CMPO Working Paper Series No. 02/49

Does Performance Monitoring Work? A Review of the Evidence from the UK Public Sector, Excluding Health Care

Simon Burgess[1]
Carol Propper[2]
Deborah Wilson[3]

[1]CEP and CASE, LSE, CEPR and CMPO and Department of Economics, University of Bristol
[2]CEP and CASE, LSE and CMPO and Department of Economics, University of Bristol
[3]Department of Economics and International Development, University of Bath and CMPO, University of Bristol

July 2002

Abstract
This paper reviews the use of performance monitoring in the UK public sector, excluding its use in health care. Our focus is on finding robust evidence that evaluates the success of the introduction of performance monitoring in terms of its impact both on behaviour and on final outcomes. We begin with a general discussion of performance monitoring (hereafter PM), before considering the nature of the public sector and the implications of this for the implementation of such schemes within it. We then review the evidence and find a general lack of quantitative evidence on the impact of PM schemes on outcomes. This is partly due to the problem of attributing changes in outcome to the introduction of a specific PM scheme. One of our recommendations, therefore, is to consider piloting of PM schemes more widely in order to provide such evidence prior to national implementation.

JEL Classification: D23, J33, J45
Keywords: Performance Monitoring, Public Sector.

Acknowledgements
This paper was originally commissioned by CHI, the Commission for Health Improvement. The views expressed within the paper and any errors which remain are the responsibility of the authors.

Address for Correspondence
Department of Economics
University of Bristol
8 Woodland Road
Bristol
BS8 1TN
Tel: +44 (0)117 928 9844
D.Wilson@bristol.ac.uk

CMPO is funded by the Leverhulme Trust.

Summary

Performance monitoring has increased in the UK public sector. Our main findings from this review are:

- Performance monitoring can be used for different purposes. It may be intended to improve whole organisation performance, or to be more focused on individual units within organisations,or to achieve both ends at the same time.
- Results may or may be made public. If not published, then improved performance comes about by individuals' concerns over how their current performance will affect future pay or rewards. If made public, then the schemes may be linked to implicit or explicit incentives.
- There is little theoretical guidance as to when schemes should be introduced in the public sector, whether they should be linked to incentive schemes and how performance management schemes should interact with other implicit and explicit incentives designed to improve performance of the public sector.
- It has not been possible to find either theoretical guidance or evidence on the level of organisation (whole organisations, teams within organisations, the individual) within the public sector at which incentives should be linked to performance management.
- Recent analysis of incentives in the public sector stresses that those who are monitored will respond to monitoring in ways that maximises their benefits, which are not necessarily the ones of those designing the performance monitoring scheme; that different organisations within the public sector will have to be monitored in different ways, and that in general, incentives in the public sector may need to be less linked to performance than is desirable in the private sector.
- In practice, in the UK there has been a move towards more focused measures of performance, ones that are designed for explicit comparison between units within organisations and ones that are linked to sanctions or rewards.
- There are many examples of individuals responding to performance management. Such responses are not always what those implementing the scheme wanted or intended. But there is a general lack of quantitative evidence on the impact of such schemes on outcomes. Where it has been possible to identify improvements in performance in a specific context, it is often not possible to attribute such a change to the introduction or implementation of performance monitoring.
- There are clearly problems in the setting of targets with data that can be manipulated by those being monitored. Gaming responses appear to be common. However, there is also evidence that public sector employees care about more than the bonuses they may earn from incentive schemes.

On the basis of this survey, we would recommend that:

- Piloting of performance measurement schemes should be considered more widely. As seen from the Best Value examples discussed in section 4 of this review, pilot programmes did provide useful evidence that in turn informed the national implementation of the schemes.
- It may be important to distinguish process from outcome. Changes in both are considered as part of many of the performance monitoring schemes discussed in this review, while the objectives of such schemes are often stated in terms of

improved outcomes (increased quality or reduced costs of service delivery, for example). We need to better understand the link between process and outcome in order to ensure that monitoring of the former has the desired result.

- There may be scope for the development of targets based on alternative, independent information sources such as, for example, the British Crime Survey to set targets for police authorities, or the use of general household surveys to measure the health of people living in an area. The reason is that these are "non-corruptible" indicators or performance; ones that are not subject to manipulation by the individuals whose actions are being measured. Their use would force the relevant organisation to focus on what really mattered (for example, crime prevention, illness prevention) and it would also encourage them to find out what really mattered.

1. Introduction

This paper reviews the evidence on the use of performance monitoring in the public sector, excluding its use in health care. The aim of the review is twofold: to review the general features of performance monitoring as it has been used in the public sector, and to review the lessons from the experience of its use in the public sector outside health, primarily in the UK. The questions we address are:

- What is performance monitoring and what are the generic problems associated with it?
- Are there special features of the public sector that are likely to affect when and how performance monitoring should be used?
- Has performance monitoring in the UK public sector had any impact on behaviour, particularly final outcomes?
- Is there any evidence about the efficacy of linking performance monitoring to explicit financial rewards?

Section 2 is a general discussion of issues in performance monitoring (hereafter PM), covering what is performance monitoring, how might it be used, what form it may take and issues in implementing performance measurement. Performance measurement can take place in both public and private sectors. Section 3 therefore followed with a brief discussion of the nature of the public sector, and the implications of this for performance monitoring. Section 4 reviews the experience of performance monitoring in the UK public sector. This literature is large, but contains relatively little material to answer the particular question we address here: that of whether performance monitoring had an impact on final outcomes. In Section 5 we present the limited evidence on the use of individual incentives in the public sector, as performance measurement may be combined with the provision of direct financial incentives. Section 6 concludes.

2. Issues in Performance Monitoring

What is performance monitoring for?

Performance monitoring may be used to achieve several aims. These include:

- To improve the performance of individual units (such as particular schools, hospitals, police forces). This is often linked to 'best practice' exercises.
- To improve the performance of the overall organisation. In this case, the focus of the exercise is to improve the performance of the parent organisation as a whole, as well as possibly providing some developmental information for a single unit. For example, PM may improve the overall performance of the education system even if it does not give many clues of itself to the problems within any one school.
- To foster or generate pseudo-competition, for example, where purchasers in health care buy care from providers on the basis of measures of performance

- To improve accountability in the public sector (for example, to highlight "failing schools").

In the use made of performance monitoring in the UK public sector, we can see elements of all these aims, but they are often not clearly separately identified.

Given what it is for, how might performance monitoring work?

Figure 1 presents the possible ways performance monitoring may be used. The figure draws attention to the link between performance monitoring and the incentives the monitoring gives to individuals in the monitored organisations to improve performance.

