



HM TREASURY

Public Services Productivity

Papers presented
at a seminar
held in HM Treasury
on 13 June 2002

Foreword

The central objective of this Government is to build a stronger, more enterprising economy, and a fairer, more equal society; extending economic opportunity and supporting those in need to ensure that rising national prosperity is shared by all.

To increase the productivity of the UK economy, since 1997, we have introduced important reforms to promote macroeconomic stability, reform the labour market and improve the microeconomic environment. Tough decisions, but they worked, and so now we have low inflation, low interest rates, and record levels of employment.

A productive economy and a productive public sector are complementary and mutually reinforcing. So the next step is investment and reform in the public sector. The monolithic structures of the past will not stand. We need now to adopt new approaches, to ensure value for money for the taxpayer, and to deliver public services flexibly, around the needs of the individual.

This is a time of unprecedented interest in public policy and economic governance. In think tanks and in universities, up and down the country and around the world, a new generation of academics is coming forward with the next generation of policy proposals. In Government, we have to draw on that expertise, and build the innovative ideas emerging from the academic community into the policy process. As part of that programme, on 13 June 2002, we invited a number of leading academics to a seminar at HM Treasury.

I am grateful to all who attended the seminar, especially to those who presented papers and have kindly agreed to make them available for publication in this volume. As these papers illustrate, there is a full programme of work ahead. Improving public service productivity means clarifying service providers' objectives, establishing appropriate incentive mechanisms, and obtaining good information on performance - all important objectives for policymakers. The challenge for us, in Government, is to work together to turn those ideas into reality – raising the standard across the board.



Paul Boateng

Chief Secretary to the Treasury

Public Services Productivity

Papers presented at a seminar
held in HM Treasury on 13 June 2002

Contents

Economic Principles of Reform

1. Making Government responsive
Timothy Besley
(Professor of Economics, London School of Economics) page 7
 2. Incomplete contracts and public ownership: Remarks, and an application to public-private partnerships
Oliver Hart
(Professor of Economics, Harvard University) page 11
 3. Models of public service provision: Command and control, networks or quasi-markets?
Julian Le Grand
(Professor of Social Policy, London School of Economics) page 15
-

Theory into Practice – Evidence on Reform

4. The response of the Los Angeles Police Department to increased oversight
Canice Prendergast
(Professor of Economics, University of Chicago) page 19
 5. Is performance related pay for teachers possible?
Peter Dolton
(Professor of Economics, University of Newcastle) page 23
 6. Value for money, educational resourcing and pupil attainment
David Mayston
(Professor of Public Sector Economics, University of York) page 27
 7. Public purchasing of health services
Jim Malcomson
(Professor of Economics, University of Oxford) page 33
 8. Performance management in health care: Information, incentives and culture
Peter Smith, Russell Mannion and Maria Goddard
(Centre for Health Economics, University of York) page 37
-

Making Government Responsive

Timothy Besley

Professor of Economics
London School of Economics

Introduction

This paper deals with the question of how to provide public services that are responsive to the needs and wishes of consumers. For private goods and services traded in markets, responsiveness is achieved through the price mechanism. If a firm tries to produce and sell a good that nobody will buy, then it will go out of business. An entrepreneur who innovates to fill a market niche can charge the consumers and may thereby make a profit. Thus pursuit of profit in the market place can engender a degree of responsiveness to consumer needs.

The important characteristics of the market as an accountability mechanism are:

- **Exit** – a dissatisfied consumer can purchase from an alternative supplier.
- **Financial incentives** – a responsive firm can earn higher profit.

This kind of incentive structure works best when consumers are mobile and well informed. There are cases where even for private goods, the market can fail to achieve a highly responsive outcome. Goods which are purchased infrequently or whose usefulness becomes apparent only after a prolonged period of use may see poor quality firms survive for long periods. With appropriate legal protection it may also be difficult for consumers to call poor performing firms to account for supplying low quality or faulty products. There is a well-established body of regulations which underpin the functioning of markets and which try to prevent such problems. The market may also provide its own response in the form of information providers such as the Consumers' Association and professional associations such as the Law Society. Recent policy discussions have raised the possibility of using more market like mechanisms for the delivery of public services. We frame this in the context of general discussions of what makes government responsive. For both historical reasons and, given the nature of the services in question, government takes a key role in service provision. Traditionally, this has been primarily as the service provider. However, the broader role of government is to ensure that these services are provided efficiently for its citizens. Direct public provision is only one way of doing this.

The observations below present ideas for discussion and are at a high level of generality – unsuitable for giving policy advice on any particular issue. However, they will hopefully provide a useful starting point and overview of some general principles that deserve attention in the on-going debate about public service reform.

