiveness to incentives. The key insights are that the more risk averse the agent, the more the output depends on factors outside the agent’s control, and the less accurate the measurement of outputs, the less “sharp” the optimal incentive structure.²

Multiple Tasks

9. Holmström and Milgrom (1990, 1991) examine the implications of extending the basic analysis to a world where agents perform multiple tasks. In this case, interaction between the tasks can affect the optimal incentive scheme. If the different tasks are substitutes, higher marginal incentives in one task will skew effort towards that task. If, moreover, some tasks are more accurately measured than others, then the principal may need to weaken the incentives for those tasks to avoid effort being directed too much towards the more measurable tasks, to the detriment of the overall outcome.
10. Marx and MacDonald (2002) consider activities which are substitutes from the perspective of the agent (more time on one means less time on others) but complements from the perspective of the principal (i.e., she wants high effort on all the tasks). They show that if the principal is uncertain about the agent’s preferences between the various tasks, linking rewards to success in individual tasks may be sub-optimal, since the agents will focus “too much” on the tasks which are less costly to them. The optimal incentive structure will in this case be related to overall success across all tasks.

² See Milgrom and Roberts (1990) and Burgess and Ratto (2003).

11. Lazear (1989) reaches a similar conclusion by analysing the case where agents can use two types of actions to win a tournament: effort or sabotage. A big prize will induce a lot of effort, but also a lot of sabotage, so the efficient reward is smaller when sabotage actions are possible. As Gibbons (1999) puts it: “weak incentives may be more efficient than strong but dysfunctional incentives”.

Intertemporal Issues

12. When the basic model is extended to incorporate the possibility that the principal and the agent interact repeatedly, it illuminates some reasons why agents may “game” the incentive system.³ The principal will try to infer the difficulty of the agent’s task from their performance in the initial periods; this gives the agent an incentive to underperform early on, because they fear that otherwise the principal will toughen up the reward system later on. The principal may be able to avoid this to the extent she can credibly commit to not revising the incentive contract in later periods.
13. However, where career concerns dominate for the agent, they may have the incentive to work harder in earlier periods, to signal to the principal their suitability for promotion. In other words, career concerns can create powerful incentives for agents. Holmstrom (1999) shows first that the incentive to exert effort is stronger if there is a lot of uncertainty about talent, but weaker if there is a lot of random factors influencing measured output. But second he also shows that the “career” incentive fades over time, as the agents learn more about the agents’ underlying talent. Thus implicit career incentives may need to be supplemented with explicit financial incentives for older workers.

Subjective Performance Assessment

14. The difficulties of measuring all the aspects of an agent’s performance relative to his overall contribution to the principal’s objectives mean that “relational” incentive contracts—i.e., those enforced by parties’ concerns for their reputations—will often be preferable to explicit contracts enforced by a court. Relational contracts can be based on subjective assessment of the agents’ performance in the round, whereas explicit contracts have to be based on objective, measurable criteria (Gibbons 1999). The parties begin by cooperating and continue to cooperate until one party calculates it would do better by renegeing. Hence explicit (financial) incentives are likely to be more important towards the end of a career.

Employee Selection Effects

15. The above has focused on the extent to which incentive mechanisms can improve productivity by inducing the optimal level of (unobserved) effort. But incentive schemes can also have an impact on productivity via selection effects. For instance, the introduction of piece-rates will attract more hard-working employees and “encourage” slower workers to leave the firm (Lazear, 2000).

2.2 Public Sector Insights

16. The considerations above apply to incentive design in private and public sectors. But the theoretical economics literature has some additional insights which are particularly relevant to a public sector context. These generally indicate reasons why it may be optimal to apply less “sharp” performance incentives than in the private sector.

Multiple principals and conflicting objectives

17. A key feature of the public sector is that it is characterised by multiple stakeholders (principals) – service users, taxpayers, local and national politicians – whose priorities for a given public sector body may not be fully aligned. Indeed, Propper and Wilson

³ See Gibbons (1987) and Tirole and Laffont (1993)

(2003) point out that public agencies exist precisely because there are conflicting goals amongst the stakeholders they represent. There can of course be multiple stakeholders affected by the actions of private sector firms, but in the private sector the owners interests are paramount and generally focussed on the single objective of value maximisation (Dixit, 2002).

18. Due to multiple principals, public sector bodies may be required to achieve a number of different and potentially conflicting objectives. The most general case is that public sector bodies are often expected to simultaneously achieve cost-effectiveness as well as equity objectives. Local authorities, for instance, are required to provide quality social care services to all in need, yet remain within a tight financial envelope⁴. The implication is that, other things equal, the presence of multiple principals means that optimal incentive schemes will be less intense in the public sector, but also harder to design and monitor.

Multiple tasks and overlapping roles

19. Multiple principals with different objectives intensify the difficulty of prioritisation for agents and add to the complexity of designing incentive schemes. Furthermore, defining the different dimensions of a public service can be difficult; for example, a school's objective of providing "a good education" cannot clearly be broken down into subcomponents (Prentice et al. 2007). And it is often not even clear which part of the delivery chain is responsible for each task, as there is often overlap between the roles and responsibilities of front line units, intermediate organisations and departments. For example, the tripartite system in the police has led to confusion regarding the roles of the Police Forces, Police Authorities and the Home Office (HM Treasury, 2004). Hence the scope for an explicit sanction/reward mechanism to skew agents' actions towards those most contributing to the measured outputs, at the expense of the desired outcome. Real world examples of this phenomenon are discussed further in Section 3.2 below.

Measurement problems and team performance

20. The economics literature on incentives has generally highlighted the difficulties of finding objective measures as the basis for explicit incentive schemes. In the public sector these problems are compounded when outcomes may not be apparent for many years (e.g., the effectiveness of a school in terms of maximising a pupil's employment potential over a number of years). At the same time, the basis for decisions made by agents in the public sector is often even less observable to the principal than in the private sector. Prendergast (2002) examines the case of police officers, whose decision whether or not to arrest a suspect depends on his suspicions, which are not easily observable.
21. In cases where neither the effort nor the output is observable, one option is use detailed audits of cases handled by agents (Burgess and Ratto, 2003). But this is expensive and will contain a degree of subjective judgment. However, attempts to summarise complex outputs or outcomes in relatively simple, measurable performance indicators can lead to familiar problems of distorting activities towards those things that are measured. This and related issues are discussed in Section 4.2 below.
22. The relative difficulties of measuring inputs and outputs in a public sector context make the collection of monitoring information more expensive. This, combined with the multiplicity of tasks and roles needed to deliver complex public services (such as

⁴ A further example is that of the US Job Training Partnership Act ("JTPA"), which was the first large-scale, federally funded programme in the US to include performance measures in state and local programmes and link payments to them. This programme mandated the provision of employment and training opportunities to "those who can benefit from, and are most in need of, such opportunities" (Heckman et al., 2002). This created a conflict between equity (those in need) and efficiency (those who benefit).

health care) means that there is often more of a reliance on performance at the aggregate rather than individual level. This in turn may make the use of team-based reward structures more relevant in the public sector; this is discussed further in Section 3.3 below.

Intrinsic motivation and risk aversion

23. Because performance is typically easier to measure in the private sector, and hence “sharper” incentive structures are more prevalent, we would expect more risk averse individuals to opt for the public sector. The objectives of public sector organisations will also tend to attract individuals with intrinsic motivation which is more aligned to those of the principal than may be the case in the private sector, reducing the scale of the underlying principal-agent problem. Indeed, Kreps (1997) argues that setting financial rewards based on performance may actually be counter-productive as it can dilute or undermine the workers intrinsic motivation by signalling that the relationship is financially motivated. The issue of motivation among public sector workers, and the implications for sanction/reward mechanisms, is discussed further in Section 3.4 below.
24. The risk aversion of the principal can also influence the optimal incentive design. Dixit (2002) highlights the case where the principal is highly risk averse – as might be a politician because she cannot easily diversify the risks of bad policy outcomes. In such cases, incentive schemes which relate rewards to achievement of a minimum threshold of performance (i.e., a step function rather than a linear function) will be optimal.

Employee Selection Effects

25. Just as the introduction of explicit performance-related pay can enhance productivity through employee selection in a private sector context, some theoretical papers have shown that emphasizing the “social good” aspects of a job can screen in socially-motivated agents and thereby reduce the importance of designing an effective incentive system to motivate the agent⁵. In the same spirit, Besley and Gahtak (2005) show that matching principal and agent “mission preferences” reduces the need for high powered incentives to leverage agent effort.

2.2.1 Section summary

26. Economic theory indicates some general pointers to the design of incentives, which “principals” can use to align the actions of “agents” with their own interests when the agent’s actions or outputs cannot be easily observed:
- The more risk averse the agent, the more that outputs or outcomes depend on factors outside the agent’s control and the less accurate the measurement of outputs, the less “intense” the optimal incentive scheme, ie the smaller the (explicit) performance-related component of remuneration should be.
 - When career concerns are important to the agent, such as for younger employees, implicit incentives can be more powerful than explicit (financial) incentives, even when there is uncertainty on the part of principals about the agent’s underlying talent. But implicit career incentives become less important over time and will need to be supplemented with explicit financial incentives.
 - When the agent has to perform multiple tasks, it will usually be better for the principal to (a) link the sanction/reward mechanism to performance across all tasks, to avoid the risks of the agent’s actions being skewed towards those tasks which are most easily measured; and (b) use a range of different types of rewards, from cash payments to promotion criteria and job design.

⁵ See Brekke and Nybord (2004) or Delfgaauw and Dur (2007)

- As a result, (a) “objective” performance measures, where the link between performance and reward is explicit, are unlikely to be feasible, and efficient (financial) incentives will typically be small; and (b) subjective assessments are likely to be important in many incentive contracts, even in private sector jobs where objective performance measures are available.

27. Economic theory also suggests some reasons why optimal incentive structures will tend to be different in a public sector context:

- Multiple principals, and hence multiple (and conflicting) objectives, tend to be more common in the public sector; especially when combined with the multiple and complex tasks that are common in the public sector, the optimal incentive structure is less intense or high powered than in the private sector;
- The relative complexity of accurately measuring outputs and outcomes in the public sector complicates the design of optimal incentive structures and increases the risk of inadvertently distorting agents’ behaviour in undesirable ways;
- Differences in intrinsic motivation and attitude to risk matter for the design of incentives, and there are sound theoretical reasons why these will be systematically different in private and public sectors.
- Matching principal and agent “mission preferences” is feasible in the public sector and will reduce the need for high-powered incentives.

3 Design features which lead to effective sanction/reward mechanisms

28. This section, and the following four sections, will examine the issues and requirements for an effective sanction/reward mechanism during the five stages of a sanction/reward scheme: design, measuring, application, review and implementation.
29. There is no “one-size-fits-all” design of sanction/reward mechanisms, and mechanisms need to be carefully tailored to local conditions to be effective. Nevertheless, the literature suggests certain design features which facilitate an effective system of rewards and sanctions and can help to overcome some of the barriers discussed in the previous section. These apply to all types of sanction/reward mechanisms.
30. Our literature review has highlighted the most significant of these:
 - a. The sanction/reward mechanism should be developed together with the agents so as to ensure consensus and appropriateness;
 - b. The sanction/reward mechanism should incentivise the right performance measure;
 - c. The agents must be able to influence the outcome which is being incentivised;
 - d. The sanction/reward mechanism should consider the motivations of agents;
 - e. The sanction/reward mechanism should take into account the team structures of the organisation; and
 - f. The sanctions or rewards must be of sufficient value or consequence to incentivise behaviour.
31. We will discuss each of these aspects in turn. To illustrate each of these aspects, we will also provide evidence through examples drawn from the literature.