Following the left-hand branch of the figure, the performance indicator (PI) information may be kept internal to the organisation, and not published. In this case it is a management tool. If a manager is given a task with a measurable PI, this might make them more likely to attempt to achieve it. Even if the PI is not linked to current rewards (either at the individual or organisational level), the fact that managers in the public sector often have career concerns may give the PI some 'bite' in that good performance against the PI will lead to a better job in the future. (Individuals have career concerns where their performance in the current job is positively correlated to the rewards they will get from future jobs.) A scheme that is not linked to direct rewards is clearly implementable and will have effects, provided managers have career concerns. Whether the effects are what is desired will depend on how well outcomes can be measured: we return to this below.

Following the right hand branch of the figure, the PI may be made public. In this case it may be linked to an incentive scheme. If so, the scheme may be explicit or implicit. In an explicit scheme a direct financial reward is made available to either the individual, a subgroup of the organisation (if one can be defined) or the whole organisation. Under an implicit scheme, the organisation (and not the individual) gets a financial reward as a result of the response of others to the PI. A classic example of this is a 'quasi-market', in which providers of services are rewarded for good performance by getting more contracts. In all these cases PM is intended to provide competitive pressure on organisations to improve, but the precise way in which PM brings about better results differs.

- If the PIs come with an attached explicit incentive scheme, then it is basically pay for performance at the organisational (or sub-organisational) level. The Public Service Agreements used by Treasury to give resources to government departments is an example of such a scheme: departments are meant to achieve targets with the resources they are given. Explicit incentive schemes linked to team performance are currently being piloted under the 'Makinson Report' pilots.
- To date, it is more common for PIs in the public sector in the UK to be linked to an incentive scheme that is implicit: one given in the form of client/service user/customer choice. The PIs then empower the client to make an informed choice. The classic examples of this are the quasi-market reforms to the UK public sector. In health, community care, housing and education, provider organisations were to get contracts on the basis of their performance. Initially,

there were few measures of performance, but over time, measures have increased (and are reviewed below).

- Even where there is no incentive scheme, explicit or implicit, publication of PIs may still have an effect on behaviour, for example, through individuals' pride in their 'league position', or avoiding a label of being a "failing" organisation. This is the idea behind 'name and shame' policies applied to schools.

These different ways of implementing PM are all seen in the UK public sector.

What form might PM take?

This paper is about PM for organisations in the public sector, not individuals. We have identified two main ways of instituting such PM:

- An in-depth evaluation of an organisation's processes and outcomes, typically involving a site-visit and large amounts of documentation. Examples are OFSTED visits, police inspections, QAA in universities, HMI Prison reports.
- The collection and publication of summary performance indicators. These can be broad or narrow in focus. For example, schools essentially face just three: truancy rates and two measures of GCSE pass rates. Local Authorities face a long list.

The more detailed measures are more expensive to collect, and if it can be shown that the summary measures provide as good a measure as more detailed ones, there is then a case for moving to such measures. In general, this has not been shown for the UK[1].

Whatever forms the PMs take, economic theory suggests that actors will respond to these in a way that will maximise their own personal benefit. Any scheme that is implemented must recognise that this will happen, and therefore there will be unintended outcomes.

Changes in the form and use of PM in the UK public sector

Over time, there has been change in the form of PIs used in the UK public sector. Mannion and Goddard (2000) find that across all the sectors they reviewed, there have been clear shifts in what data has been collected. These shifts are: from collection of data on narrow range of dimensions of performance towards development of indicator packages which reflect a broader assessment of organisational activity; from gratuitous collection of performance data towards the development of more streamlined and focused indicator packages; and some development of cross-sector or interface indicators where it has been recognised that organisational performance is partly reliant on actions of other agencies.

[1] Recent work in health in the US by McClellan and colleagues shows that some summary measures may be as good as much more detailed expensive measures for one particular treatment (Acute Myocardial Infarction).

There has also been change in how PIs are used. For example, open enrolment and overlapping catchment areas following the Education Reform Act of 1988 made it possible for schools to "compete" and thus the PI (exam pass rates) were then useful. Without those, the publication of PIs would have had to rely on "warm glow" effects. In general, Mannion and Goddard find there has been a general shift in use of information on performance away from primarily being used for internal management control purposes towards use of these data for external accountability and control. Performance data has been increasingly used to mediate contractual relations. There has been a shift away from informal performance assessments based on peer review or sample based inspection towards increased reliance on published performance league tables. Finally, there has been a shift towards use of performance information to facilitate participatory form of democracy and active citizenship.

While these changes in the form and aims of PM have occurred, these changes are based on experience in the use of PM tools, rather than on solid theoretical foundations of when and which PM tools should be used and when they should not. The question of when each type of performance management scheme should be used has not been systematically addressed. While there is a large literature in economics on the use of incentive schemes in the private sector, and a small but growing one of the use of schemes in the public sector, there is as yet no clear body of work that examines the conditions under which a scheme linked to incentives is desirable, or whether those incentives should be implicit or explicit rewards. Similarly, there is little work on whether such schemes should be used in conjunction with each other, or separately.

These issues remain to be resolved. The best that can be done is to learn from the scattered evidence that exists and we review this in Section 4. However, the use of PM in the public sector also requires careful consideration of how PM should be tailored to use in a public sector setting. This requires some consideration of what, if anything, is different about public sector provision.

3. The nature of the public sector

Performance monitoring is used in both the private and public sector. While many of the issues that arise in its use are common to both sectors, researchers studying the behaviour of public sector organisations have recently drawn attention to the fact that the public sector is different to the private sector and therefore a public sector organisation faced with a change in incentives will not necessarily behave in the same way as a private sector one. (As an example, see the influential case study of US bureaucracy, Wilson 1989). From this literature we identify some issues that appear particularly salient for the issue of performance monitoring.

Economists analysing the behaviour of individuals subject to different incentives have used a principal-agent framework. In this framework applied to public services, the principal is the user or the taxpayer and the agent is the provider of services. Both parties are motivated by self-interest, but the agent has better information that the principal. So, for example, a tax inspector working for the Inland Revenue has better information about whether a particular case needs investigation than the taxpayer. The issue is that the principal needs to design incentives schemes so that the agent uses this better information to achieve the goals of the principal, rather than the goals of

the agent. Within this framework, Dixit (1999) stresses two important features of the public sector. The first is that bureaucrats often serve several masters: these may include users of the service, payers for the service, politicians at different levels of government, professional organisations. The second, in part a consequence of the first, is that bureaucrats often have several ends to achieve. For example, they are often expected to increase efficiency whilst simultaneously increasing the equity of the delivery of public services. Dixit argues that these features (known as multiple principals and multiple tasks) mean that the provision of high- powered incentives (the use of contracts which reward individuals in a direct financial manner for particular outputs) are less likely to be suitable for the public sector than in the private sector where individuals may have to perform fewer, better defined tasks.