We discuss two main paths for improving responsiveness – increasing consumer choice in public service provision and changing the governance structure for provision of services so as to give consumers a greater say in how services are provided. We emphasize that the success of either rests on how well proposed changes allow private and social benefits/costs of public services to be aligned. We emphasize that the organization of production can change the way in which producers respond to consumers.

We use the metric of responsiveness to consider public service reform. This is closely linked to the idea that providers of services are more accountable for the decisions that they make. Accountability is a means to an end with responsiveness being the desired outcome.

What Makes Public Services Different?

Before discussing government responsiveness, it is important to see why public services are different and why traditional market provision is unlikely to be the solution of choice. Public services are best defined by that set of activities where social and private returns diverge. The main reasons identified for this in the literature are three fold: (i) positive externalities in consumption or production; (ii) minimum service constraints – such as providing universal access to health care or education, and (iii) ignorance and/or bounded rationality of consumers.

In all three cases, private and social returns to providing and consuming public services need not be aligned. This provides the basic rationale for public intervention. Institutions for delivery need to align private and social returns. The fundamental challenge of public service provision is to find the most efficient mechanism for equating private and social returns.

Viewing public services as defined by the divergence of social and private returns cuts across the traditional view of public services as those activities provided by publicly owned firms. The mode of provision is contingent on the policy being pursued whereas the nature of the good and the benefits that it generates is a more fundamental characteristic. On this view, even private health and education may legitimately be thought of as being part of the public sector while a publicly produced car (as we had in the days of British Leyland when it was nationalized) was never a part of the public sector according to this definition. This output oriented definition is useful in suggesting that the public ownership question is a sort of veil which sometimes masks the real economic issues. What really counts is efficient and equitable provision not the mode of delivery *per se*.

Pure market provision is unlikely to achieve the socially desirable outcome – responsiveness based on profit seeking need not achieve what is good for society. To illustrate, consider the case of market driven incentives for rail safety. Society as a whole (not just rail passengers) may care about rail safety. However only the willingness to pay for safety of passengers is reflected in ticket sales. This will imply that the values of a whole group of beneficiaries are not taken into account by the market mechanism. Hence, responsiveness based on sales of rail tickets will not lead to the right level of safety being chosen.

Responsiveness in public services has consequences for both production and consumption. It means providing services in the most cost-effective manner and adjusting the delivery and provision of public services to changes in social values. If the providers of services were motivated solely to pursue the public good, then there would be no issue of providing specific incentives to motivate them. However, even though there is a good deal of evidence to suggest that public sector workers are, in part, motivated by broader social goals than purely private gain, there may still be reasons to build institutions that improve responsiveness.

There are two key dimensions of public service reform which from an intellectual view are best kept separated:

- **Vertical dimension** – the classic principal-agent problem. The main concern is that providers put in less effort than consumers would wish (shirking) or else may find ways of

diverting resources to their own ends (rent extraction). The reform problem is to reduce both of these problems.

- **Horizontal dimension** – the mix of services and the priorities attached. The issue here is making the service mix more closely correspond to what consumers’ desire.

One of the main ideas in the debate about public service reform is that service providers should be made more accountable to consumers and that this will result in services that are more responsive to consumer needs and demands. The optimal provision of public services requires the (weighted) sum of benefits created by such services to equal their cost. Responsiveness can be interpreted in this context as being a measure of how changes in demands for public services translate into service provision. At the heart of this interpretation is the notion that consumers know their own best interests. Clearly some degree of paternalism has permeated the traditional view of public services with educators and health professionals being given primary authority for the design of public services. A key issue in the debate about customer focus is just how far the wishes of consumers are to be weighted.

First, they may have different tastes than do some of the beneficiaries. For example, health care professionals may attach greater weight to the health effects of intervention whereas the consumers may care also about the conditions of the waiting rooms in which they wait or better ‘hotel’ benefits for in-patient hospital stays. A system of provision that is responsive to consumer wishes will respond to such concerns. Second, some of the providers may be partially self-interested and hence fail to provide services in the most efficient way. Most contracts for construction workers build in incentives for timely completion and to provide high quality. This is because we expect construction workers to be mostly motivated by private gain. To the extent that this is true of public sector workers, we would expect incentive arrangements to be put in place.

The Role of the State

Three main roles for the state in the provision of public services can be differentiated:

- **Finance:** the state can choose to use public finance (through the tax system) to pay for public services. The main alternative would be private finance or some kind of system of user charges.
- **Provision:** The state may choose to contract for the provision of public services through the private sector or directly provide the service. This is like the make-or-buy decision discussed further in Oliver Hart’s contribution to this seminar.
- **Regulation:** The state could choose to regulate private provision without engaging in any kind of direct public provision or financing.

To illustrate the distinction between state finance and provision, consider the case of health. It is the key distinction between the National Health Service and the continental European social insurance schemes. The NHS provides health care publicly and finances it through the public purse. A pure social insurance scheme uses public finance, but relies on private provision of hospitals and doctors’ services.