3.1 The sanction/reward mechanism should be developed together with the agents and all other stakeholders

32. Involving all agents and stakeholders in the development of the sanction/reward mechanism helps to ensure consensus among stakeholders. This is particularly important to sanction/reward mechanisms applied to the public sector, where a common issue is that of multiple principals with multiple, potentially conflicting, goals. Involving all stakeholders ensures that that trade-offs between different goals can be recognised and quantified and clear overall objectives be set.
33. Furthermore, developing a mechanism together with all agents and stakeholders can help to win their endorsement. In his theory on goal-setting, Locke (1968) argues that if there is employee involvement in the process of goal setting and scheme design, then the employees will be much more committed and motivated towards the achievement of those mutually agreed goals.
34. Giving agents some degree of control of the sanction/reward mechanisms also ensures that targets set through the sanction/reward mechanism are achievable and relevant locally. For example, the successful Carl D. Perkins Vocational and Technical Education Programme in the US gives states responsibility for developing their own performance measures. This programme provides grants to states for career and technical education based on their performance against the agreed

measures. These are related to four required performance indicators⁶ and are negotiated to ensure sufficient rigour. Involving the states in the development of the indicators helps to win their endorsement and ensures that the measures are appropriate for local conditions (GAO, 2006).

35. Similarly, the UK government has recognised the importance of involving agents in the development of incentive mechanisms. In a report on devolving government, giving local organisations the authority and freedom to create local incentives has been identified as key to driving improvements (HM Treasury, 2004).
36. The potential of involving local organisations in the designing of sanction/reward mechanisms is highlighted through local Public Service Agreements. These are voluntary negotiated PSAs between local authorities and government. They are the result of detailed negotiations between the LA and government, and include agreement on the indicators measuring outcomes and on the performance to be rewarded (ODPM, 2003). They have been widely welcomed and are viewed by the Local Government Association as being extremely effective in taking local priorities into account (HM Treasury, 2004). For example, Hammersmith and Fulham's local PSA to tackle the causes of criminal behaviour, disorder and unemployment has led to a number of innovative solutions through freedoms agreed through the local PSA, such as the pooling of crime and disorder budgets (ODPM, 2004).
37. Centralisation appears to be a barrier in the UK to the development of effective performance-related-pay schemes, as it removes the autonomy of public sector managers and their ability to tailor a scheme for local conditions. For example, the lack of progress in the UK in linking pay to performance for most public sector groups (PWC, 2006) may be partly attributed to the inability of local managers to create pay schemes in response to local market forces, caused by the continued centralisation of government (Prentice et al. 2007). This reinforces the results of the OECD (2005) report on performance pay, where it was found that a strong positive relationship existed between delegation in human resource management in an organisation and the effectiveness of a performance pay scheme.
38. Involving agents can also ensure that the burden imposed by sanction/reward mechanisms is manageable. Targets and incentive mechanisms in the UK have tended to be considered in isolation, so that the cumulative burden on any organisation is often not understood (HM Treasury, 2004). This means that the aggregate number of sanction/reward mechanisms and targets may not be manageable or be aligned. Working with local organisations and agents will ensure a better understanding of the total performance burden and a coherent framework.

3.1.1 Section summary

39. Involving all agents and stakeholders in the development of the sanction/reward mechanism helps to ensure consensus and mitigate the problems associated with multiple principals with multiple, potentially conflicting, goals.
40. Giving agents some degree of control over the sanction/reward mechanisms ensures that targets set through the sanction/reward mechanism are achievable and relevant locally, and that the scheme can be tailored to local conditions.

3.2 Incentivise the right performance measure

41. One of the most important features to concentrate on when designing a sanction/reward mechanism is the specification of the outcome measure being incentivised. Wrongly measured performance measures lead to strategic behaviour

⁶ These are academic and technical skill attainment, completion, placement and retention, and non-traditional participation and completion (GAO, 2006).

among agents and possible perverse incentives. The issue was clearly identified in a classic paper by Kerr (1975) titled “On the Folly of Rewarding A, While Hoping for B”.

42. However, measuring outcomes accurately can be problematic, especially in the public sector. Many public sector bodies are “decision makers” (Burgess and Ratto, 2003). The outcomes of these organisations are difficult to measure as their decision often depends on factors which are not easily observable. Wilson (1989) characterises these organisations as “coping organisations”, where both actions and outcomes cannot be accurately measured or observed. Examples include agency managers, tax inspectors, and social care workers. Burgess and Ratto (2003) provide the illustration of police officers, where the decision to arrest a suspect is based on the officer’s opinion of the suspect’s guilt, which is not easily observed by superiors. The authors note that the primary way of controlling behaviour in coping organisations is through costly audits of the cases handled by individuals.
43. Measurability of output is also made more difficult by the multiple tasks prevalent in the public sector. The role of a public sector worker is likely to consist of a number of tasks of different measurability. This makes it likely that a sanction/reward scheme will base incentives on the easily measurable tasks, motivating employees to focus effort only on the tasks being assessed. For example, a school’s provision of a “good education” may lead to exam marks being focussed on as a more easily measured task than other possible components of education, such as the teaching of life skills.
44. However, it is important to note that much of the evidence of “strategic” behaviour, particularly for the UK, is anecdotal. Indeed, one of the few systematic studies of strategic behaviour (in the context of the introduction of wait time targets for A&E cases in NHS hospitals) finds significant improvement in the targeted measure with no evidence of “dysfunctional” consequences (Kelman and Friedman, 2007). This suggests that it is possible to address the risks of gaming or strategic behaviour. There are a number of areas to focus on when choosing output measures, some of which involve trade-offs between different considerations.
 - Ensuring that there are no unmeasured outcomes while still keeping the number of measures in the sanction/reward mechanism manageable;
 - Choosing an intermediate outcome which can feasibly be measured, but which still leads to the desired overall long-term outcome;
 - Minimising strategic behaviour which may improve measured outcomes without improving overall outcomes; and
 - Avoiding selection bias, or “cream skimming”.

Ensuring that there are no unmeasured outcomes while still keeping the number of measures in the sanction/reward mechanism manageable

45. There are generally several aspects to overall outcomes, particularly in the public sector as described above. A successful sanction/reward mechanism must therefore be designed so as to capture all of these; otherwise, certain aspects will be focussed on at the expense of others.
46. There are numerous examples in the literature of altered behaviour to improve performance to the detriment of unmeasured outcomes. A study by Kane and Staiger (2002) examined the assessment of school performance through standardised test grades in some US states in the 1990s. They found that perverse incentives were created through different subjects being assessed in different years⁷. Consequently, teachers used considerably more teaching time on the subject that was tested in the year concerned rather than allocating effort evenly, thus improving test performance without necessarily advancing overall pupil attainment.

⁷ Specifically, Science was tested in Year 4, while Mathematics was assessed in the following year.

47. Similarly, Prendergast (2001) found that unmeasured outcomes in a Los Angeles police force sanction scheme led to a focus on only some aspects of desired outcomes. The scheme was designed to combat violence against criminals during arrests, yet led to a disproportionate fear of disciplinary action in case of complaints from incorrectly arrested citizens. Given the trade-off between arresting the innocent and not arresting the guilty, the scheme resulted in officers being strongly biased towards the latter. Here, the overly punitive system led to the inefficient outcome of letting too many criminals free or not arresting them in the first place. This hindered the achievement of the institution's overall objective of crime-fighting, evidenced by the rapid increase in homicide rates after the introduction of the scheme.
48. It is likely that a range of measures, rather than a single performance measure, will be needed in order to capture all aspects of the overall desired outcome through the sanction/reward scheme. A single performance measure will tend to lead to distortions, and so will not be adequate (Pidd, 2005). For example, the US Child Support Enforcement incentive programme awards incentives based on meeting performance targets in a number of outcome areas such as paternity establishment, current collections, past due collections, and cost-effectiveness. A formula determines the relative weights of each of these outcome areas (GAO 2006). Similarly, the contract designed by Ontario Realty Corporation ("ORC") with a facilities management company sets thirty key performance indicators. These are accumulated and weighted to determine a total performance rating (GAO, 2006).
49. The use of a range of measures is also recommended by Makinson (2000), who puts forward the balanced scorecard approach of combining financial goals with other indicators of performance. This concept was originally developed by Kaplan and Norton (1992) in order to relate key drivers of organisational performance directly to objective measures.
50. At the same time, however, care must be taken to keep the total number of performance targets and measures in a sanction/reward scheme manageable. A report from HM Treasury (2004) notes that at that time the frontline in health faced more than 200 targets, measures and compliance requirements. Aggregating the frontline targets and measures into a single agreed set is crucial (HM Treasury, 2004). For example, Makinson (2000) suggests a maximum of five targets should be used in a performance pay incentive scheme for junior staff and a maximum of eight for senior staff.
51. A further solution is suggested by Kelman and Friedman (2007). Sanction and reward mechanisms should seek to ensure "complementarity across performance dimensions", so that more effort to achieve the measured target may enhance rather than substitute for other desired (but unmeasured) activities.

Measuring an intermediate outcome which leads to the overall outcome

52. In addition to capturing all aspects of the overall outcome, the sanction/reward mechanism must ensure that the measured outcome leads to the overall desired outcome. For example, the JTPA programme measured the outcomes of participants shortly after they completed the programme, as an indication of the long-term benefits of the programme in terms of employability and earning. However, one benefit of training is to encourage further training and schooling, which increases employment and earnings in the long-run but depresses it in the short-run. This meant that the short-term measurement of outcomes was actually found to be perversely related to long-term benefits (Heckman et al., 2002).
53. Focussing on long-term outcomes is more likely to ensure the overall objectives of the sanction/reward scheme are met. However, this conflicts with the need for measures to be reasonably achieved and evaluated within the specified time frame, in order to provide immediate feedback on the programme (GAO, 2006). While focussing on short-term outcomes has obvious benefits, these must be demonstrated to have clear correlation with positive long-term outcomes. Although long-term

studies are required to convincingly determine which short-run measures are strongly related to long-term overall outcomes, an interim solution could be to base the choice of short-term measures on theory or expert opinion, and revise these during the course of the programme.

Minimising strategic behaviour which improves measured outcomes but not overall outcomes

54. Furthermore, the choice of intermediate outcome should take into account any incentives that agents have to “game” the system. This can result in the intermediate outcome being achieved, but not the overall objectives of the programme. There are many examples of agents gaming sanction/reward schemes. For instance, Smith (1995) discusses the massaging of waiting lists in UK healthcare. He finds that targets in the early 1990’s to reduce waiting time for elective surgery led to surgeons delaying placing patients on the waiting list after a consultation, in order to improve their waiting time performance. Jacob (2002) points to a test-based accountability policy introduced in Chicago schools. This led to the improvement of maths and reading test scores, but these improvements were largely driven by improvement in test-specific skills rather than in actual student learning. Petersen et al. (2006) highlight a number of studies of incentive schemes in healthcare whereby the scheme leads to an improvement in documentation rather than a change in the quality of health care delivered to patients. For example, a US programme to increase immunisation rates led to improvements which were mainly due to more up-to-date records of immunisation status rather than in actual numbers of vaccinations given (Petersen et al., 2006)
55. In particular, threshold schemes⁸ can lead to strategic behaviour on the part of agents which improves measured outputs without improving long-run overall outcomes. For example, the JTPA programme allows agency workers to measure performance data⁹ on a participant in the programme at any point up to 90 days after completing the training programme. Because targets are measured on an annual basis, Prentice et al. (2007) note that workers may have incentives to use strategic behaviour to spread good and bad performance over multiple years. This is because the programme bases bonus payments on reaching a specified annual target, with no additional benefits being offered for exceeding targets. Evidence of such strategic behaviour is found in a study of the JTPA programme by Courty and Marschke (2004).