In the context of performance management, the Dixit argument suggests that linking performance monitoring directly to individual reward may be less desirable in the public sector than in the private. In addition, precisely because this is the case, the type of individuals found in the public sector may be more risk averse than those in the private sector, as the more risk averse will want to work in an environment where employment contracts are less high powered.

Le Grand (1997) argues that the view of the motivations of those providing, funding and receiving welfare from the UK welfare state has changed. From its inception in the late 1940s to the mid-1970s providers and funders of welfare services were seen as 'knights', eschewing self-interest to achieve the collective good. The users of the service, in contrast, were seen as passive 'pawns' prepared to take what they were given without complaint. This view then changed to one where in which all parties were viewed as pursuing their own self-interests: in Le Grand's terms, they behaved as 'knaves'. Le Grand points out that in fact that it is likely that individuals have a mixture of motivations and that design of the welfare state is better when it allows for this mixture of motivation. He also points out that the design on incentives may make individuals change their motivations. For example, he argues that giving high-powered financial rewards to doctors may turn them from knights to knaves, or at least increase the amount of knavish behaviour.

In the context of performance management, the Le Grand perspective emphasises the endogeneity of provider motivation to the type of performance management scheme. In other words, not only may individuals 'game' the system but the introduction of different methods of measuring performance and rewarding performance may attract different types of individuals to provide public services.

In an influential study of bureaucracy, Wilson (1989) argues that the public sector can be seen as encompassing four different types of organisation. The four different types of organisation arise according to whether the activities of providers can be observed or not, and whether the results of these activities (outcomes) can be observed.[2] The design and method of monitoring the provider organisation will depend on what can be observed. In the simplest case, labelled by Wilson as production organisations, activities and outcomes can both be observed. Examples of such organisations are the postal service and the tax collection service. He argues that such organisations can be

[2] Note that Wilson refers to activities as outputs, while in the principal-agent literature this is generally referred to as effort.

monitored in terms of outputs: in other words, performance monitoring of outcomes should be feasible. However, if there are several outputs, performance management can lead to problems when outcomes are defined too narrowly, so giving the organisation has a distorted focus.

Procedural organisations are ones in which activities can be observed, but outputs cannot. Examples cited by Wilson are health care providers, or the army in peacetime. Because outcomes are not observable, but activities are, such organisations will have both standard operating rules and have within them strong professional associations. However, the lack of measures of outcomes means too much emphasis is placed on activities and not enough on outcomes.

Craft organisations are ones in which activities are difficult to observe, but the results of these activities can be observed. Examples are the army at war, or governmental organisation that operate a long way from the centre. Wilson argues that these organisations will tend to develop strong decentralised structures, as the centre cannot easily measure activities. However, to prevent the wrong kind of actions from being undertaken, such parts of government need to develop a strong sense of mission.

Finally, coping organisations are ones in which neither actions nor outcomes are observed. Wilson suggests much police work falls into this category. In these it is difficult to generate objective, reliable measures of what is done. The best that management can do is to focus on recruitment, the generation of an atmosphere that is conducive to good work, and to react to complaints. These organisations are difficult to manage and there may often be conflict between managers and front line staff. In measuring activities and outcomes there will be a focus on what is most easily measured, and employees will be able to tailor their activities so they meet these targets, without necessarily improving the output of the organisation.

Wilson's analysis implies that different performance management strategies will be appropriate to, and needed for, different parts of government. It should be relatively easy to put in performance management in production organisations, and craft organisations can be assessed against performance against outcome targets. Of course, the issues of responses to these targets (gaming and concentration on measured behaviour/outcomes) will still remain. It will be more difficult to measure the outcomes of procedural organisations and performance assessment that is activity orientated may only increase the over-emphasis such organisations have on activities, rather than outcomes. Finally, coping organisations cannot be easily monitored. However, complaints systems may be used. (An analysis of the impact of complaints is provided by Prendergast (1999) and is reviewed in Section 4.)

4. Evidence on the Value of Performance Monitoring in the Public Sector

Introduction

This paper is not about simply documenting the existence of performance monitoring in the public sector. It is now widespread. The Public Service Agreement (PSA) structure also provides in principle a way of linking specific PIs to the wider aims of

government departments. Certainly this is how it is being used in the introduction of team-based performance pay in a number of Departments (the so-called "Makinson" Departments) in the "Incentives for Change" programme. Rather, we are interested in finding robust evidence that evaluates the success of the introduction of PM. It seems reasonable to assume that the general goal is to improve public service delivery and raise public sector efficiency. This assumption provided the focus for our search for evidence: we did not carry out a systematic review of the literature on performance monitoring; rather we looked for evidence on whether or not it had worked. We hence focused on existing reviews and evaluations of PM, discussions with key commentators in the specific areas as well as relevant Government documentation including commissioned reports. In addition we drew on our own previous work on benchmarking for privatised utilities and the use of individual incentives in the public sector.

One of the problems of such an evaluation is the lack of experimentation in government policy. Performance measures have been introduced, generally not in a controlled trial manner, but as a result of a policy change. Often they are accompanied by other changes in incentives. So, for example, league tables in schools were introduced across all schools, and come as part of the general reform of schooling provision. So it can be difficult to isolate the impact of introduction of PM from other policy changes that are implemented at the same time.

Given this caveat, we present our review by area of government. We also present a brief review of the lessons learnt from performance monitoring in the Utilities. Note however, that the use of PMs in this sector is for a slightly different purpose: PMs are used by the regulator as part of the regulatory toolkit, not by the organisation to raise its own performance.

Before we review the specific experience of different parts of government, we highlight general issues with the two types of PI distinguished above. Both have their problems. The in-depth, periodic, detailed, process and outcome (the OFSTED) type can encourage non-productive activity (i.e. trying to appeal to the inspectors rather than necessarily doing things that improve outcomes). They are very judgemental, and expensive to collect. It can be hard to ensure precise comparability across units. The detailed measures may also suffer from a more general problem which arises in the context of subjective performance appraisal - that of a tendency of the appraisers to rate everyone as the average. This bias is greater the longer the relationship between the appraisers and the appraised.