The UK has traditionally used a system of public finance and public provision of services. The United States has toyed with use of workplace mandates – demanding that employers provide public services to their workers as part of their employment contracts. In terms of the above outline, this is the regulatory solution.

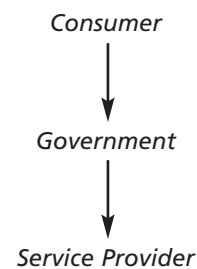
Overview

There are three key actors to consider in the delivery and financing of public services:

- Government.
- Service providers (e.g. schools, hospitals).
- Consumers.

There are a variety of ways of organizing interactions between them. The traditional model in the UK is depicted in Figure 1. The key feature of hierarchical design is that customers get very little ability to influence the design and organization of public services except through their influence over government policy. This traditional model of service provision in the UK has given very little power to beneficiaries. There has been limited choice and little direct input for consumers/beneficiaries. The main choices are confined to periodic elections where there are few salient issues. The system has fostered a generation or more of poorly informed consumers. Monitoring of service provision has mostly been left to bodies with limited sanctions such as the press, the Audit Commission and watchdogs such as OFSTED. In terms of principal agent theory, the traditional conception is that the beneficiaries/consumers are the ultimate principals — they elect the Government that designs policies which affect providers. From the point of view of the providers, the government is the only principal with which they deal.

Figure 1: Hierarchical Design



The new view of public services is to use greater customer focus. This is depicted in Figure 2. The idea is to give the consumers greater influence over policy. This may involve introducing some elements of choice and/or a reformed governance process to give consumers a direct say in how the service is provided.

Figure 2: Customer Focus



Two key notions in debates about public service reform are the need to make government more accountable and more responsive. We begin with a definition of accountability drawn from Fearon (1999). He suggests that A is accountable to B if two conditions are met: (i) there is an understanding that A is obliged to act in some way on behalf of B; and (ii) B is empowered by some formal or informal institutional rules to sanction or reward A for her activities or performance.

Responsiveness is an equilibrium outcome which is encouraged by improved accountability. Thus, a government is more responsive

when a shift in the preferences of the electorate or a crisis in the provision of a service leads to a reform. There are good reasons to suppose that increased accountability will promote greater responsiveness as an agent is forced to respond to the interests of the principal.

We discuss below two main dimensions in which reforms can be designed to produce greater government responsiveness:

- Improving consumer choice.
- Improving governance.

Choice

As we discuss above, choice is at the heart of market provision. Even though there are reasons to doubt that unfettered market provision will lead to efficient and equitable provision, that does not mean that introducing mechanisms to enhance consumer choice are unimportant.

There is a large literature in economics that looks at how decentralized provision of public services can be used to enhance choice. The traditional model is Tiebout's (1956) model of decentralized provision of (local) public goods. He envisaged (local) public goods being provided by a variety of providers located in jurisdictions with consumers choosing to live in the jurisdiction that came closest to representing their preferences. In a fully decentralized system, this would involve choosing both a tax and service provision level. Some consumers who did not value public services at all could, in principle, opt out altogether.

This is a very stylized model. In reality, localities have a variety of services which need to be weighed up. It is rarely possible for any consumer to get exactly what they want. In Tiebout's world, taxes cleared the market for locations – moving to a higher service area means paying higher taxes and will put off those who do not value high quality provision from living there. Those who want more and better public services will be willing to pay for them. In a world where taxes are centralized, other mechanisms for chocking off demand for services will emerge. House prices may have some role to play in this. Gibbons and Machin (2002) argue persuasively that this is already happening in the UK with education.

Decentralized solutions have a tendency to accentuate inequality in public service provision. This is particularly true if there are externalities in the production of public services so high income individuals prefer to "share" their use of public services with other high income individuals. This can be a particular issue in education given the widespread evidence of peer group effects on attainment.

Choice also creates a premium for informed consumers – those who know about the quality of public services or are willing to invest in information. The system of public service provision that has dominated in the UK for more than fifty years has not been one in which beneficiaries have significant gains from being informed. This suggests the need to focus on information provision for consumers as part of any reform strategy where choice is enhanced.

However, this is not easy. The parallel with private pension provision is telling. Leaving individuals to exercise private choices on such important lifetime decisions must be put alongside the evidence of the huge extent of continuing consumer ignorance on pension issues. In general, we might also expect the importance of information in choice to accentuate further the tendency towards inequality if the higher income and more educated consumers are also those with greater knowledge. Restricting choice may be paternalistic, but it could save certain groups from making the wrong decisions. This raises philosophical issues that go beyond the economics of the situation.

The possibility that choice will result in failing public service institutions closing down also presents dilemmas for those who do not leave the institution early on. They may be a premium and

greater ease of finding alternatives for those who exercise their choice early. Whether or not it leads to better long run outcomes this process will tend to lead to short term costs to be born by those who are forced to quit rather than doing so by exercising their choice.