Avoiding selection bias, or “cream skimming”

56. Cream-skimming is a particular example of strategic behaviour caused by incorrectly specified incentive measures. If measured outputs do not take into account differences between users, this can lead to agents using “cream skimming” to select people with high levels of target outputs, rather than those who would benefit most, or receive the most impact from the programme. There are numerous examples of this in the sphere of education, where schools face incentives to tailor their pupil population to improve performance measures based on raw test scores. Propper and Piebalga (2008) point to anecdotal evidence from the UK that schools encourage weaker pupils to sit for vocational exams rather than for GCSE courses.¹⁰

⁸ These are schemes where the sanction/reward is applied once a specified level of performance is achieved.

⁹ Performance data consists of employment status and earnings.

¹⁰ This may make sense in the context of scarce resources, where the effort required to help a weaker pupil through a GCSE course may lead to abler pupils being neglected. There is a clear trade-off in this instance between equity and efficiency. However, this can still be regarded as a perverse incentive, as altering the balance between equity and efficiency in favour of efficiency was not what was intended by the designer of the sanction/reward mechanism, but rather an improved education for all.

Furthermore, schools may attempt to only admit pupils who are likely to improve target outcomes. A number of authors¹¹ have examined ways in which UK schools design their admission policies in order to select particular types of pupils and parents, including the use of pre-admission interviews and complicated admission forms.

57. Examples of cream-skimming, also exist in sectors outside education. The incentive scheme associated with PSA 20 (Increase long term housing supply and affordability) awards a Planning Delivery Grant based on performance criteria related to the speed of processing planning applications. There is evidence that a focus on speed of application processing has led to a bias against more complicated planning applications in some authorities, with these applications being refused or encouraged to withdraw to ensure targets are met (ODPM, 2005). Furthermore, Besley and Ghatak (2003) point to health providers who compete for the healthiest customers, and nursing homes which prefer the relatively healthier. Similarly, Leventis' (1997) study of New York cardiac surgeons pointed to surgeons who refuse to operate on risky patients once their mortality rates come close to a certain threshold.
58. Along with the strategic behaviour discussed in the previous section, threshold sanction/reward schemes can also lead to cream-skimming. Threshold levels, although easier to monitor, can lead to agents' attention being focussed on the borderline. This can lead to dysfunctional behaviour. Wiggins and Tymms (2002) examine the behaviour of UK primary schools concerned with hitting their Key Stage 2 targets, which are reported in terms of a percentage of a school's population attaining the target grade. They discuss how a focus on the achievement of the specific grade level may not benefit pupils at either end of the ability distribution, because schools might target their resources to those who are borderline rather than to all pupils.
59. A solution to the problem of selection bias is the use of adjusted outputs - which take into account characteristics of the individual or area-specific characteristics - or net outcomes, which seek to isolate the impact of the programme. Although these are more difficult to observe, they can reduce the incentives to focus on a subset of potential users. These are discussed further in Section 3.3.
60. Selection issues may also be mitigated by the natural inclination of workers in the public sector to help the most disadvantaged. Heckman et al. (2002) found in their analysis of the JTPA programme that cream-skimming was not much of a problem and that the natural inclinations of employees tended to dominate, except in the presence of very high-intensity incentive systems.

Implications for choice of sanction/reward mechanism

61. The choice of sanction/reward mechanism is heavily dependent on whether measures of output can be correctly specified so as to lead to overall desired outcomes. Outcomes may be easier to specify and measure in some conditions. For example, outputs from parts of the delivery chain which fall into the private sector may be easier to define due to the more clearly defined goal of value maximisation. Furthermore, measurement difficulties vary across the public sector. Pidd (2005) categorises a spectrum of types of public service. On the one extreme are departments where most of the staff are low paid and relatively unskilled, whose work involves limited discretion. Correctly specifying outputs for this group is likely to be far more straightforward than for the other extreme, which is typified by health and education, where workers are well qualified professionals with considerable scope to vary the protocols that define their work.
62. If outputs cannot be easily specified, or if there are multiple objectives or tasks where some are more easily measured than others, it is likely that a high-intensity sanction/reward mechanism, where compensation depends strongly on performance,

¹¹ In particular, see West and Pennell (2000).

will not be appropriate (Prentice et al., 2007). In this case, such schemes are likely to give rise to ineffective incentives or to encourage strategic behaviour such as the misallocation of effort by agents. In this case the theory suggests that lower intensity sanction/reward mechanisms should be used, such as softer operational or reputational incentive scheme or financial incentive schemes where only a limited proportion of compensation is tied to performance (Burgess et al., 2005b).

63. A potential solution in the case where outcomes are not easily measurable may be a subjective judgement appraisal, as this provides a more holistic picture of performance. Gibbons (1999) and Prendergast (1999) both emphasize the extensive use of subjective performance assessments in the private sector, even in occupations where clear objective measures of performance are available. However, a sanction/reward mechanism based on this approach is not only less straightforward to implement, but might also be negatively influenced by factors such as leniency or centrality bias (Prendergast, 1999). Leniency bias refers to a situation where management is reluctant to give negative appraisals to employees, while centrality bias involves managers concentrating appraisal outcomes around a certain (expected) standard. Both of these distort the accurate assessment and reward of employee performance, and thus incentives for effort. Additionally, employees' resources such as time and energy are sacrificed towards encouraging favouritism from their superiors rather than spent on the socially beneficial tasks that the sanction/reward scheme is trying to incentivise. (Prendergast, 1999).
64. A sanction/reward mechanism which combines an objective performance measure supplemented by a subjective performance appraisal may provide a partial solution. The performance pay scheme for civil servants in the Brazilian Tax Collection Authority used a combination of group rewards based on objective measures of the amount of fines collected, together with individual rewards based on subjective assessments (Kahn, Silva and Ziliak, 2001).
65. Furthermore, inputs and process¹² may provide useful alternatives to measuring outcomes in certain limited circumstances. Measures of inputs or processes are generally ineffective in providing information on the effectiveness of a sanction/reward mechanism, as for example the number of staff by itself has little informational value on the quantity and quality of the work performed (Propper and Wilson, 2003). However, where outcomes cannot readily be measured, processes and inputs may provide a useful alternative if there is a clear correlation between an input or process and an outcome. In health care, for example, there may not be agreed consensus on what constitutes a good outcome, but if certain processes are clearly correlated with good outcomes then there may be merit in measuring these processes. Indeed, Marshall et al. (2000) point to evidence that processes can be a more appropriate measure of quality of care in health care than outcomes.

3.2.1 Section summary

66. The specification of the outcome being incentivised in a sanction/reward scheme is very important. Wrongly measured outcomes can lead to strategic behaviour among agents and possible perverse incentives.
67. Outcomes can be difficult to measure, particularly in the public sector where many public sector bodies are "decision makers". Measurement difficulties are exacerbated in the public sector by the multiple tasks that many public sector bodies and employees are required to perform, making it likely that a sanction/reward scheme will base incentives only on the easily measurable tasks.
68. Where there are several aspects to overall outcomes, a successful sanction/reward mechanism should capture all aspects of these objectives; otherwise, certain

¹² These include measures such as the number of staff or the number waiting for treatment.

objectives will be focussed on at the expense of others. It is likely that a range of measures rather than a single performance measure will be needed. At the same time, however, care must be taken to keep the total number of performance targets and measures in a sanction/reward scheme manageable.

69. A sanction/reward mechanism must also ensure that the measured, intermediate, outcome leads to the overall desired outcome. In particular, if short-term outcomes are measured, then these should be demonstrated (empirically or through expert opinion) to have a clear correlation with positive long-term outcomes.
70. The choice of intermediate outcome should also take into account possible gaming by agents. This can result in the intermediate outcome being achieved, but not the overall objectives of the programme.
71. A particular form of gaming is cream-skimming, where agents select people with high levels of target outcomes, rather than those who would benefit most. This occurs where measured outputs do not take into account differences between users. Cream-skimming can be mitigated by adjusting outputs to take into account individual or area-specific characteristics.
72. Where outcomes cannot be easily specified or measured, it is likely that a high intensity sanction/reward mechanism will not be appropriate.
73. A sanction/reward mechanism which combines an objective performance measure supplemented by a subjective performance appraisal may help to overcome difficulties in measuring outcomes. In some limited cases measurement of inputs or processes can be used, where there is a clear correlation between an input or process and a desired outcome.

3.3 The agents should be able to influence the outcome which is being incentivised

74. A sanction/reward mechanism should seek to incentivise outcomes which can be sufficiently influenced by the agent's actions. Without this, the agent may feel that changing his behaviour will have little impact on the outcome and so will have little motivation to strive for the desired levels of performance. This means that the sanction/reward mechanism should strip out aspects which cannot be controlled by the agent from the measured outcome, and seek to measure only the value being added by the agent. By doing so, the incentives for cream-skimming are also reduced.
75. There are a number of examples of incentivised outcomes being adjusted for sources of variation. Mortality data published in the US as part of health care performance reports is adjusted for the clinical mix of patients treated (Marshall et al., 2000). Similarly, in an attempt to account for differences between pupils, the UK school performance tables published net outcomes for the first time in 2002 in the form of measures of the progress of a cohort between two points in time¹³ (Wilson and Piebalga, 2008). By incorporating prior attainment, this measure helps to account for factors beyond the school's control and is therefore theoretically more suitable as a performance measure used for incentivising teaching staff, as they should be more able to influence measurement outcomes.
76. Further solutions may include measuring and incentivising performance compared to historic results, rather than against peers. This adjusts for local conditions. An example is the highly successful New York Compstat programme, which set crime

¹³ The measures were: an indicator of the average value added by the school between Key Stage 2 and Key Stage 3 at 11 and 14 years respectively, and the other between Key Stage 3 and GCSE, at 14 and 16 years respectively. (Wilson, 2003)

reduction targets for each precinct based on historical numbers (HM Treasury, 2004). This allowed the netting off of community-specific factors which could affect crime levels.

77. Sanction or reward mechanisms should also ensure that even those starting from the lowest base are able to influence measured outcomes. For example, threshold schemes which do not take past performance into account may not reward agents who improve the most, because they are starting from the lowest base. Allowing agents to be measured against their own past performance can help to resolve this issue. A further solution is a scheme which rewards a sliding scale of achievement up to a maximum target, as in the UK Quality and Outcomes Framework for General Practitioners. This incentive scheme, introduced in 2004, links 25 percent of GPs' incomes to the quality of their care. There are 146 quality indicators, relating to clinical care, practice organisation and patient experience. Financial rewards increase linearly up to a maximum threshold, which rewards for different indicators weighted by order of clinical value¹⁴ (Gravelle et al., 2007).
78. If the influence of the agent on the outcome cannot be satisfactorily isolated, then a lower-intensity sanction/reward mechanism should be used. High-intensity sanction/reward mechanisms such as performance pay schemes which link a large proportion of pay to performance can impose too much risk on agents if outcomes are beyond their control (Grout and Stevens, 2003). For financial sanction/reward mechanisms in general, risk increases as the amount of money tied to performance increases. Bonuses represent the least risk, as the reward is on top of any basic payment, while a mechanism where the entire financial payment is based on performance represents much higher risk (GAO, 2006).
79. Alternatively, if an agent has limited influence over outcomes, other aspects of performance over which the agent has more control could be incentivised. If certain inputs or intermediate outputs are clearly correlated to outcomes, these could be the focus of a sanction/reward mechanism rather than outcomes.