Regular (usually annual) summary, outcome-based indicators can be 'corruptible' and 'corrupting' (terms taken from Cooley, 1983). This means that the indicators themselves can be altered, and that they change behaviour, possibly in dysfunctional ways. Examples abound in the public sector: they include massaging of truancy rates in UK education (see below), massaging of waiting lists and treated cases in UK healthcare (see Smith 1995), unnecessary changes in the timing of graduation of workfare enrolees from schemes in the US (Courty and Marschke 1997). In economic terms, these indicators can be (and are) 'gamed'. Gaming can take many forms.

More generally, economic and other analyses have stressed that individuals will respond to performance indicators in ways that maximise their own utility or benefit.

This is not necessarily consistent with performance indicators improving welfare, and nor is it necessarily in ways that are expected by those that design the system. In economics, this issue is dealt with in terms of principal-agent model. The idea is that the supervisor (the principal) has to design a system so that the agent will do the right thing. The literature has many examples of both distorted indicators, and altered behaviour to improve the indicator at the expense of unmeasured things. In health Smith (1995) has given a list of unintended consequences of publishing PIs. These include tunnel vision; myopia; measure fixation; sub-optimisation; gaming; misrepresentation and misinterpretation[3]. While these are different forms of behaviour, all of these are due to the fact that the agent has different aims from the principal. As the principal tries to get higher effort (and so better public services) by implementing PI, the response may be better services but also may be other less desired behaviour. Note also that there is nothing that says that these responses are confined to the public sector.

Education

The indicators

Education is a very large and complex system. There are currently (2001) over 400,000 fte teachers, and 8.4 million pupils in 25,760 schools (of which 7% are independent). Each pupil receives at least 15,000 hours of compulsory 'treatment' from the system (Fitz-Gibbon and Tymms, 1999). Education, until recently, has been treated as a procedural organisation with concomitant emphasis on the role of professionalism. Now it has become relatively highly monitored by outsiders. There are two main systems of measuring performance in education. These are reports from the Office for Standards in Education (OFSTED), and summary performance indicators. These correspond to the two types of PM identified above.

OFSTED was set up in 1992 as part of the drive to raise "standards" in education. It replaced the system of inspection by HMIs (Her Majesty's Inspectors). It conducts pre-announced 4-day site visits to schools. Reports are published on the web, and all parents are sent a summary of the findings. The reports focus particularly on *process*. The decisions are necessarily judgmental. The tender to carry out the visits are competitively tendered, and carried out by teams of individuals. The cost of an OFSTED report on a school is not trivial: it averages at £60,000, which is equal to 2 – 3 annual teacher salaries. It also has potentially large indirect costs: the work undertaken specifically to create a picture for the OFSTED team but that may not necessarily benefit the pupils' education and may also add to teacher stress (Fitz-Gibbon and Tymms, 1999).

Performance indicators (PIs) appeared quite early in the UK, as an outcome of the school effectiveness research and as part of the move to introduce "quasi-market" (Le Grand, 1991) forces in education following Education Reform Act of 1988. This introduced local management of schools (devolved budgets), open enrolment and over-lapping catchment areas. These are clearly all crucial to giving parents choice over where the children go to school. Another key component is for parents to have

[3] Goddard, Mannion and Smith (2000) show how these can be derived from a principal-agent model.

the information on which to make a choice. This role was played by the introduction of the publication of performance indicators from 1992 (GCSE exam pass rates). Examples of PIs in education and their associated problems include:

- Truancy rates. Schools are required to publish these and it has been shown that schools reclassify truancy to be excused absences (Fitz-Gibbon 1996).
- Pass rates at GCSE. These may focus attention on the borderline students. While these are issued alphabetically, the press quickly turn them into league tables, ranked by score. There are a number of problems with the indicators, perhaps the most important being that they currently give 'raw' GCSE scores, not the value-added that schools actually provide. Thus they are a statement about the intake into schools as well as the effectiveness of the school in educating its pupils.

The evidence on effectiveness

With respect to OFSTED, the relevant question is does the existence of OFSTED monitoring raise the performance of the education system? This is both relative to no monitoring and relative to other forms of either detailed or summary types. And if it does, how does the benefit compare to the (large) costs?

In fact there is very little evidence and nothing to directly get at the above questions. There is a lot of evidence on how people feel about OFSTED. This is obviously useful as feedback for OFSTED in its operations, but not so useful as for gauging whether having OFSTED is a good thing or not. This evidence says that people found OFSTED very judgmental, demanding, often inaccurate and not effective in helping to develop schools. There are some differences between the views of parents, governors and teachers. OFSTED is not seen to develop schools and while it does put a lot of pressure on schools, such pressure is not necessarily of a useful kind (Centre for Evaluation of Public Policy and Practice, and the Helix Consulting Group 1999).

There is little evidence on the validity of OFSTED reports, for example, through linking actual data on pupil progress to inspectors' judgements on pupil progress. There are judgements about processes, but the problem in judging effectiveness by processes is that we do not really understand the link between processes and outcomes (Fitz-Gibbon 1999, and Fitz-Gibbon and Tymms, 1999). Fitz-Gibbon notes that the OFSTED judgements are often inaccurate (Fitz-Gibbon 1998). They have declared schools to be failing when in fact the pupils make average or better progress as measured by the YELLIS value-added score. The lack of inter-inspector reliability has been highlighted as a particular problem (Fitz-Gibbon 1998, Fitz-Gibbon and Stephenson-Forster 1999). The perception is that the judgements may not always be sound, and that therefore the process may not be contributing a lot to performance management. As there is no trial of OFSTED versus other detailed types of measures, it is not clear how much of this is generic to this mode of PM, and how much is due to the approach of OFSTED and its first Director.

With respect to the summary PI information, the precise question we aim to answer is does the use of PI information enhance the effectiveness of the service? If so, how? This might be because of direct or indirect financial incentives associated with customer choice or because people do not like being low down a league table, or classified as "failing". And do the benefits outweigh the costs?

The direct costs of the summary PIs in education are pretty minimal. GCSE exams, for example, are sat anyway, and are high-stakes exams whose integrity and marking are not in question. So there is little extra cost in publishing some aggregates of these numbers, nor in validating their authenticity. Computing value-added is a little harder, but again is not that costly.

While there is evidence that a particular school responds to an adverse performance indicator (whether it is part of the obligatory set of government benchmark results, or to the voluntary YELLIS system) this is interesting and useful, but what we are mainly interested in is whether there are systemic effects of PM.