All of these issues imply that balancing the trade-off between choice and inequality is one of the major policy challenges in introducing choice into public service provision.

Governance

Governance refers to the rules that are used to organize accountability structures within organizations. In relation to public services, this can involve improving monitoring and providing better incentives for service providers. It can also find better ways of selecting individuals who have public service motivation and/or greater competence.

In terms of incentives, a key issue concerns the kinds of sanctions that can be used for poor performance. The obvious analogy is with democratic governance with threat of replacement if performance is poor. However, for this to be an effective mechanism, it needs to be the case that there is something in the system to motivate office holders. We would expect this to be some combination of rents/returns to reputations/career concerns. Often this leaves the nature of the performance standard implicit rather than explicit. Even when performance standards are not implicit, many dimensions of performance are hard to measure.

The right principal-agent model to be used in studying the public sector is not altogether clear given the multiple tasks that agents are asked to perform (many of which are hard to measure) and the fact that they can face incentives from multiple principals (some of whom may have different objectives). The multi-task literature emphasizes why focusing only on measurable aspects of performance can be counter-productive when the non-measurable components are important. This is an important lesson in considering the achievements of regulation by performance indicators and league tables.

A well-designed governance structure may help to alleviate the fact that many consumers are ignorant about the nature and quality of public services. If the more informed consumers can represent the interests of the less informed, then giving the former the ability to monitor quality can be a substitute for choice.

The role of trustees in pension fund governance gives a model that may have wider applicability in public services. However, to be an effective responsiveness mechanism, it has to be the case that trustees credibly represent beneficiary interests. One possibility is to create explicit accountability mechanisms that try to forge this link. One example would be the greater use of directly elected representatives in governance structures.

Above we discussed how choice and decentralization may go together. Decentralization may also play a role in governance. A good example is the use of yardstick competition – comparisons across service providers as a means of improving accountability. However, it is now well appreciated that the theoretical merits of such comparisons are, by no means, straightforward when the incentives faced by those who are being compared are fuzzy or incomplete – see, for example, Dewatripont, Jewitt and Tirole (1999).

It is clear that reforms which improve responsiveness by changing governance are a real option. To be effective, they are mostly likely to require some form of direct accountability. A scheme in which a publicly elected official, first and foremost a constituency MP, takes primary responsibility provides a very weak link with interests of beneficiaries. It is clear that this line of discussion quickly leads to fundamental questions about the structure of democratic institutions and the role of beneficiaries in making social decisions.

The UK has minimal schemes for direct accountability, although the drift through regional decentralization is clearly towards a structure with more potential for the popular exercise of influence. However, the possibilities for strengthening direct accountability are many.

References

Dewatripont, M., Jewitt, I. and Tirole, J. (1999), "The Economics of Career Concerns" *Review of Economic Studies*, 66(1):189-217.

Disney, R. and Gosling, A. (1998), "Does it pay to work in the public sector?" *Fiscal Studies*, 19:347-374.

Fearon, J. (1999), "Electoral Accountability and the Control of Politicians: Selecting Good Types versus Sanctioning Poor Performance" in Przeworski, A., Stokes, S. and Manin, B., eds, *Democracy, Accountability and Representation*, Cambridge University Press.

Gibbons, S. and Machin, S. (2001), *Valuing Primary Schools*, DP 15, Centre for Economics of Education, London School of Economics.

Tiebout, C. (1956), "A Pure Theory of Local Expenditures" *Journal of Political Economy*, 64(5):416-24.

Incomplete Contracts and Public Ownership: Remarks, and an Application to Public-Private Partnerships

Oliver Hart*

Professor of Economics
Harvard University

Introduction

The question of what should determine the boundaries between public and private firms in an advanced capitalist economy is a highly topical one. In this paper I will discuss some recent theoretical thinking on this issue. The paper is divided into two parts. First, I will make some general remarks about the relationship between the theoretical literature on privatisation and incomplete contracting theories of the firm. Second, I will use some of the ideas from this literature to develop a very preliminary model of public-private partnerships.

Parallels between Theories of the Firm and of Privatisation

Let me begin by discussing the very close parallel between the theory of the firm and the theory of privatization.¹ In the vertical integration literature one considers two firms, A and B. A might be a car manufacturer and B might supply car-body parts. Suppose that there is some reason for A and B to have a long-term relationship (e.g. A or B must make a relationship-specific investment). Then there are two principal ways in which this relationship can be conducted. A and B can have an arms-length contract, but remain as independent firms; or A and B can merge and carry out the transaction within a single firm. The analogous question in the privatisation literature is the following. Suppose A represents the government and B represents a firm supplying the government or society with some service. B could be an electricity company (supplying consumers) or a prison (incarcerating criminals). Then again, there are two principal ways in which this relationship can be conducted. A and B can have a contract, with B remaining as a private firm, or the government can buy (nationalise) B.