Directing a sanction/reward mechanism at the correct organisational level

80. The agent's degree of influence over an outcome should also be considered in terms of organisational level, with the correct level of agent within an organisation being incentivised. If front-line workers are those who will deliver the outcome, then incentivising management may be ineffective unless they have the flexibility to filter the incentive scheme down through the organisation in some way. It may be appropriate to have a number of performance measures for different tiers of the organisation. For example, a school league table can be supplemented by an individual sanction/reward mechanism such as award ceremonies for outstanding teachers.¹⁵
81. In particular, influence over an outcome may be at the level of the team rather than at the level of the individual agent. In this case, focussing the sanction/reward mechanism at the team as a whole is likely to be more effective than at the individual level. Directing the sanction/reward mechanism at the team is also able to protect agents from poor outcomes beyond their control through risk pooling.
82. Team work is very prevalent in the public sector and we discuss in detail the interaction between teams and sanction/reward mechanisms in section 3.5

¹⁴ Note that in practice, however, many GPs were at or very close to the max threshold before the scheme started, which limited the benefits of the scheme.

¹⁵ An example of such an individual sanction/reward mechanism is given by GAO (2006): the National Association for State Directors of Career and Technical Education Consortium annually recognises high-performing career and technical administrators

3.3.1 Section summary

83. A sanction/reward mechanism should seek to incentivise outcomes which can be sufficiently influenced by the agent's actions.
84. This can be done by adjusting incentivised outcomes for sources of variation, by measuring and incentivising performance compared to historic results, rather than against peers, or by incentivising aspects of performance over which the agent has more control, such as inputs or processes.
85. Additionally, the sanction/reward mechanism should be directed at the correct organisational level, so that the level of agent who will actually deliver the outcome is incentivised.

3.4 Motivation of agents

86. The motivations of the agents in the public service must be taken into account when designing a sanction/reward scheme. In general, there are two main types of motivation, described by Armstrong (1999). Intrinsic motivation is where the reward is inherent to the activity. Extrinsic motivation is what is done to people to motivate them; namely, rewards or sanctions. For the public sector in particular, Le Grand (1995) distinguishes between workers motivated primarily extrinsically through their own self-interest, "knaves", and those motivated intrinsically by public-spiritedness and altruism, "knights".
87. Despite the gradual shift in belief towards the view that public officials are more likely to be self-interested than public-spirited (Le Grand, 1995), public sector workers may still be motivated by non-financial reasons, such as the welfare of their clients or a professional ethic. There is therefore likely to be a mixture of motivations present in public sector agents.
88. In the case of uncertainty about motivations, it may seem appropriate to assume all public sector agents are knaves and design incentive schemes to appeal to their self-interest. In this way, those agents who are knaves will be motivated by the mechanism, while those who are knights should be unaffected by the scheme and continue to work towards the public good.
89. However, an incentive scheme based on knavish behaviour may negatively impact the motivation of the knights. Sanction/reward mechanisms, particularly those which are financial in nature, may act to weaken intrinsic, knightly motivation and so be counter-productive. Baron and Kreps (1999, quoted in Bénabou and Tirole, 2003) conclude that the benefits of incentive devices can be considerably compromised when the systems undermine workers' intrinsic motivation.
90. There are a number of possible reasons for this. Placing an explicit value on effort may suggest to agents that their employer recognises "no association between output and effort other than a pure, market relationship" (Prentice et al., 2001). Gneezy (2003) suggests that explicit incentives might change the perception of the activity and destroy the intrinsic motivation to perform it. In particular, individuals may work in some areas of the public sector in order to signal altruism, and the introduction of incentives may make signalling more difficult. This might cause agents to stop doing what they were previously doing without reward, or "crowding out"¹⁶ (Bénabou and Tirole, 2005, Seabright, 2004). Bénabou and Tirole (2005) additionally suggest that incentive schemes may weaken intrinsic motivation by undermining agents' confidence in their own abilities or in the value of the rewarded task. This occurs where agents are unsure about the difficulty of the task or in their own abilities.

¹⁶ Crowding out is the finding that people who are willing to perform certain actions for free (or for no explicit reward) will refuse to perform them for a low price. (Seabright, 2004)

Agents might therefore interpret the explicit incentives offered by the principal as an indication that the task is difficult or that the principal does not trust the agent.

91. Furthermore, the reaction to incentives by knights is not necessarily monotonic. Gneezy (2003) demonstrates how a movement from no incentives to small incentives may lead to reduced performance, for the reasons outlined in the previous paragraph. However, when the value of the incentives is then increased, performance improves. This highlights that while small incentives are not necessarily better than no incentives, once the incentive becomes large enough, performance will eventually improve.
92. Gneezy and Rustichini (2003) describe an example of the negative impact of a financial sanction/reward mechanism on “knightly” behaviour. Researchers conducted an incentive experiment at daycare centres in Israel, where fees were charged to parents who arrived late to pick up their children. The introduction of a small penalty fee resulted in more parents arriving late than before. Before the fee was introduced, the only penalty was the social or moral one associated with making the daycare providers wait longer. By the daycentres allowing parents the possibility of paying for late collection, they reduced the perceived element of social disapproval (Seabright, 2004).
93. An additional example is that of blood donation, where providing financial incentives to donate blood might reduce the supply of blood donors. Researchers in Sweden found that providing compensation for donating blood reduced the number of donors by “crowding out” those who had originally donated for reasons of altruism and intrinsic motivation.¹⁷ (Mellström and Johannesson, 2005).
94. Interestingly, intrinsic motivations on the part of agents may in fact reduce the need for explicit incentive schemes. Besley and Ghatak (2005) argue that matching principal and agent “mission preferences reduces the need for high powered incentives to leverage agent effort”. In similar vein, Kelman and Friedman (2007) highlight the possibility that public sector managers can harness the “public service” motivation of their employees to minimise or offset gaming responses to performance targets.¹⁸
95. In addition to intrinsic or self-interested motivations, agents’ attitudes to risk should also be taken into account when designing a sanction/reward mechanism. A number of authors have argued that workers in the public sector tend to be more risk-averse than those in the private sector. For example, Mayntz’s (1985) survey of school leavers found those opting for public service jobs were a self-selective sample that preferred jobs that required them to be competent but risk-averse.

Implications for the choice of sanction/reward scheme

96. In parts of the public sector where intrinsic, “knightly” motivations dominate, the introduction of financial incentives may have a counterproductive effect, particularly those which are of small value. However, incentives may still be helpful in focussing agents on tasks or outputs that would have been neglected if intrinsic motivation alone had been relied upon (Prentice et al., 2007). In this case, low-intensity incentives or softer sanction/reward mechanisms such as operation or reputation incentive schemes may be more effective. In particular, reputational sanction/reward

¹⁷ More precisely, there was a significant crowding out effect for female donors in the field experiment, but this was not observed to any great extent for men.

¹⁸ According to Kelman and Friedman (2007), “A manager... would point out to employees that gaming does nothing to improve real performance and thus runs counter to the service or mission goals of the organisation.” They posit this as one explanation for their finding of no gaming or “dysfunctional” effects of introducing waiting time targets in A&E departments in English hospitals.

mechanisms which recognise good achievement (or point out poor performance) may act to strengthen intrinsic motivation.

97. Where there may be a mixture of motivations, incentive schemes should be designed to allow for the possibility of both types of motivation being present and incentivise individuals of either type. For example, in the example of blood donation given earlier, researchers in Sweden found that offering monetary compensation for donating blood led to a significant crowding out effect. However, by allowing individuals to donate the payment to charity, the crowding out effect was fully counteracted (Mellström and Johannesson, 2005). This incentive scheme attracted those motivated by self-interest – through the monetary compensation – while simultaneously not repulsing those who would have donated out of reasons of intrinsic motivation.
98. In the public service context, Le Grand (1995) provides the example of the General Practitioner (“GP”) fund-holders scheme, where GPs are allowed to keep any surplus on their funds but must use this surplus for any purpose that is beneficial to patients. He argues that this scheme allows for the motivations of both knights and knaves, in that the surplus can be used to benefit both the GPs and the public (for instance, through improving premises or employing more staff, thus easing work pressure and providing a better or more relaxed service.)
99. Other examples include incentivising the introduction of fuel efficiency measures by providing subsidies. Individuals are motivated both by the opportunity to avert climate change and to benefit themselves (for example through reduced fuel bills resulting from improved insulation).
100. Consideration of agents’ attitude to risk also has implications for the choice of sanction/reward scheme. One of the determining factors of the intensity of a sanction/reward scheme is the ability of agents to bear risk. Risk increases as the degree to which an agent’s compensation depends on his performance increases. In particular, financial sanction/reward mechanisms can impose additional risk on workers. In her discussion of performance-related pay, Prendergast (1999) notes that the additional risk imposed by such schemes on workers represents the primary constraint on their use. In environments where agents are risk-averse, low-intensity sanction/reward schemes may be more effective.

3.4.1 Section summary

101. There is likely to be a mixture of intrinsic, public-spirited motivation and self-interested motivation present in public sector agents.
102. Sanction or reward schemes which appeal only to agents’ self-interest may weaken intrinsic motivation where it is present and so be counterproductive.
103. If intrinsic motivations dominate, lower-intensity financial incentives or softer sanction/reward mechanisms such as operational or reputational incentive schemes may be more effective than high-intensity financial incentives which link a large proportion of compensation to performance.
104. Where there may be a mixture of motivations, incentive schemes should be designed to allow for the possibility of both types of motivation being present and incentivise individuals of either type.
105. The sanction/reward mechanism should also take agents’ attitudes towards risk into account. Where agents are very risk averse, low-intensity sanction/reward mechanism are likely to be more effective.

3.5 Teams and incentives

106. Sanction/reward mechanisms aimed at the level of the team have become increasingly popular (Prentice et al., 2007). Makinson (2000) notes that teamwork

reflects better the way in which most public servants work. Furthermore, measures of output are often only available at the level of the team, particularly in the public sector (Makinson, 2000).