Pupil performance, as measured by key stage results, has improved in the UK during the period in which PM has been implemented. What the evidence does not conclusively show is whether this improvement in performance can be linked to either PIs or OFSTED. Taylor and Fitzgibbon (1998) suggests that the use of summary performance indicators, while imperfect, in conjunction with the local management of schools initiative have all put pressure on schools to exert more effort[4].

There are two other sources of evidence that examine whether there are systematic results. The first is Bradley *et al* (2000) for the UK. They examine the impact of the publication of league tables and ask whether this system does put competitive pressure on schools, whether school enrolment does respond to PI information, and whether the pressure does help to raise performance. They analyse data from the School Performance Tables, combined with data on new admissions and other data. This data is now available for a run of years – a panel of schools from 1992 to 1998 – and panel data techniques were used. Because of the problem of omitted local context variables, analyses were also made looking at changes in outcomes.

The main findings were that:

- New admissions are positively related to school's own exam performance, and negatively related to exam performance of its competitors in same school district;
- The impact of the school's comparative exam performance on new admissions increased after the introduction of quasi-market forces;
- Schools achieve better exam results when they are in competition with schools with good exam performance but the impact of this is small;
- Excess demand for places in popular schools has led to an increase in capacity at those schools;
- Greater parental choice and increased competition have led to some polarisation with respect to family background.

The use of PIs allied with (albeit implicit) incentives appears to have produced an effect on outcomes. Note that these are the outcomes measured by the PIs (which may not be all the outcomes that are desired by the government, teachers or parents). On the negative side, there are possible countervailing effects on equity. Note that the PIs are not directly corruptible as GCSE exam scores are external. But they are indirectly

[4] OFSTED, by contrast, may not have additionally improved performance, particularly given the costs – direct and indirect – of its implementation.

corruptible by the school's choice of who to submit for examination, and there is some evidence of this selection effect. Finally, what the Bradley et al analysis does not tell us is whether this competition on an imperfect indicator has improved or worsened performance on other indicators, including a better indicator of school effectiveness, namely value-added.

Minter Hoxby (2001) examines the introduction of a "report card" scheme for schools in the USA. These carry no explicit incentives, their aim is simply to inform. They report the result of testing in schools on a state-wide basis, and are intended to be user-friendly. She asks "How much can one expect from a policy that just informs, with few stakes?" In fact, she shows that this reform does appear to have had an effect. States that introduced testing and report cards early saw reading and maths scores improve faster than those states that chose to stay out of the scheme till later. The evidence for this conclusion is based on national standardised testing and not the tests that were used in the report cards, so the teachers were not "teaching to the test".

Local Government

Local authorities are currently subject to several elements of external review of their performance, including Audit Commission national thematic studies, performance aspects of local audit, Best Value inspections, as well as service specific inspections such as OFSTED and SSI (Social Services Inspectorate) (Byatt and Lyons 2001). In this section we concentrate on the evidence regarding the impact of Best Value.

In 1999 the government's Best Value legislation created the Best Value inspection service that is responsible for a comprehensive inspection of all local authority services (Davis *et al* 2001). Under this legislation, local authorities (as well as police and fire authorities) are required to continually improve performance with regard to a combination of economy, efficiency and effectiveness (Mannion and Goddard 2001). Each authority now has to implement a Performance Management Framework, involving the establishment of objectives and performance measures as well as a programme of annual review (*ibid*, Figure 5.1, page 125). The emphasis is on improvement of service delivery through the setting of targets and independent inspection. So this is mainly PM by means of the "internal" branch in Figure 1. External routes might be seen as less important here because the scope for customer choice is lower. People are unlikely to change where they live because LA services are poorer, and local elections are not a very powerful force for improving local services.

Best Value Performance Indicators (BVPIs) are a central part of the new performance monitoring scheme. The purpose of these indicators is threefold: first, to provide information to the public; second, to be used comparatively in order to improve authority performance over time (this is envisaged to happen through the sharing of best practice and benchmarking clubs); third, as part of inspection procedures by, for example, OFSTED or SSI (*ibid*; see appendix three for a full list of (over two hundred) BVPIs). One feature of BVPIs is that they have been designed and

developed to reflect local priorities (partly through the pilot process discussed below) as well as national targets.

Best value performance indicators were only introduced across all authorities in 2001, hence it is too soon to assess their impact (Mannion and Goddard 2001). The Best Value scheme was, however, piloted in 42 local authorities and the pilot evaluated on behalf of the DETR (now DLTR). Martin *et al* (2001) provides some evidence on the impact of this system of performance monitoring. The authors distinguish three areas in which the impact of the scheme may be evaluated: learning outcomes, process outcomes and service outcomes. The first area is particularly relevant to how the pilot programmes have been able to inform the national implementation of the scheme and will not be discussed further here. With regard to process, several changes were noted.

- Pilot authorities developed and used more standardised procedures, including the introduction of a five-year review period as part of a new set of strategic priorities, as well as Best Value training programmes and a Best Value "toolkit" for staff.
- There was an increase in the sharing of information and best practice between authorities, including the formation of benchmarking clubs and performance networks. In addition, the new regime enabled staff to raise issues that previously had not been fully addressed by the authority.
- One problem identified by participants, however, was the fear that focus would be given only to those activities that demonstrated compliance with the review process. This is again an example of tunnel vision.

A key question is whether the introduction of the Best Value regime has had an impact on service outcome. Even given the short duration of the pilot programmes, Martin *et al* (2001) state that tangible service improvements can be linked directly to the implementation of the new system of inspection and review.

- For example: Camden made productivity increases which enabled the provision of an additional 70,000 hours of care at no extra cost; in Surrey, the introduction of joint caretaking arrangements led to improved use of community buildings; Portsmouth doubled the number of dyslexic children being taught for the same cost (*ibid*, page 5).
- In many pilot authorities, the implementation of Best Value led to improvements both in terms of increases in quality and/or responsiveness of service delivery as well in terms of cost savings achieved. Services that had previously failed relative to target were particularly improved, and some authorities additionally set more demanding targets regarding quality of service delivery. Unfortunately, the authors were unable to determine a cost/benefit ratio associated with the piloting.
- While the authorities did identify significant costs, a proportion of these were recognised to be set-up costs, with the implication that, in the long term, the benefit from the Best Value regime would outweigh the (direct) costs of its implementation.

The evidence regarding the impact of Best Value from the pilot authorities, therefore, seems generally positive, although neither costs nor benefits have been able to be quantified, and the problem of tunnel vision remains. The short duration of the pilot

programmes means that any evaluation thus far is necessarily preliminary and it will be important to monitor the impact of the evolving statutory regime.