There are, of course, some important differences between the two situations. First, if B is an electricity company, it will likely have direct dealings with consumers, independent of its relationship with the government. In this case the contract the government has with a private electricity company can be thought of as an attempt to regulate the company's dealings with consumers. There is no obvious analogy in the case of vertical integration. Second, decisions to privatize or nationalize are often highly political, presumably because of the government's unique position in society, whereas vertical integration decisions are usually strictly economic. Third, the government is often thought of as a very different agent from a private firm: it is concerned with social welfare rather than just profit. Here, however, the distinction is less sharp than it might seem at first sight since there are a number of firms (particularly nonprofits or cooperatives) that have broader concerns than just profits.

In spite of these differences, the issues of vertical integration and privatisation have much more in common than not. Both are concerned with whether it is better to regulate a relationship via an

arms-length contract or via a transfer of ownership. Given this, one might have expected the literatures to have developed along similar lines. However, this is not so. Whereas much of the recent literature on the theory of the firm takes an "incomplete" contracting perspective, in which inefficiencies arise because it is hard to foresee and contract about the uncertain future, much of the privatisation literature has taken a "complete" contracting perspective, in which imperfections arise solely because of moral hazard or asymmetric information.

My own view is that this is unfortunate. One of the insights of the recent literature on the firm is that, if the only imperfections are those arising from moral hazard or asymmetric information, organizational form – including ownership and firm boundaries – does not matter: an owner has no special power or rights since everything is specified in an initial contract (at least among the things that can ever be specified). In contrast, ownership does matter when contracts are incomplete: the owner of an asset or firm can then make all decisions concerning the asset or firm that are not included in an initial contract (the owner has "residual control rights").

Applying this insight to the privatisation context yields the conclusion that in a complete contracting world the government does not need to own a firm to control its behaviour: any goals – economic or otherwise – can be achieved via a detailed initial contract. However, if contracts are incomplete, as they are in practice, there is a case for the government to own an electricity company or prison since ownership gives the government special powers in the form of residual control rights.

Even if this position is accepted, it does not follow that one can take an "off the shelf" model from the theory of the firm literature and apply it to privatisation. In the standard "property rights" model found in that literature, ownership serves to elicit appropriate *ex ante* investments, particularly those in human capital.² If firm A acquires firm B, then A, having more residual control rights, has greater bargaining power when uncontracted – for contingencies arise; A earns a greater return on her investment and therefore invests more. Conversely, B's incentive to invest falls since B's bargaining power is lower. The optimal allocation of ownership trades off these two effects.

Applying this logic to the privatisation context, one concludes that, if the government buys an electricity company or prison, the benefit is that some government bureaucrat who is in charge of the prison will invest more (have more ideas, be more entrepreneurial); but the cost is that the manager of the prison – who used to be an owner but is now an employee – will invest less. The latter effect – that a government employee will be less entrepreneurial than a government employee – seems very plausible, but the idea that government ownership leads to more entrepreneurship by bureaucrats seems less so.

For this reason the literature has explored other trade-offs.³ Consider, for example, the model in Hart, Shleifer and Vishny

* I am grateful to the National Science Foundation through the National Bureau of Economic Research for financial support, and to Andrei Shleifer for helpful comments. This paper will also appear in the Economic Journal conference volume 2003 (forthcoming).

1 For an excellent recent summary of thinking about privatisation, see Shleifer (1998).

2 For a summary of the property rights literature, see Hart (1995).

3 For some representative contributions, see Schmidt (1996), Laffont and Tirole (1993), Besley and Ghatak (2001), and King and Pitchford (2001).

(1997) (HSV). HSV compare two cases. The government can own a facility, a prison, say, and employ a manager to run it; or, alternatively, the government can contract with a company owned by the prison manager to run the prison for a period of time. HSV ignore investments on the government side, but suppose that the prison manager can make two kinds of investment. He can invest in efficiency-enhancing ideas that raise the quality of prison services, e.g. develop new rehabilitation programs; he can also spend time figuring out how to cut costs and quality, while staying within the letter of the contract. A government employee has little incentive to engage in either activity since it is easy for the government (as possessor of residual control rights) to “hold up” the employee without rewarding him appropriately. In contrast, a private prison owner-manager is less subject to hold up. The good news about this is that private ownership encourages the first, innovative type of investment. The bad news is that private ownership also encourages the second, quality-shading kind of investment. The choice between public and private ownership depends on which of these effects is more important.

In summary, the HSV model differs from the property rights theory of the firm in two ways. First, only one party (the prison manager) invests, but he makes two kinds of investments (as in the multi-tasking model of Holmstrom-Milgrom (1991)). Second, the contract between the government and the prison provider plays a crucial role – it defines the extent to which quality shading can occur. In contrast, in the property rights model, long-term contracts are assumed to be sufficiently incomplete to be useless.