107. As described in the previous section, influence over an outcome may be at the level of the team rather than at the level of the individual agent. In this case, fostering competition between teams through team-based sanction/reward mechanisms will tend to be more productive than internal competition within the team (Makinson, 2000). For example, attempting to attribute performance in education to an individual teacher may not be correct. Even if a value-added measure is used, learning is a cumulative process rather than one which can be attributed to an individual teacher (Koretz, 2002), particularly if a pupil is taught simultaneously by a number of teachers. In this case, it may be more appropriate to reward the entire organisation rather than a particular individual. For example, Glendale Union School District in the US provide significant financial incentives to all school staff members, including support staff such as maintenance staff and bus drivers, upon the school's achievement of a number of academic, involvement and satisfaction-related measures. This is based on the premise that all employees have an impact on the school's achievement of these measures, and so should share in the reward. The district has seen academic improvement, which is partly attributed to the camaraderie and collaboration fostered by the scheme among school employees (GAO, 2006).
108. Team-based sanction/reward mechanisms can create the issue of free-riding. Because the contribution of individual agents within a team is not rewarded or sanctioned, agents may feel they exert insufficient influence over the outcome and may be tempted to let others do the work for them (Holmström, 1982)¹⁹. The larger the team, the more serious the free riding problem (Burgess et al, 2005b), as it becomes increasingly easy for an individual to disguise his lack of contribution.
109. Free-riding can be mitigated by peer pressure, if team members can observe each other's effort. This means that the smaller the team, the more likely team members are to be able to monitor each other and reduce free-riding through peer pressure (Kandel and Lazear, 1992). Evidence supporting the relationship between team size and peer monitoring comes from the assessment by Burgess et al. (2004) of the Makinson Incentive Scheme piloted in Jobcentre Plus. The incentive scheme was based on team performance, and paid all members of a team a one percent increase in salary for each of five performance targets met. Meeting all five targets additionally led to a further 2.5 percent bonus. The authors found that team size made a considerable difference to the response rate to incentives. Although overall there was almost no response to incentives, small teams showed ten percent increases in output. This suggests that smaller teams were more successfully able to use peer monitoring to prevent free riding.
110. Free riding can also be mitigated through incentive schemes which combine both team and individual performance. For example, the Kowloon-Canton Railway in Hong Kong introduced a successful incentive scheme, based on individual and team performance targets in four strategic areas. Deductions from monthly pay of between 15 percent to 20 percent were taken from each of the members to create a variable pay pool. If both individual and team targets were met, full contributions were returned. If not, the member would lose part of their contribution depending on the degree to which targets had been met. If targets were exceeded, they would receive pay in addition to their pool contributions (Indian Sixth Central Pay Commission, 2008).

¹⁹ This is because the individual bears the full cost of his own effort, but shares the reward with the rest of the team. He therefore experiences the full gain from reducing his effort in terms of reduced cost, but only some of the resulting reduction in reward.

Teams and motivation

111. The use of team-based measures may also help to overcome the difficulties of retaining intrinsic motivation within a sanction/reward mechanism, discussed in the previous section. A team-based sanction/reward scheme can maintain the collaborative, collective ethos of a public sector organisation without too great a focus on individual financial gains or achievements (Makinson, 2000). Furthermore, if team work is required for some tasks, team rewards foster a spirit of co-operation. Milgrom and Roberts (1991) suggest that rewarding individual performance in this case can reduce team performance by increasing the cost of co-operating.
112. Furthermore, Makinson (2000) notes that it is easier to integrate explicitly financial sanction/reward mechanisms with non-financial incentives in a team framework. For example, having an “office of the month” is less divisive than an “employee of the month”, yet the former can still be rewarded by a shared financial bonus.
113. An example of this is the JTPA scheme, where the financial bonuses received by training centres were not allowed to be used to supplement individual salaries, but only for things such as employing more workers or improving facilities (Prentice et al., 2007). Workers with intrinsic motivation were therefore rewarded indirectly through better conditions and potentially a lessened work load.

3.5.1 Section summary

114. Team level sanction/reward mechanisms have become increasingly popular, as influence over an outcome may be at the level of the team rather than at the level of the individual agent.
115. Team-based sanction/reward mechanisms can create the issue of free-riding. This can be mitigated by peer pressure, if team members can observe each others' effort. Smaller teams are more likely to be able to monitor each other and so reduce free-riding through peer pressure.
116. The use of team-based measures may also help to overcome the difficulties of retaining intrinsic motivation within a sanction/reward mechanism by maintaining the collaborative ethos of a public sector organisation.

3.6 The sanctions or rewards must be of sufficient value to incentivise behaviour

117. Whether the sanction/reward mechanism is financial, operational or reputational, the value or consequences of the sanction/reward needs to be sufficiently high so as to motivate the desired behaviour. Indeed, inappropriately valued incentives appear to be one of the main reasons sanction/reward mechanisms fail (GAO, 2006).
118. Several of the UK performance pay schemes have resulted in only small changes in behaviour as the value of the pay scheme is correspondingly low (Prentice et al., 2007). Examples include performance-related awards of between £50 and £150 paid in the Defence Aviation Repair Agency; a bonus of £400 paid to all staff awarded an “exceeding” performance rating in the Department for Constitutional Affairs; and top ranked performers in the Home Office being awarded 2 percent of salary (Prentice et al., 2007). In contrast, Makinson (2000) suggests that performance pay-related bonuses should represent at least 5 percent of base salary in order to be effective.
119. The JTPA programme gives a further example of the importance of adequately valued incentives. The relative success of the JTPA programme is attributed to the size of the incentives, whereby the programme awarded local training centres as much as a 20 to 30 percent increase in their regular budgets in the event of success in training and placing customers. These awards were particularly valued as training

centres could use them more flexibly than regular budget allocations (Heckman et al., 2002).

- ¹²⁰. As well as rewards being of adequate value, sanctions should correspondingly be stringent enough to incentivise behaviour. Propper et al. (2008) discuss the hospital waiting time target scheme in England, which used the dismissal of key managers of hospitals as a sanction for poor performance. The aggressiveness of these sanctions led to the scheme being dubbed “targets and terror”. In comparison to Scotland, which had a similar focus on reducing hospital waiting times yet without the sanction regime, the authors find that the English sanctions scheme significantly reduced waiting times.²⁰

Trade off between value of sanction/reward and affordability

121. There can be a trade off between a scheme which is of sufficient value to incentivise behaviour and yet is still affordable²¹. For example, in schools in England and Wales, a new upper pay scale for teachers related to performance was introduced in 1998 but abandoned six years later as it became apparent that the scheme was unaffordable. This was due to a flawed scheme design, which initially allowed around a third of teachers in the highest pay scale, but over time permitted far more individuals than expected to enter the highest pay bracket. This eventually made the scheme financially unrealistic, and resulted in the two top scales being taken away (PwC, 2006).
122. Modelling the costs and benefits of an incentive scheme can help to determine the cost-effective level of a sanction/reward. An earlier NAO (2001) report provides the example of HM Customs and Excise, who developed a model to support their objective to reverse the trend of tobacco smuggling. The model draws on a strong evidence base to estimate the current and future levels of smuggling and assess the impact of HM Custom and Excise activities on the level of smuggling. This assists the department in making resource decisions, and allows them to understand the benefits of achieving particular performance measures. The model would therefore also assist the department in determining the appropriate level of an associated sanction/reward associated with particular performance measures.
123. Where no explicit monetisation is possible, understanding how the sanction/reward mechanism can support the strategic priorities of the public sector body may assist in determining the appropriate value of the mechanism (GAO, 2006). This is because the size of the sanction or reward should be in relation to the importance of the priority. This requires that sanction/reward mechanism designers have a clear understanding of the priorities of users being served, so that the incentive scheme accurately reflects the outcomes valued by society (Prentice et al., 2006) In particular, in a multi-dimensional sanction/reward mechanism, the most important priorities should be associated with the largest reward. In the example of the Ontario Realty Corporation (“**ORC**”), the contract designed by the ORC with a facilities management company set thirty key performance indicators. The priorities of the ORC were then used to inform how these indicators were scored and weighted, with greatest weight given to the most important performance objectives (GAO, 2006).

²⁰ The authors note that this does not necessarily point to the overall success of the programme in meeting hospitals’ objectives, as they were not able to test for unintended consequences of the scheme such as the targeting of less needy patients or the reduction in other activities (Propper et al, 2008).

²¹ We note here, however, that it is possible to avoid this trade-off. For example, an sanction/reward mechanism that uses financial penalties in conjunction with financial rewards may avoid this trade off by using penalties paid by some agents to finance the rewards to others. However, considerations of political acceptability mean that reward-only schemes are used more frequently than combined penalty and reward schemes.

3.6.1 Section summary

124. Whether the sanction/reward mechanism is financial, operational or reputational, the value of the sanction or reward needs to be sufficiently high to motivate the desired behaviour. Inappropriately valued incentives may lead to the failure of the sanction/reward mechanism.
125. At the same time, schemes still have to be affordable. The cost effective level of a sanction or reward can be determined by monetising the costs and benefits of a scheme. Understanding the priorities of the public sector body can also assist in achieving cost-effectiveness.

4 Measurement issues and requirements

126. The second stage in a system of sanctions and rewards is to measure performance against the parameters defined in the sanction/reward mechanism. Ensuring that the incentivised outcome can be clearly measured and that systems are in place to collect good quality performance data on the outcome is key to the success of the sanction/reward mechanism.

4.1 Systems are in place to collect good quality performance data

Development of data systems

127. Once the performance measures to be incentivised have been chosen, it is necessary to ensure adequate systems are in place to collect data on the chosen measures. Good quality, reliable data is crucial to the success of a sanction/reward mechanism. Indeed, the GAO (2006) note that ensuring good quality measures and data precedes attaching these to any incentive scheme. They point to the Office of Management and Budgets, the US body responsible for improving the performance of federal programmes, who have to date focussed on improving agency ability to develop high-quality, results-based performance measures as a prelude to introducing incentive schemes and accountability provisions. Similarly, the UK government has identified accurate and timely publication of performance data as essential to strengthening incentives through sanction/reward mechanisms (HM Treasury, 2004).
128. Developing systems to collect good quality performance data involves considerable effort. Data controls and quality assurance arrangements need to be in place to ensure the required level of data, common data definitions need to be developed to allow comparable data to be collected across an organisation, and the data collected needs to be clear and transparent enough to allow reliable decisions on whether to apply the sanction/reward to be made. Additionally, the provision of data also needs to be frequent enough to inform the use of sanctions and rewards on a timely basis, so as to motivate the delivery chain. In particular, data needs to be collected on a regular basis so that action can be taken before issues develop into significant problems (HM Treasury, 2004).
129. Difficulties in developing good data systems are exacerbated by the outcome-focussed nature of many sanction/reward schemes. Internal data collection systems which concentrate on inputs, processes and outputs may not allow the impact of a department or organisation's activities in terms of outcomes to be understood. In this case, information may need to be derived from external sources or collected, for example through surveys (NAO, 2001). This may increase challenges associated with data assurance and transparency.
130. Furthermore, data standards should be kept constant over time so that historic comparisons can be made (HM Treasury, 2004) and performance over time assessed.
131. For example, the critical success factor in New York's Compstat programme is arguably the collection of timely, robust precinct level data (HM Treasury, 2004). This allows crime patterns to be analysed and appropriate, robust crime reduction plans to be developed. The scheme was introduced in 1994, and by 2003, a 67 percent reduction in recorded crime had been achieved. While the impact of the programme cannot be isolated, and other factors such as an increase in police numbers and economic growth certainly contributed to the success of the scheme, Compstat is seen to have played a key role in this reduction (HM Treasury, 2004).

Development of supporting structures

132. Data systems cannot be developed in isolation, but should be supported by appropriate structures. Collected data needs to be audited, interpreted and then presented in a concise and clear way so that all stakeholders are able to understand performance.
133. In the earlier example of the Compstat programme (HM Treasury, 2004), a twenty-five person team is responsible for central analysis and quality assurance of the data collected from each local precinct. This then paves the way for the weekly accountability meeting. Indeed, Garicano and Heaton (2007) demonstrate that the supporting structure around the data collection system was crucial for the success of the Compstat programme. In other areas, the adoption of new IT systems by US police forces had no crime reduction effect, unless it was accompanied by new incentives schemes as in the Compstat case.

4.1.1 Section summary

134. Good quality, reliable data is crucial to the success of a sanction/reward mechanism.
135. Adequate systems should be in place to collect data on the chosen measures, so that the collected data is robust, clearly defined, transparent and frequent.