Housing

The targets set by the Best Value regime at local authority level include ones specific to council housing services. In this section we focus on evidence regarding the impact of the same performance monitoring scheme during a pilot of registered social landlords (RSLs).

The Housing Corporation issued guidance in February 1999 which welcomed the "principles underpinning the Government's proposals [regarding Best Value] as just as relevant to the way RSLs run their businesses and provide services as they are to local authorities" (Walker *et al* 2000, page 1). In October 1998, 23 RSLs were accepted as Best Value pilots, and a team from the University of Birmingham evaluated the impact of the scheme through 1999/2000 (see the appendix in Walker *et al* (2000) for a full list of the pilots and their main purposes). A striking feature of this pilot process was the range and diversity both of the RSLs that participated and of the type of activity they chose to review within the framework provided by the Best Value pilot programme. One implication of this wide range is that general lessons regarding implementation of a national scheme are necessarily more difficult to draw (*ibid*).

The authors concede that it did not prove possible to measure impact in any kind of structured, formal way, but they were able to identify and assess four types of impact resulting from the review process, namely: organisational learning, changes in process and milestones achieved, measured performance change, measured changes in service costs. As organisational learning primarily relates to the lessons learned from the pilots for national implementation of the RSL Best Value regime it will not be discussed further here.

Changes in process and milestones achieved accounted for the most widespread type of impact (partly due to the nature of activities undertaken by the pilots). These occurred when a specific activity was carried out or some procedure changed as a consequence of the new system, and the authors identify many positive changes within this category. For example, one objective for the Anchor Retirement Trust was to set up a database to monitor energy consumption at each of its schemes. In November 1999 it became accredited for Energy Efficiency. Eastleigh Housing Association made the delivery of services to tenants in extra care schemes the focus of its Best Value pilot. As these examples illustrate, such changes often relate to improvements in service quality and/or in RSL-user relations and thus are ongoing in nature and difficult to quantify.

In addition to such procedural changes, approximately one third of the pilots developed and used performance indicators, enabling changes in performance to be measured and quantified. While the authors state that the one-year time frame of the pilot study is too short to determine whether Best Value had a "significant" impact on performance, they make several observations.

- Almost all pilots were able to demonstrate improved service on at least some measures, even if they had not fully met the target(s).
- While some RSLs responded slowly to the targets, others showed an initial improvement in performance that then tailed off.
- A conflict in performance between different measures was observed in some pilots, and the authors noted the difficulty in interpreting performance figures in isolation from general trends. This appears to be a more general concern with the measurement of the impact of such performance targets.

Changes in service cost proved both the least widely included in individual RSL's pilot programmes and the most difficult to gather any evidence on. The problem of obtaining any quantifiable information regarding the relative costs and benefits of the Best Value programme appears to be as much of an issue in the social housing sector as it is for local authorities.

Police

The police service is subject to inspection by Her Majesty's Inspectorate of Constabulary (HMIC). HMIC is an independent inspectorate, funded by central government and located outside the tripartite policing structure but working closely with the three parties (Home Secretary, police authorities, police forces) (Vass and Simmonds 2001).

There is little substantive evidence of the impact of the monitoring process to which police forces are subjected, or indeed of the costs and benefits of this process. The system has become more focused in recent years, targeting resources more towards poorer performing forces and measuring performance or outcomes relative to a number of specific protocols (Vass and Simmonds 2001). In part this is in response to the perception that the previous system – comprehensive review of each police force every 18 months – was too burdensome, although there is no evidence of any evaluation to support such a perception (*ibid*).

It is certainly the case that in general terms police performance has improved over time as measured by, for example, general crime rate indicators (Mannion and Goddard 2001). It is difficult, however, to directly attribute such improvements to the impact of any inspection or monitoring process. While this is a general problem with the evaluation of the impact of performance monitoring schemes, the level of complexity of policy activity makes it a particular issue in this context. Such complexities also suggest that a simple set of performance measures will not be sufficient: rather they should be used in conjunction with HMIC in-depth reports and means of sharing best practice (*ibid*).

This view is supported by the findings of Policing for London (FitzGerald and Hough 2002), a report commissioned by the Metropolitan Police in response to the report of the MacPherson Inquiry into the murder of Stephen Lawrence. A key finding of this report is that the introduction of the performance management schemes in the 1990s which emphasised quantified performance targets while ignoring the complexities of police work has led to a reduction in responsiveness to local need and a fall in staff morale. The authors subsequently recommend a performance management system

which redresses the balance between the achievement of professional standards and the hitting of numerical targets. This will help reduce the problem of tunnel vision (Mannion and Goddard 2001), whereby the incentive is created to focus on measurable aspects such as crime rates at the expense of non-measurable activities such as crime prevention. Moreover, numerical targets involving crime rates are particularly open to manipulation. Burrows et al (2001) investigate the "recording shortfall", i.e. the discrepancy between the British Crime Survey's estimate of the number of crimes and the figures appearing in police recorded crime statistics. The authors identify the exercise of "police discretion" in terms of what is recorded as a crime as the major reason for this discrepancy. A key recommendation of the report is that counting rules need to be clarified in order to achieve more consistency in the recording of crimes across police forces. This seems to be a necessary pre-requisite for published measures or targets incorporating such statistics to provide meaningful comparisons of performance.

Wilson (1989) identifies the police as a coping organisation, one in activities and outcomes are difficult to observe. So we would expect it to be difficult to implement useful PM. Wilson argues that in such organisations, one way of ensuring that aims are being achieved is to respond to complaints. Prendergast (2000) uses principal-agent theory to analyse the effect of responding to complaints on public servants behaviour, and empirically examines the effect of a new complaints system on behaviour of the Los Angeles Police Department (the LAPD). He argues that public officials allocate goods to consumers, and that consumers get rents (benefits) from these allocations. As they do not pay for them, they will only complain if they are denied the service, and not if the service is incorrectly allocated. Thus consumers cannot be relied upon to state when allocations are correct or not. Investigation of complaints harms public officials on the grounds that even if the complaint is shown to be incorrect, 'some mud always sticks'.

Using this analysis, Prendergast argues that external monitoring of complaints is not necessarily a good incentive for better performance. As consumers only complain when they have unfairly been denied a good, the only complaints that are investigated are those where the bureaucrats are wrong. This will mean that external investigators become biased against the public officials. In addition, in order not to get investigated, the bureaucrats are likely to do less (so they have less chance of being investigated). They may also ignore legitimate complaints (as they fear that some mud always sticks), accede to consumer demands to avoid complaints (i.e. give the good when they shouldn't) monitor good decisions too much and delay decision making to be more confident (this doesn't harm them, as they aren't paid depending whether they give the good or not, but does harm the consumer).