Public-Private Partnerships (PPPs)

In this section, I use an HSV-type model to understand the costs and benefits of PPPs.⁴ To repeat what was said in the introduction, this model is extremely preliminary. I will take the defining property of a PPP to be that facility construction and service provision are bundled, i.e., in the case of a prison the government contracts with one party – henceforth known as the “builder” – to build and run the prison (the builder may then subcontract with someone else to run the prison). In contrast, under “conventional” provision, the government contracts with the builder to build the prison and then later on with another party to run it. For simplicity, in order to focus on the bundling issue, I will ignore the choice between public and private ownership and assume that all provision is private.

There are three dates (0, 1, 2) as in Figure 1.

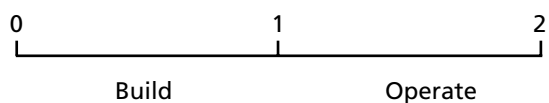


Figure 1

The government and builder contract at date 0, the prison is built between dates 0 and 1, and the prison is operated between dates 1 and 2. The contract specifies either the basic characteristics of the prison that should be delivered at date 1 (in the case of conventional provision) or the basic prison services that should be provided between dates 1 and 2 (in the case of a PPP). In each case the contract is assumed to be incomplete in the sense that the builder can modify the nature of the prison or the nature of prison services in various ways, without violating the contract. Specifically, the builder can make two investments, i and e , that have consequences for the costs and benefits of running the prison between dates 1 and 2. We write:

$$B = B_0 + \beta(i) - b(e)$$

$$C = C_0 - \gamma(i) - c(e)$$

where $\beta, b, \gamma, c > 0, \beta' > 0, b' > 0, \gamma' > 0, c' > 0, \beta'' < 0, b'' > 0, \gamma'' < 0, c'' < 0$. Here B represents the (unverifiable) benefit to society (measured in money) from running the prison and C represents the costs from doing so. The builder's total investment costs equal $i + e$.

We can interpret these investments as follows. Investment i is a productive investment that makes the prison more attractive and easier to run (it raises B and reduces C) – it could correspond to a higher quality / more pleasant / airier building. In contrast, e is an unproductive investment that reduces costs and quality. It corresponds to the quality-shading investment in HSV. For example, in the process of building the prison, the builder might realize that he can install an electric fence that reduces the likelihood of escapes. This reduces prison operating costs since fewer guards have to be hired, but may not be what the government had in mind (it reduces quality).

In the first-best, i and e are chosen to maximize net benefit ($B - C - i - e$), i.e.:

$$B_0 + \beta(i) - b(e) - C_0 + \gamma(i) + c(e) - i - e$$

The first-order conditions are:

$$\beta'(i^*) + \gamma'(i^*) = 1,$$

$$c'(e^*) - b'(e^*) \leq 1 \text{ with equality if } e^* > 0$$

We will assume that $c'(0) - b'(0) \leq 1$, (i.e. e is socially unproductive). This means that we get a corner solution in the first-best: $e^* = 0$. (In contrast, all other first-order conditions are supposed to have interior solutions.)

We now consider the second-best, where the builder's investments are non-verifiable and hence can't be contracted on. We assume, however, that the provider of prison services observes i and e ; in particular, he knows what his costs will be. We compare two cases.

Case 1: Separate contracts to build and operate, or “unbundling”

In this case, the government contracts with a builder at date 0 to build a basic prison for price P_0 . At date 1 the government auctions off the contract to operate the prison. If there is a competitive supply of contractors, the government will pay the prison operator a price equal to his operating cost $C = C_0 - \gamma(i) - c(e)$, where \hat{i} and \hat{e} are the builder's equilibrium choices of i and e . At date 0, the builder chooses i and e to solve:

$$\text{Max } P_0 - i - e$$

The solution is extremely simple: $\hat{i} = \hat{e} = 0$. That is, the builder builds the cheapest prison possible (while staying within the contract). Note that, although i and e affect the operating contract price the government has to pay, the builder does not internalize this externality.⁵ If there is a competitive supply of builders, $P_0 = \hat{i} + \hat{e} = 0$, and so the government's net payoff is:

$$B - C - P_0 = B - C - i - e \text{ evaluated at } \hat{i} = \hat{e} = 0.$$

Case 2: PPP, or “bundling”

In this case the government offers a contract at date 0 that specifies the basic quality of the service to be provided between dates 1 and 2 and a price P . Now the builder internalizes the cost of service provision since either he provides the service himself or he subcontracts the service (in the latter case, given competition, he will pay the subcontractor a price equal to the subcontractor's cost, $C = C_0 - \gamma(i) - c(e)$).