5 Application issues and requirements

136. The key conclusion from the literature is that the sanction/reward mechanism must be consistently applied.

5.1 There must be consistency in the application of the reward/sanction mechanism to create credibility

137. In order to credibly motivate agents, a sanction/reward scheme must be consistently implemented. If rewards or sanctions are not awarded as expected, agents will learn that their additional efforts are not worth the cost or risk. GAO (2006) notes that if rewards are paid indiscriminately or sanctions are not levied as expected, agents can learn that no additional effort is required to benefit.

138. Similarly, consistency in application is one of the principles highlighted by Macrory (2006) in his report on effective regulatory sanctions. He notes that failure by regulators to follow up low-level enforcement actions such as warning letters with the threatened sanction means that they will not be taken seriously and credibly by firms.

139. GAO (2006) provides the example of the Federal Transit Administration (“FTA”) in the United States. Reviews showed the FTA’s oversight was inconsistent, and that it seldom used the sanctions at its disposal to deal with grantees’ non-compliance. This meant that long-standing weaknesses among grantees were seldom corrected and that federal dollars were placed at risk.

140. Consistency is particularly important in the context of the multiple principals prevalent in the public sector. Given the many, and potentially conflicting, objectives held by different principals, it is possible that agents may play off one principal against another (Marsden et al., 2002, in Brown et al, 2002). For example, Ong (2006) gives the example of Rural Credit Cooperatives (“RCCs”), which are microcredit institutions in rural China. Ong argues that RCC officers are not accountable to member households because of multiple principles with conflicting objectives. For example, member households are unable to enforce appropriate sanctions against poorly-performing officers as this is generally blocked by unions and local party secretaries who influence personnel appointment and dismissal. Issues with multiple principles can be mitigated by ensuring that levels of performance which trigger a sanction or reward are clearly defined and understood, and are consistently applied without exception.

5.1.1 Section summary

141. Phased implementation allows the design of a sanction/reward mechanism to be tested. It also allows data systems to be developed and those affected to develop the necessary knowledge about the mechanism and the ability to implement it.

142. There is a strong case for initially implementing a sanction/reward mechanism on a pilot basis.

143. A sanction/reward scheme must be consistently implemented in order to credibly motivate agents.

6 Issues and requirements during the review phase

144. The key conclusions from the literature are:

- a. The sanction/reward mechanism should be regularly reassessed; and
- b. Assessment of effectiveness is difficult but the approach taken needs to be fit-for-purpose.

6.1 The sanction/reward mechanism must be regularly reassessed

145. During and beyond the implementation phase, scope must be built in to revise and update the sanction/reward mechanism. Formal assessment and revision points should be built in to allow for this. For example, in the Carl D. Perkins Vocational and Technical Education Programme (GAO, 2006), states can periodically revise their target and measures during annual negotiations of their state plans. Officials have attributed the programme's success to this ability to revise and renegotiate targets and measures.

146. The effectiveness of a sanction/reward programme can often only be assessed after it has been in place for a substantial period of time. In the case of the JTPA, for example, the long-term consequences of the programme could only be assessed after the programme had been in operation for several years.

147. Furthermore, changing needs over time may require the mechanism to be revised. Technology may be introduced that alters performance expectations (GAO, 2006) or the priorities of an organisation may change. Additionally, certain performance standards may become the norm and may no longer need to be incentivised. Conversely, some aggressive targets may not be sustainable in the long-term and may need to be scaled down.

6.2 Assessment of effectiveness should be fit-for-purpose

148. Despite the increasing interest in sanction/reward schemes, there is little empirical evidence as to whether these mechanisms actually achieve the goals of the programme or department (Burgess et al., 2002, Prentice et al., 2007). As an illustration, Prentice et al. (2007) found only seven empirical studies on UK public sector financial incentive programmes after 1999.

149. This is partly to do with the difficulty of robustly assessing a sanction/reward mechanism's effectiveness. Fully understanding the effectiveness of sanction/reward mechanisms requires experimentation and the formal assessment of government policy through randomised trials. Sanction or reward mechanisms must be introduced on a controlled trial basis in order to develop a counterfactual of what would have happened in the absence of the sanction/reward mechanism. This allows the impact of the mechanism to be understood. However, sanction/reward mechanisms tend to be introduced on a wide-scale basis as a result of a policy change, rather than on a controlled trial basis (Burgess et al., 2002). Furthermore, their introduction is often alongside other policy changes, making it difficult to isolate the impact of a particular mechanism. For example, UK school league tables were introduced across all schools, and as part of the general reform of schooling provision (Propper and Wilson, 2003). This means that the effectiveness of sanction/reward mechanisms can often only be determined in terms of their impact on measured outputs, rather than through isolating their specific impact on outcomes.

150. In general, there are a range of approaches to assessing effectiveness; these are on a spectrum from purely qualitative approaches to data driven econometric analyses. As described above, the most accurate assessment of the effectiveness of a

sanction/reward mechanism is one which involves the estimation of a counterfactual through randomised trials or experiments. This approach is technically difficult and requires much data, which often needs to be collected over a prolonged period of time. In some cases, “natural experiments” where a control group can be identified can provide a good proxy for randomised trials.

151. In other cases, where there is reliable data on the output or outcome of interest and other factors which impact on the output/outcome, econometric or statistical analysis can be used to estimate the counterfactual outcome of the sanction/reward mechanism through controlling for other factors influencing the output or outcome. The robustness of these kinds of studies depends critically on the quality and quantity of data; but even when the data is good, causation is much harder to determine than correlation.
152. Data on outcomes and on potential explanatory variables other than the sanction/reward mechanism is often difficult, expensive or slow to obtain. In these cases, the assessment of effectiveness is limited to examining the path of output measures over time and hypothesising links with the introduction of (or refinements to) sanction/reward mechanisms. Net output measures, which attempt to measure the value added by the programme, are preferable to gross or raw output measures.
153. Finally, in the absence of quantitative outcome measures, the effectiveness of a sanction/reward mechanism may be assessed on a purely qualitative basis. This may be through canvassing the opinions of participants in the sanction/reward scheme.

Assessments of effectiveness which involve counterfactuals

154. An effectiveness assessment involving a counterfactual is the most robust approach to assessing effectiveness, but is very expensive and time-consuming. A counterfactual outcome is often not available as sanction/reward mechanisms are generally introduced on a wide-scale basis, as mentioned earlier. Furthermore, estimating counterfactual outcomes can take a long time, often years (Heckman et al., 2002). This is due to the time associated with collecting, cleaning and analysing comparison group data. Moreover, methods of counterfactual estimation are also technically demanding. An additional concern is that there may be moral objections to introducing a sanction/reward mechanism on only a trial basis if it is believed to substantially improve performance.
155. A number of assessments involving counterfactuals have been conducted on private sector incentive schemes, which found performance gains resulting from the introduction of the schemes (see studies by Paarsch and Shearer (2000), Shearer (2003) and Bandiera, Barnakay and Rasul (2005), quoted in Prentice et al. (2007)). A good illustration comes from Lazear (2000), who examined the introduction of a piece-rate payment scheme²² in a windscreen installation firm in the United States. As this scheme was introduced gradually over time, Lazear was able to control for outside influences on worker productivity. He found that the productivity per worker improved by 44 percent, partly due to the incentive effect and partly due to a selection effect, whereby more able employees were attracted to and retained by the firm. The scheme controlled for multiple dimensions of output by including quality standards: if a windscreen that a worker installed broke, this had to be replaced without wages being received.
156. The US Job Training Partnership Act (JTPA) provides one of the few instances where a public sector sanction/reward measure has been assessed through the estimation of a counterfactual. Introduced in 1982, it was the first large-scale, federally funded programme to include performance measures in state and local programmes and link payments to them. This programme provided employment and training opportunities to low income users through local training centres. It focussed on outcomes such as job placements and trainee earnings, and based budgetary

²² Workers' pay was linked to the number of windscreens they installed.

incentives for managers on these outcomes. As part of the programme, data was collected on the long-term earnings and labour market participation of enrolees to the programme, as well as on control groups. This enabled an assessment of the effectiveness of the programme.²³

157. A number of studies have examined the effectiveness of the JTPA programme. In particular, Heckman, Heinrich and Smith (2002) find the programme was effective in as much as managers responded to the incentive scheme, and the output-based measures of performance improved. However, as previously examined in section 3.2, this did not necessarily mean that the real goals of the organisation were achieved. The short-term outputs of placements and earnings which were measured were found not to predict the long run placement and earnings impacts. In other words, the short-term output measures were often negatively related to the longer-term participant earnings and employment gains that are the programme's real aims²⁴.
158. A further example of an assessment involving a counterfactual is that of the Makinson Incentive Scheme in HM Customs and Excise. The Makinson Incentive Scheme was trialled in a number of teams in HM Customs and Excise, whereby bonuses were provided for greater output of some of the activities of these teams. Burgess et al. (2005b) estimate the impact of the scheme on two particular teams by using a control group, where the change in performance in the targeted group was compared against the change in performance for the control group. This "difference in differences" (Burgess et al., 2005b) approach allows definition of the counterfactual of what would have happened without the scheme. It also implicitly controls for team and individual-specific characteristics which might additionally affect performance, such as difference in staffing and local market conditions. The authors found the scheme to be effective, as both incentivised teams increased output in comparison to the control group.
159. The evaluation by Kahn, Silva and Ziliak (2001) of the performance pay scheme introduced for civil servants in the Brazilian Tax Collection Authority is an example of the use of econometric analysis to isolate the impact of a sanction/reward mechanism. This scheme rewarded tax officials for their performance in collecting overdue federal taxes. The sanction/reward mechanism paid bonuses which increased with the amount of fines collected, on both a team and individual basis. The group award was based on relative efficiency with respect to other tax collection groups, and paid officials in a group up to 30 percent of the additional fines collected. The individual bonuses were paid out of the remaining 70 percent of additional fines collected, and were based on subjective individual monthly evaluation of performance. Prentice et al. (2007) comment that this scheme is relatively unique due to the high-intensity of the incentives. Indeed, bonuses frequently triple the basis salary of officials.
160. Kahn, Silva and Ziliak (2001) found that this scheme was very effective, leading to a 75 percent increase in fines collected per inspection. This conclusion was reached through an econometric assessment of the scheme, conducted on a panel set of data from 1987 to 1992, three years before and after the introduction of the scheme. The authors were able to isolate the impact of the incentive scheme from a concurrent income tax reform through an additional data set, which looked at collection activities for a range of specific taxes including income tax.
161. The assessment of the English hospital waiting time target scheme provides an example of counterfactual estimation using a "natural experiment". Propper et al. (2008) examine the effectiveness of the hospital waiting time target scheme in England, in the "targets and terror" scheme discussed previously in Section 3.6. This scheme used the dismissal of key managers of hospitals as a sanction for poor

²³ See Heckman et al. (2002) for a detailed description of the data used.

²⁴ See Section 3.2 for a more detailed discussion of this.

performance in measured waiting time for elective hospital procedures. They compare reduction in waiting times in England to those in Scotland, which had a similar focus on reducing hospital waiting times yet did not have a sanction regime. In a difference-in-differences estimator, they compared the reduction in waiting times while controlling for total healthcare expenditure per capital (as increased health spending could account for part of the difference in reduction rates), total NHS staff per capita and differences in population health through standardised mortality ratios. This allowed them to isolate the impact of the sanction scheme, with the conclusion that the English sanctions scheme significantly reduced waiting times.