Prendergast (2000) found that when external monitoring of the LAPD was increased, all these outcomes occurred. The police did less (they took up a policy of 'drive and wave') and this resulted in a decrease on assault rates on officers but an increase in homicides. One interpretation of this is that since 1998 officers have been responding to increased oversight by actions, which, although keeping them out of trouble, also results in higher crime.

Social Services

Personal social services are monitored by the Social Services Inspectorate (SSI), for which the Select Committee on Health has responsibility. The Select Committee has yet to undertake any review of the impact of the SSI (Vass and Simmonds 2001). Again there has been a general improvement in performance in this sector (Mannion and Goddard 2001), but as with the police service, this is difficult to attribute to the implementation of any specific performance monitoring scheme. We have found no quantitative evidence on the relative costs and benefits of the SSI.

Privatised Utilities

The objective of performance monitoring of the privatised utilities is slightly different. Specifically, the regulator may impose such a PM system in order to obtain a framework within which cost data can be compared and thus a pricing regime imposed and/or pseudo-competition created within the sector. A relatively crude comparative cost analysis underpinned the initial price limits set for the water industry at the time of privatisation in 1989, for example (Grout *et al* 2000). These comparative cost data are now embedded in a system that has the central objective of providing incentives for efficiency improvement. Companies which appear, from comparisons, to be operating inefficiently are penalised by not being allowed to increase prices by as much as the more efficient companies. Furthermore, the comparative data, much of which is in the public domain, provide information to others such as shareholders, analysts and customers who can also apply pressure to companies that appear to be inefficient to improve their performance.

There is strong evidence that the efficiency of the utilities has been improving since they were privatised. What is less clear is the extent to which such efficiency benefits can be attributed to the comparative competition or benchmarking frameworks employed. In this context (as in those discussed above) it has proved difficult to show causality between the implementation of a performance monitoring scheme and an improvement in outcome.

There are, however, useful lessons that can be drawn from the experience of benchmarking in the utilities sector for PM in the public sector:

- Getting comparable data and a generally acceptable framework of analysis has taken a long time. For example, the water industry was privatised in 1989 but the comparative competition framework was still being adjusted and refined significantly in the 1999 price review.
- The related issue of setting the appropriate benchmark has proved difficult, as has knowing how far to push the quantitative comparative analysis. But the degree of judgement required has declined as the quality of data has improved.

This suggests PM in the public sector will take a while to ‘bed down’. In addition, the level of complexity of some of the activity being monitored in different areas of the public sector could dictate that some degree of judgement, in addition to the setting of specific targets, might be required in the long term.

5. Evidence on Individual Incentives in the Public Sector

Here we look briefly at the evidence on the response of individuals to incentives with in the public sector. This is relevant because whether organisations respond to PM is likely to be closely related to whether individuals respond to incentives. In addition, PMs may be linked to explicit incentives as discussed above.

There is a well-developed theoretical literature on incentives for individuals. Using a principal-agent approach, economists have characterised the nature of optimal incentive schemes. The precision of output measures, the importance of outside factors in influencing those outcomes, and individuals' attitudes to risk all matter. The greater the precision of the performance measure, the less risk averse the monitored individual and the lower the importance of outside factors in determining outcomes, the more incentives should be related to measured output (the more 'high powered' the incentive scheme should be). The theory has been extended to cover the case of individuals with different facets to their jobs (or different tasks), some easily measured (for example, quantity of output) and others more costly to measure (quality). This shows that agents will divert their activities to those that are measured and those that are more easily done (so we would expect to see tunnel vision).

Almost all of this work has been developed for the private sector, for incentives set by profit-maximising organisations. More recently, a literature has developed that studies the same problem of incentives located in the public sector. This literature shows that, in the presence of features common to the public sector (for example, multiple principals, multiple tasks, measurement problems), incentives need to be designed carefully; what works in the private sector may have different effects in the public sector. In general, many of the current analyses support the use of less high powered schemes in the public than in the private sector (e.g. Dixit 1999).

While there is now a wealth of evidence on incentives in the private sector (see Prendergast, 1999, for a recent review), there is very little evidence for the public sector (Burgess and Metcalfe, 1999a, review this). We summarise the few studies that have been undertaken.

- Workers do react in significant ways to financial incentive schemes. The evidence suggests that, in general, workers do work harder and produce more output when they are incentivised to do so.
- Workers react in sophisticated ways, manipulating the quality or timing of what they do. These are generally responses that the organisation neither intended nor wanted. For example, in studying a job training scheme in the United States (the JTPA), Courty and Marschke (1997) found that the incentive scheme led to 'gaming' to achieve targets and so bonus payments, and that this behaviour was welfare decreasing. For the UK healthcare sector, Croxson et al (2001) provides evidence that GP Fund Holder manipulated the timing of referrals to increase practice income above what is otherwise would have been.
- Some public sector workers are motivated by more than just their own income. Case workers in the JTPA in the United States systematically took on the hardest-to-place workers even though their narrow financial interest (and possibly social

welfare) was better served by selecting more employable workers (see Heckman et al, 1996). For the UK healthcare sector, Propper et al (1999) provides evidence that GP Fund Holders decreased the waiting times for some of their patients, even though they received no direct financial benefit for this.

- Theoretical work has proposed a set of factors that may influence whether any particular organisation would find it optimal to use incentive pay. There is some evidence to support these hypotheses, in the form of detailed case studies and a broad cross-section of UK establishments, both public and private (Burgess and Metcalfe, 1999b).

Finally, individuals may be rewarded financially on the basis of hitting team targets. The literature on teams in private sector organisations, in which there is a monetary output that accrues to the owner of the firm, is well developed. Various solutions have been proposed in this literature to overcome the problems of free-riding in teams (for a review see Ratto et al 2001). The analyses of teams and team-based incentives in the public sector is much less well developed than that in the private sector. The nature of the public sector – in which there is no owner of the enterprise, and no monetary output – means the solutions suggested for the public sector cannot be directly applied. These is little theoretical guidance on the optimal size of teams in the public sector (Ratto et al 2001). However, in practice, team rewards have been used in the public sector outside the UK.