4 For related work, see Bentz et al (2002) and Bennett and Iossa (2002). Bennett and Iossa's model is based on incomplete contracts and is similar in a number of ways to the model developed here. In contrast, Bentz et al's model emphasizes asymmetric information.

5 One can imagine that one way for the externality to be internalised is for the builder's payment P_0 to be made contingent on the price the government pays to the prison operator at date 1. One problem with such an arrangement is that this price may not be verifiable. In any event, we will see shortly that a PPP achieves a similar outcome.

At date 0 the builder chooses i and e to solve:

$$\text{Max } P - C - i - e = P - C_0 + \gamma(i) + c(e) - i - e$$

The first order conditions are:

$$\begin{aligned} \gamma'(\hat{i}) &= 1 \\ c'(\hat{e}) &= 1 \end{aligned}$$

If there is a competitive supply of builders, $P = C + i + e$, and so the government's net payoff is :

$$B - P = B - C - i - e, \text{ now evaluated at } i = \hat{i}, e = \hat{e}$$

The PPP equilibrium is illustrated in Figure 2 along with the first-best and the unbundling equilibrium.

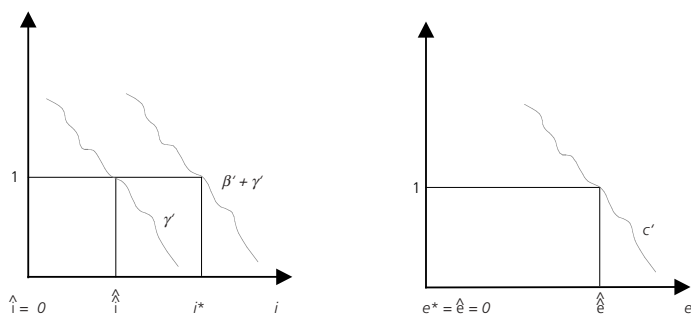


Figure 2

The trade-off between unbundling and bundling is simple. Under unbundling, the builder internalises neither the social benefit B nor the operating cost C . By setting $i = e = 0$, he does too little of the productive investment, i , but the right amount of the unproductive investment, e . In contrast, under bundling or PPP, the builder again does not internalize B , but does internalize C . As a result, he does more of the productive investment, although still too little, but also more of the unproductive investment.

The model yields a simple conclusion. PPP is good if the quality of the service can be well specified in the initial contract, whereas the quality of the building can't be. Under these conditions, underinvestment in i under conventional provision is a serious issue, while overinvestment in e under PPP is not. In contrast, conventional provision ("unbundling") is good if the quality of the building can be well-specified, but the quality of the service can't be. Under these conditions, underinvestment in i under conventional provision is not a serious issue, while overinvestment in e under PPP is.⁶

Final Comments

In this paper I have discussed some recent theoretical thinking on public versus private ownership, and I have developed a simple incomplete contracting model to analyze PPPs. The model of PPPs is far from general. For example, it ignores financing issues, which often seem to be of great concern to policymakers. In addition, the model takes the length of the contract with the service provider as given (implicitly it is assumed that the world ends at date 2); and it ignores the issue of who owns the asset (prison) at the end of the contract.

It may be interesting to extend the model to allow for these and other issues. Among other things, the model may be useful for a more general understanding of optimal contract length and

bundling issues: for example, it may throw light on the choice a purchaser in the private sector faces between hiring a general contractor to do a job and contracting separately with specialists.

References

- Bennett, J. and Iossa, E. (2002), "Building and Managing Facilities for Public Services" mimeo, Brunel University.
- Bentz, A., Grout, P. A., and Halonen, M. (2002), "Public Private Partnerships: What Should the State Buy?" mimeo, Dartmouth College.
- Besley, T. and Ghatak, M. (2001), "Government versus Private Ownership of Public Goods," *Quarterly Journal of Economics*, 116(4):1343-72.
- Grout, P. A. (1997), "The Economics of the Private Finance Initiative" *Oxford Review of Economic Policy*, 13(4):53-66.
- Hart, O. (1995), *Firms, Contracts, and Financial Structure*, Oxford, Oxford University Press.
- Hart, O., Shleifer, A., and Vishny, R. W. (1997), "The Proper Scope of Government: Theory and an Application to Prisons" *Quarterly Journal of Economics*, 112(4):1126-61.
- Holmstrom, B. and Milgrom, P. (1991), "Multi-Task Principal-Agent Analyses: Incentive Contracts, Asset Ownership and Job Design" *Journal of Law, Economics and Organization*, 7:24-52.
- King, S. and Pitchford, R. (2001), "Private or Public? A Taxonomy of Optimal Ownership and Management Regimes" mimeo, Australian National University.
- Laffont, J.-J., and Tirole, J. (1993), *A Theory of Incentives in Regulation and Procurement*, Cambridge, MIT Press.
- Schmidt, K. (1996), "The Costs and Benefits of Privatisation: An Incomplete Contracts Approach," *Journal of Law, Economics and Organization*, 12:1-24.
- Shleifer, A. (1998), "State versus Private Ownership" *Journal of Economic Perspectives*, Autumn:133-50.