Other assessments of effectiveness

162. The difficulty and expense of robustly estimating a counterfactual means that the effectiveness of a sanction/reward mechanism is generally estimated in terms of short-term measured outputs based on administrative data.
163. At the simplest level, gross outputs can be used to determine effectiveness. Identified by Barnow (1992), gross outputs are a measure of raw outputs of a programme and usually have the advantage of being easy to collect and understand. Examples include the number of people receiving social care services or the number of individuals given hip replacements. The UK school league tables are also an example of gross outputs, as these are based on pupils' exam results at a number of key learning stages.
164. Gross measures can be used to demonstrate evidence of effectiveness. For example, Marshall et al. (2000) discuss studies of the New York Cardiac Surgery Reporting system, and find a significant decline in mortality after mortality rate data was published.
165. However, gross outputs may not be the most accurate measure of effectiveness. They can be manipulated by agents. Furthermore, they do not necessarily measure the contribution of the programme to the output.
166. Net outcome measures can be more useful in measuring the contribution of a programme to the outcome. These measures seek to estimate the value that is added by the programme. As with a full counterfactual assessment, these outcomes are generally difficult to measure, in that it is again hard to define what would have been the counterfactual in the absence of the programme. However, net measures are available in some sanction/reward mechanisms. For example, indicators used in education can be designed so as to account for differences between pupils by including measures of the pupil's prior performance, as is the case in the net outcome measure incorporated in the UK school performance tables. As described in Section 3.3, this measures the progress of a cohort between two points in time and thereby incorporates prior attainment.
167. In the absence of net measures of effectiveness, gross output measures can be adjusted to take into account characteristics of the individual or area-specific characteristics. In the US, performance measures for medical outcomes are risk-adjusted by being adjusted for the health of the individual treated (Propper and Piebalga, 2008). The JTPA programme used a regression model which allowed states to adjust performance targets for local differences in economic conditions and participant characteristics (Heckman et al., 2002).
168. Finally, in the absence of quantitative outcome measures, the effectiveness of a sanction/reward mechanism may be assessed on a purely qualitative basis. This may be through canvassing the opinions of participants in the sanction/reward scheme.

6.2.1 Section summary

169. Frequent reviews of the sanction/reward mechanism are required in order to assess the effectiveness of the mechanism, amend the mechanism for design flaws and take into account changing needs over time.

170. The most accurate assessment of the effectiveness of a sanction/reward mechanism is an assessment which estimates a counterfactual of what would have happened in the absence of the mechanism.
171. Assessments involving counterfactuals are technically difficult, require much data, and can take a long time. Methodologies include the use of control groups and the econometric analysis of comprehensive data sets.
172. The difficulty and expense of conducting a full counterfactual assessment means that the effectiveness of a sanction/reward mechanism is generally estimated in terms of short-term measured outputs based on administrative data.
173. At the simplest level, gross outputs are a measure of raw outputs of a programme and have the advantage of being easy to collect and understand. If available, net outcome measures can be more useful in measuring the contribution of a programme to the outcome. These measures seek to estimate the value that is added by the programme.
174. In the absence of other data, the effectiveness of a sanction/reward mechanism can be assessed on a qualitative basis, for example through canvassing the opinions of participants in the scheme.

7 Implementation

175. The final stage in the life cycle of a sanction/reward mechanism is the implementation period. Key to this stage is allowing the mechanism to be phased in over a period of time. This allows a planned and incremental change which is important to the scheme's success. Indeed, Armstrong and Murlis (1998) argue that a hasty implementation is often responsible for a scheme's failure, rather than any faults with the scheme itself.
176. A phased implementation allows organisations to ensure that the scheme is effectively designed. There are often unintended consequences associated with a sanction/reward mechanism, meaning that allowing sufficient time to test the mechanism is vital before tying it to rewards and sanctions. A phased implementation also allows time for data collection systems to be developed, and confidence in the quality and reliability of the collected data to be built up. In particular, there is a strong case for a sanction/reward mechanism to be initially implemented on a trial or pilot basis (Armstrong and Brown, 2001). This allows the impact of the mechanism to be accurately assessed, as it can be compared to a counterfactual.
177. Furthermore, phasing in a sanction/reward mechanism over time allows agents the opportunity to develop the necessary knowledge about the mechanism and the ability to implement it. This is critical to the success of the scheme. For example, the US Air Force found that the most important factor in implementing its successful performance-based contracting programme was employee training focussing on how the performance mechanism worked (GAO, 2006).

Appendix: Bibliography

1. Armstrong, M. & Murlis, H. (1998). *Reward Management*, 4th ed., London: Kogan Page.
2. Armstrong, M. (1999). *Employee Reward*, 2nd ed., Institute of Personnel and Development
3. Armstrong, M. & Brown, D. (2001). "New Dimension in Pay Management". CIPD.
4. Bandiera, O.; Barankay, I. & Rasul, I. (2005). "Social Preferences and the response to incentives: Evidence from Personnel Data". *Quarterly Journal of Economics*, 120(3), 917-962.
5. Barnow, B.S. (1992). "The Effect of Performance Standards on State and Local Programs", in Manski, C & Garfinkel, L. (eds) *Evaluating Welfare and Training Programme*, Cambridge, MA: Harvard University Press.
6. Baron, J. and Kreps, D. (1999). *Strategic Human Resources*. New York: John Wiley.
7. Bénabou, R., & Tirole, J. (2003). "Intrinsic and Extrinsic Motivation". *Review of Economic Studies*, 70(3), 489-520.
8. Bénabou, R. and J. Tirole (2005). "Incentives and Prosocial Behavior." *American Economic Review* 96(5): 1652-1678
9. Besley, T., & Ghatak, M. (2003). "Incentives, Choice and Accountability in the Provision of Public Services". Working Paper No. WP03/08, Institute of Fiscal Studies.
10. Besley, T. and M. Ghatak (2005). "Competition and incentives with motivated agents." *American Economic Review*, 95(3), 616-636
11. Bradley, S., Crouchley, R., Millington, J. & Taylor, J. (2000). "Testing for Quasi-Market Forces in Secondary Education". *Oxford Bulletin of Economics and Statistics*, 62(3), 357-390
12. Bradley, S., & Taylor, J. (2007). "Diversity, choice and the quasi market: An empirical analysis of secondary education policy in England". Department of Economics, Lancaster University.
13. Brekke, K. A. and K. Nyborg (2004). "Moral hazard and moral motivation: Corporate social responsibility as labor market screening." 13th annual EAERE conference in Budapest, June: 25-28
14. Burgess, S., Metcalfe, P. (1999) "The use of incentive schemes in the Public and Private Sectors: Evidence from British Establishments". CMPO Working Paper 9(015).

15. Burgess, A., Propper, C. & Wilson, D. (2002). "Does performance monitoring work? A Review of the evidence from the UK Public sector, excluding healthcare". CMPO Working Paper No. 02/49, CMPO, University of Bristol.
16. Burgess, S., & Ratto, M. (2003). "The role of incentives in the Public Sector: Issues and Evidence". CMPO Working Paper 03/071, CMPO, University of Bristol.
17. Burgess, S., Propper, C., Slater, H. & Wilson, D. (2005a). "Who wins and who loses from school accountability? The distribution of educational gains in English secondary schools". CMPO Working Paper 05/128, CMPO, University of Bristol.
18. Burgess, S., Propper, C., Ratto, M., Scholder, S. & Tominey, E. (2005b). "Evaluation of the Introduction of the Makinson Incentive Scheme in HM Customs and Excise". CMPO, University of Bristol.
19. Courty, P., Marschke, J. (1997) "An empirical investigation of gaming responses to explicit performance incentives". Department of Economics, University at Albany.
20. Roderick, M., Jacob, B. & Bryk, A. (2002). "The Impact of High-Stakes Testing in Chicago on Student Achievement in the Promotional Gate Grades." *Educational Evaluation and Policy Analysis*. 24(4), 333-357.
21. Gorard, S. (2002). "Let's keep it simple: the multilevel model debate". *Research Intelligence*, 81, 24.
22. Cabinet Office (2006), "The UK Government's Approach to Public Service Reform – A Discussion Paper"
23. Cullen, J.B., & Reback, R. (2006). "Tinkering towards accolades: school gaming under a performance accountability system". NBER Working Paper No. 12286, NBER, Cambridge MA.
24. Dixit, A. (2002). "Incentives and Organizations in the Public Sector: An Interpretive Review". *Journal of Human Resources*, 37(4), 696-727.
25. Deere, D., & Strayer, W. (2001). "Putting schools to the test: school accountability, incentives and behaviour". Department of Economics, Texas A&M University
26. Delfgaauw, J. and R. Dur (2007). "Signalling and screening of workers' motivation." *Journal of Economic Behavior & Organization* 62(4): 605-624
27. Fairbrother, G., Friedman, S., Hanson, K.L., & Butts, G.C. (1997). "Effect of the vaccines for children program on inner-city neighborhood physicians". *Archives of pediatrics & adolescent medicine*, 151(12), 1229-35.
28. Figlio, D.N., & Getzler, L.S. (2002). "Accountability, ability and disability: gaming the system". NBER Working Paper No. 9307, NBER, Cambridge MA.
29. Fitz-Gibbon, C.T. (1998) "OFSTED: Time to go?". *Managing Schools Today*, 7 (6), 22-25.
30. Fitz-Gibbon, C.T. & Stephenson-Forster, N.J. (1999). "Is OFSTED helpful? An evaluation using social science criteria". In Cullingford, C. (Ed) *An Inspector Calls: OFSTED and its effect on school standards*. London: Kogan Page.

31. Fitz-Gibbon, C.T. (1999). "Education: High Potential Not Yet Realized". *Public Money & Management: Integrating Theory and Practice in Public Management*, 19(1), 33-40.
32. Fitz-Gibbon, C.T. and Tymms, P. (2002). "Technical and Ethical Issues in Indicator Systems: Doing things right and doing wrong things". In *Education Policy Analysis Archives*, 10 (6).
33. Gibbons, R. (1987) "Piece-Rate Incentive Schemes", *Journal of Labor Economics*, 5(4), 413-429.
34. Gibbons, R. (1999) "Incentives in Organizations ", *Journal of Economic Perspectives*, 12(4), 115-132.
35. Government Accountability Office (GAO). (2006). "Grants Management: Enhancing Performance Accountability Provisions Could Lead to Better Results". *Report to the Chairman, Subcommittee on Government Management, Finance and Accountability, Committee on Government Reform, House of Representatives*, GAO-06-1046.
36. Gneezy, U. (2003). "The W effect of incentives". The University of Chicago Graduate School of Business. Available at <http://www.cramton.umd.edu/workshop/papers/gneezy-W-effect-of-incentives.pdf>.
37. Gneezy, U. and A. Rustichini. (2000). "A fine is a price," *Journal of Legal Studies*, 29, 1-17.
38. Gravelle, H., Sutton, M., & Ma, A. (2007). "Doctor behaviour under a pay for performance contract: Evidence from the quality and outcomes framework". Research Paper No. 28, Centre for Health Economics, University of York.
39. Grout, P.A., & Stevens, M. (2003). "Financing and managing public services: an assessment". CMPO Working Paper No. 03/076, CMPO, University of Bristol.
40. HM Treasury. (2004a). "Delivering better public services: refining targets and performance management". Available at: http://www.hm-treasury.gov.uk/media/3/2/devolving_decision1_409.pdf.
41. HM Treasury. (2004b). "Devolving Decision Making: a consultation on regional funding allocations". Available at: http://www.hm-treasury.gov.uk/pre_budget_report/prebud_pbr04/assoc_docs/prebud_pbr04_adregional.cfm.
42. HM Treasury. (2004c). "SR 2004: public service agreements 2005/2008". Available at http://www.hm-treasury.gov.uk/spending_review/spend_sr04/psa/spend_sr04_psaindex.cfm
43. HM Treasury. (2007). "Review of sub-national economic development and regeneration". Available at http://www.hm-treasury.gov.uk/media/9/5/subnational_econ_review170707.pdf.
44. Hansen, D. (1997). "Worker Performance and Group Incentives: A Case Study". *Industrial and Labour Relations Review*, 51, 37-49.