6. Conclusions

Performance management is very much part of the UK public sector. This review has highlighted the fact that there are many examples of individuals responding to these schemes in different parts of the UK public sector, and some evidence on how this has affected processes. But we know little about their impact on outcomes, and the costs of achieving these outcomes. So there is almost no evidence on whether these schemes improved the efficiency of the service delivered. On the basis of these findings, we would recommend that:

- Piloting of performance measurement schemes should be considered more widely. As seen from the Best Value examples discussed in section 4 of this review, pilot programmes did provide useful evidence that in turn informed the national implementation of the schemes.
- It may be important to distinguish process from outcome. Changes in both are considered as part of many of the performance monitoring schemes discussed in this review, while the objectives of such schemes are often stated in terms of improved outcomes (increased quality or reduced costs of service delivery, for example). We need to better understand the link between process and outcome in order to ensure that monitoring of the former has the desired result.
- There may be scope for the development of targets based on alternative, independent information sources such as, for example, the British Crime Survey to set targets for police authorities, or the use of general household surveys to measure the health of people living in an area. The reason is that these are "non-corruptible" indicators or performance; they are indicators that are not subject to

manipulation by the individuals whose actions are being measured. Their use would force the relevant organisation to focus on what really mattered (for example, crime prevention, illness prevention) and it would also encourage them to find out what really mattered. As areas differ, giving all groups the same level of an unadjusted output would be unfair. However, there are now well-developed techniques for adjusting for differences in outputs between areas that are outside the control of those who performance is being measured. These have been used by OFWAT and others, and include comparative cost/benchmarking, or setting targets as changes with previous levels (so differencing out area specific effects).

References

Bradley, S, R Crouchley, J Millington and J Taylor (2000), Testing for Quasi-Market Forces in Secondary Education, Oxford Bulletin of Economics and Statistics, 62(3): 357-390

Burgess, S and Metcalfe P (1999a) The Use of Incentive Schemes in the Public and Private Sector: Evidence from British Establishments. CMPO, University of Bristol Working Paper 99/015.

Burgess, S and Metcalfe, P (1999b) Incentives in Organisations: A Selective Review of the Literature with application to the Public Sector. CMPO, University of Bristol Working Paper 99/016.

Burrows, J, R Tarling, A Mackie, R Lewis and G Taylor (2000), Review of Police Forces' Crime Recording Practices, Home Office Research Study 204, www.homeoffice.gov.uk/rds/horspubs1.html, 12/06/02

Byatt, I and M Lyons (2001), *Role of External Review in Improving Performance*, Public Services Productivity Panel

Centre for Evaluation of Public Policy and Practice, and the Helix Consulting Group. 1999. The OFSTED System of School Inspection: An Independent Evaluation.

Courty, P and Marschke, G (1997) Measuring Government Performance: Lessons from a Federal Job Training Programme. American Economic Review 87(12). Papers and Proceedings, May, 383-388.

Croxson, B, Propper C and A Perkins (2001) Do Doctors Respond to Financial Incentives: UK family Doctors and the GP Fundholder Scheme. Journal of Public Economics 79 (2), 375-398.

Davis, H, S Martin and J Downe (2001), The impact of external inspection on local government, http://www.jrf.org.uk/knowledge/findings/government/921.asp, 15/02/02

Dixit, A (1999) Incentives and Organizations in the Public Sector: An Interpretative review. Mimeo, Princeton University.

Fitzgerald, M and M Hough (2002), Policing for London: Responding to Diversity, www.policingforlondon.org, 13/06/02

Fitz-Gibbon, C.T. 1996 Monitoring Education: Indicators, Quality and Effectiveness London Cassell

Fitz-Gibbon, C.T. 1998 OFSTED: Time to go? Managing Schools Today, Vol 7, No 6, 1998, 22-25

Fitz-Gibbon, C.T. 1999 Education: High Potential Not Yet Realized . Public Money & Management: Integrating Theory and Practice in Public Management, Vol 19, No 1, 33.40

Fitz-Gibbon, C.T. 1999. Quality, Science and Soros's Reflexivity Concept: A Value-Added Approach. In Balázs, É., van Wieringen, F. and Watson, L.E. (eds) Quality and Educational Management: A European Issue. Wolter Kluwers Group.

Fitz-Gibbon, C.T. and 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

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, vol. 10 no. 6

Goddard, M., Mannion, R. and Smith, P. 2000 Enhancing Performance in Health Care; A Theoretical Perspective on Agency and the Role of Information. Health Economics vol. 9 pp. 95 – 107.

Grout, P, A Jenkins and C Propper (2000), *Benchmarking and Incentives in the NHS*, Office of Health Economics, London

Heckman, J, Smoth, J and C Taber (1996) What do bureaucrats do? The effects of performance standards and bureaucratic preferences on acceptance into the JTPA program. In Advances in the Study of Entrepreneurship, Innovation and Growth, Volume 7, JAI Press, 191-217.

Hoxby, C. M. (2001) Testing is about Openness and Opennesss Works.
http://post.economics.harvard.edu/faculty/hoxby/papers/NAEP_results_jun01.pdf
http://www-hoover.stanford.edu/pubaffairs/we/current/hoxby_0701.html,

Le Grand, J (1997) Knights, Knaves or Pawns? Human Behaviour and Social Policy. Journal of Social Policy, 26, 2 149-169.

Mannion and Goddard (2001), *The Impact of Performance Measurement in the NHS: Report 3: Performance Measurement Systems: A Cross-Sectoral Study*, Report prepared for the Department of Health, Centre for Health Economics, University of York, York

Martin, S, H Davis, T Bovaird, J Downe, M Geddes, J Hartley, M Lewis, I Sanderson, P Sapwell (2001), *Improving Local Public Services: Final Evaluation of the Best Value Pilot Programme*, Warwick Business School and DETR

Prendergast, C (2000) The Limits of Bureaucratic Efficiency NBER.

Propper, C, Croxson, B and Shearer, A (forthcoming) Waiting times for hospital admissions: the impact of GP fundholding, Journal of Health Economics.

Ratto, M et al (2001) Team based Incentives in the NHS: An Economic Analysis. CMPO, University of Bristol Working Paper 01/037.

Smith, P. 1995. On the unintended consequences of publishing performance data in the public sector. International Journal of Public Administration vol. 18 pp. 277 – 310

Vass, P and G Simmonds (2001), External Review: A Report by the Centre for the Study of Regulated Industries, Supporting Document 1 to: Byatt, I and M Lyons (2001), *Role of External Review in Improving Performance*, Public Services Productivity Panel, http://www.hm-treasury.gov.uk/pspp, 01.02.02

Walker, B, D Mullins, P Niner, A Jones and K Spencer (2000), *The Evaluation of the RSL Best Value Pilots*, University of Birmingham, School of Public Policy

Wilson (1989). Bureaucracy. New York: Basic Books.

Figure 1: Performance Monitoring