45. Heckman, J., Heinrich, C., & Smith, J. (2002). "The Performance of Performance Standards". *Journal of Human Resources*, 37(4), 778-811
46. Holmström, B. (1982). "Moral hazard in teams", *Bell Journal of Economics*, 13, 324-340.
47. Holmström, B. Milgrom, P. (1990). "Regulating trade among agents", *Journal of Institutional and Theoretical Economics*, vol. 146(1), 85-105.
48. Holmström, B. Milgrom, P. (1991). "Multi-task principal-agent analyses: Linear contracts, asset ownership and job design", *Journal of Law, Economics and Organisation*, 7, 24-52.
49. House of Commons. (2003). "Secondary Education: Diversity of Provision". Education and Skills Committee, HC94, May.
50. Garicano, L. and Heaton, P. (2007) "Information Technology, Organization, and Productivity in the Public Sector: Evidence from Police Departments". Available at http://faculty.chicagogsb.edu/luis.garicano/research/index_files/ITPaper25.pdf
51. Indian Sixth Central Pay Commission. (2008). "Formulating the concept, Principles, and Parameters for Performance-Related Incentives (PRI) in Government". Indian Institute of Management, January 2008.
52. Jesson, D. (2002). "Response to the article by Schagen and Goldstein", *Research Intelligence*, 80, 16.
53. Jesson, D. & Crossley, D. (2004). "Educational Outcomes and Value Added by Specialist Schools". Specialist Schools Trust. Available at <http://www.specialistschoolstrust.org.uk>, last accessed May 2006.
54. Kahn, C., Silva, E., & Ziliak, J. (2001). "Performance-Based Wages in Tax Collection: The Brazilian Tax Collection Reform and Its Effects". *Economic Journal*, 468, 188-205.
55. Kane, T., & Staiger, D. (2002). "The Promise and Pitfalls of Using Imprecise School Accountability Measures". *Journal of Economic Perspectives*, 16(4), 91-114.
56. Kandel, E. Lazear, E. (1992) "Peer Pressure and Partnerships", *Journal of Political Economy*, 100 (4), 801-817.
57. Kaplan, R.S., & Norton, D.P. (1992). "The balanced scorecard – measures that drive performance". *Harvard Business Review*, Jan-Feb, 71-79.
58. Kelman S, and Friedman J. N. (2007) "Performance Improvement and Performance Dysfunction; An Empirical Examination of Impacts of the Emergency Room Wait-Time Target in the English NHS". Harvard University JFK School of Government Faculty Research Working Paper RWP07-034
59. Kerr, S. (1975). "On the Folly of Rewarding A, While Hoping for B." *Academy of Management Journal*, 18(4): 769-783.
60. Laffont, J.-J. Tirole, J. (1993). "A Theory of Incentives in Procurement and Regulation". Cambridge, MA: MIT Press.

61. Lazear, E. P. (1989). "Pay Equality and Industrial Politics". *Journal of Political Economy*, 97, 561-580
62. Lazear, E. P. (2000). "Performance Pay and Productivity". *The American Economic Review*, 90 (5), 1346-1361
63. Le Grand, J. (1991). "Quasi-markets and Social Policy". *Economic Journal*, 101, 1256-67.
64. Le Grand, J. (1995), "Knights, Knaves or Pawns? Human Behaviour and Social Policy", *Journal of Social Policy*, 26(2), 149–169.
65. Leventis, A. (1997). "Cardiac Surgeons Under the Knife: A Patient Selection Hypothesis". Center for Economic Policy Studies Working Paper, Princeton University.
66. Levitt, S.D., & Dubner, S.J. (2005). *Freakonomics: A Rogue Economist Explores the Hidden Side of Everything*. New York: HarperCollins Publishers.
67. Locke, E. (1968). "Towards a theory of Task Motivation and Incentives". *Organisational Behaviour and Human Performance*, 3, 157-189.
68. Macrory, R. (2006) "Regulatory Justice: Making Sanctions Effective", London: Better Regulation Executive.
69. Makinson, J. (2000). "Incentives for change: Rewarding performance in national government networks". Public Services Productivity Panel, HMT.
70. Marsden, D. & French, S. (2002). "Performance Pay in the United Kingdom", 115-147, in Brown, M., Heywood, J. S. (Eds.). (2002). *Paying for Performance: An International Comparison*. New York, London: ME Sharpe.
71. Marshall, M., Shekelle, P., Brook, R. & Leatherman, S. (2000). *Dying to Know: Public Release of Information about Quality of Health Care*. London: Nuffield Trust.
72. Martin, S. & Smith, P.C. (1995). "Modelling waiting times for elective surgery". Occasional Paper, *Centre of Health Economics*, University of York, 24.
73. Marx, L.M. MacDonald, L. (2001) "Adverse specialization", *Journal of Political Economy*, 109(4), 864-899.
74. Mayntz, R. (1985). *Soziologie der öffentlichen Verwaltung (Sociology of the public service)*. 3rd ed., Heidelberg: Müller Jristicher Verlag.
75. Mellström, C. and Johannesson, M. (2005). "Crowding Out in Blood Donation: Was Titmuss Right?" Stockholm School of Economics.
76. Meyer, R.H. (1997.) "Value-added indicators of school performance: a primer". *Economics of Education Review*, 16(3), 283-301.
77. Milgrom P., & Roberts, J. (1990). "The economics of modern manufacturing: technology, strategy and organization", *American Economic Review*, 80(3), 511-528.
78. National Audit Office (NAO). (2001). "Measuring the performance of government departments - Report by the comptroller and auditor general". Available at http://www.nao.org.uk/publications/nao_reports/00-01/0001301.pdf.

79. National Audit Office (NAO). (2006). "Tackling Child Obesity – First Steps". Available at http://www.nao.org.uk/publications/nao_reports/05-06/0506801.pdf.
80. OECD. (2005). "Performance-related Pay Policies for Government Employees". Policy Brief. Available at: http://www.oecd.org/document/9/0,3343,de_2649_34139_35553801_1_1_1_1_00.html#ExecutiveSummary.
81. OECD. (2005). "Performance-related Pay Policies for Government Employees", 'Annex A: Performance-related Pay Policies Across 12 OECD Countries: Brief Overview'.
82. Office of the Deputy Prime Minister (ODPM). (2003). "Building on success – a guide to the second generation of local public service agreements".
83. Office of the Deputy Prime Minister, Addison & Associates with Arup. (2005). "Evaluation of Planning Delivery Grant 2004/05". London: Office of the Deputy Prime Minister.
84. Office for Standards in Education (OFSTED). (2005). "Specialist Schools: A Second Evaluation". February, Ref. HMI 2362, OFSTED, London.
85. OFSTED. (2007). "Review of the impact of inspection". Ref: 070042. Available at http://www.OFSTED.gov.uk/assets/Internet_Content/Shared_Content/Files/2007/may/impact07.pdf.
86. Ong, L. (2006). "Multiple Principals and Collective Action: China's Rural Credit Cooperatives and Poor Households' Access to Credit". *Journal of East Asian Studies*, 6 (2), 477-479.
87. Osbourne, S., Bovaird, T., Martin, S., Tricker, M., & Waterson, P. (1995). "Performance Management and Accountability in Complex Public Programmes". *Financial Accountability and Management*, 11, 19-37.
88. Paarsch, H.J. & Shearer, B. (2000). "Piece Rates, Fixed Rates and Incentives Effects: Statistical Evidence from Payroll Records". *International Economic Review*, 41(1), 59-92.
89. Petersen, L. A., Woodard, L.D., Urech, T., Daw, C., & Sookanan, S. (2006). "Does Pay-for-Performance Improve the Quality of Health Care?". *Annals of Internal Medicine*, 145 (4), 265-272.
90. Pidd, M. (2005). "Perversity in public service performance measurement". *International Journal of Productivity and Performance Management*, 54 (5/6), 482-493.
91. Prendergast, C. (1999). "The Provision of Incentives in Firms". *Journal of Economic Literature*, 37, 7-63.
92. Prendergast, C. (2001). "Selection and oversight in the public sector, with the Los Angeles Police department as an example". *National Bureau of Economic Research Working Paper* 37.
93. Prentice, G., Burgess, S., & Propper, C. (2007). "Performance-pay in the public sector: A Review of the issues and evidence". Office of Manpower Economics.

94. PriceWaterhouseCoopers (PwC) (Hibbert, D.). (2006). "Connecting Public Sector Pay to Service Delivery". Available at http://www.pwc.com/uk/eng/ins-sol/publ/pwc_publicpay-svcdelivery_nov06.pdf.
95. Propper, C. & Wilson, D. (2003). "The Use and Usefulness of Performance Measures in the Public Sector". CMPO Working Paper No. 03/073, CMPO, University of Bristol.
96. Propper, C., Sutton, M., Whitnall, C., & Windmeijer, F. (2008). "Did 'Targets and Terror' Reduce Waiting Times in England for Hospital Care?" *BE Journal of Economics Analysis and Policy* Vol 8,2
97. Seabright, P. (2004). "Continuous Preferences Can Cause Discontinuous Choices: An Application to the Impact of Incentives on Altruism". Mimeo, IDEI, Toulouse University.
98. Siegel, L.B. (2006). *Criminology :Theories, Patterns & Typologies*. 9th ed. Wadsworth Publishing.
99. Sharifi, S. & Bovaird, T. (1995). "The Financial Management Initiative in the UK Public Sector: The Symbolic Role of Performance Reporting". *International Journal of Public Administration*, 18, 467-90.
100. Smith, P. (1995). "On the Unintended Consequences of Publishing Performance Data in the Public Sector". *International Journal of Public Administration*, 18 (2/3), 277-310.
101. Upton, W. E. (1973) "Altruism, Attribution, and Intrinsic Motivation in the Recruitment of Blood Donors", Doctoral Dissertation, Cornell University.
102. Taylor, J. (2007). "Estimating the impact of the specialist schools programme on secondary school examination results in England". *Oxford Bulletin of Economics and Statistics*, 69, 445-471.
103. Titmuss, R. M. (1970). *The Gift Relationship*. London: Allen and Unwin.
104. West, A., & Pennell, H. (2000). "Publishing school examination results in England: incentives and consequences". *Educational Studies*, 26(4), 423-436.
105. Wiggins, A & Tymms, P. (2002). "Dysfunctional Effects of League Tables: A Comparison Between English and Scottish Primary Schools". *Public Money and Management*, 22(1), 43-48.
106. Wilson, D., Croxson, B. & Atkinson, A. (2006). "What gets measured gets done': headteachers' responses to the English secondary school performance management system". *Policy Studies*, 27(2), 153-171.
107. Wilson, D., & Piebalga, A. (2008). "Accurate performance measure but meaningless ranking exercise? An analysis of the English school league tables". CMPO Working Paper No. 07/176, CMPO, University of Bristol.