6 Our conclusions about the relative desirability of PPPs are similar to those in Grout (1997).

Models of Public Service Provision: Command and Control, Networks or Quasi-Markets?

Julian Le Grand

Professor of Social Policy
London School of Economics

The Problem

The provision of public services such as health care, education, social services, housing and transport has long presented a problem for government. This is for two reasons. First, the government concerned usually has broad social objectives with respect to these services – objectives that are wider than simply the provision of an efficient service. The principal one of these is equity or fairness: society has decided that, for one reason or another, these services should be distributed equitably between its members. Equity may be defined in different ways according to the service: thus health care may be provided in such a way as to promote equality of access, equal treatment for equal need, and/or greater equality of health itself; education so as to further equality of opportunity; housing, so as to ensure a minimum standard of provision; and so on (Le Grand 1982). But some conception of equity is always there – as is the perception that whatever distribution is required by that conception would not be achieved by the random forces of the market. Hence part of the case for government intervention. Other social aims may also be important, including the promotion of social cohesion or solidarity (particularly with respect to education), and the alleviation of some of the consequences of poverty (ill-health, poor access to transport); and again there is the perception that these will not be achieved by unfettered market operations.

The second reason for government involvement in these services is their association with (sometimes massive) areas of market failure. These include externalities (health, education and – particularly – transport and housing), poor information among users and consumers (education, social care and – particularly – health), and increasing returns to scale (transport). The presence of these market failures means that, *even if* the government did not have social objectives with respect to these services other than their efficient production, there would still be a *prima facie* case for government involvement in their provision. For governments could not rely on the operation of the free, competitive market to produce a socially efficient level of service.

However, the fact that there is market failure to achieve both efficiency and other social ends does not necessarily imply successful government intervention. For governments can fail too (Wolf, 1988; Le Grand 1991). More specifically, we can identify a number of models of service delivery that governments have employed, either on their own or in conjunction with other models, to achieve their ends – each of which has its problems as a means of service delivery.

Models of Service Delivery¹

In part because of the association with market failure to achieve both efficiency and equity, many governments historically have tended to reject market mechanisms entirely in their dealings with public services. Others, especially in recent years, have adopted internal or 'quasi'-market models of service delivery in an effort to obtain some of the benefits of market competition without its costs.

The most widely known non-market model that some governments have adopted instead is generally described as 'command and control' or 'hierarchy'. This relies upon the state both funding *and* providing the service concerned. The state owns all the relevant assets, and resources are allocated through instructions given to agents through a managerial hierarchy. Classic examples are, of course, the systems of economic organisation that characterised the Soviet Union and Eastern European countries before the fall of the Berlin wall.

A second model of public service delivery, perhaps less widely known (at least by economists), is the so-called 'network' model. This is similar to command and control in some respects, but differs crucially in one key factor. Here again the state owns all the relevant facilities and assets and provides the funding, and to that extent the model is similar to command and control. But there is no direct command or instruction mechanism to allocate resources. Instead, service delivery on the ground is undertaken by qualified professionals, often operating in a network for some kind, who are essentially being trusted to make the right decisions. The emphasis throughout is on social relationships rather than on hierarchical ones or on impersonal economic exchange.

An illustration of this kind of model in action was the National Health Service between 1948 and 1991. This has been described by many authors (including the present one) as an example of the first model: command and control. However, as has been argued by Exworthy, Powell and Mohan (1999), this is misleading. For there were rather few commands and precious little control. It is better described as an example of the network model, one that relied upon politicians and civil servants to allocate resources at a macro-level and gave medical professionals almost complete clinical freedom to make ground-level decisions as to what patients should receive what treatment. The British school education system pre-1989 is another case, with local authorities allocating the funds to schools but with teachers being given extensive freedoms to determine curricula and methods.

A third model of service delivery for key public services was widely adopted in the 1990s by many governments, including the British. This is called 'quasi-markets' by economists and 'the new public management' by political scientists. Here the state still funds services, as with the other two models, but separates purchasers from providers and encourages competition between the latter for service delivery. The state finances the purchase of services, sometimes through allocating resources to centralised purchasing agencies and sometimes by giving an ear-marked budget or voucher to individual purchasers (who may or may not be the users themselves). The purchasers set the objectives. Provision is undertaken by independent providers competing with one another for formal or informal contracts for service delivery from purchasers. These providers may be for-profit or private firms, not-for-profit or voluntary organisations, or independent entities within the public sector.

It is of interest to look at these models through the lens of principal-agent theory. In each case there is a principal – the government (the politicians, the civil servants) – and a set of agents

1 The classification that follows has many parents. But key references include Williamson (1983), Ouchi (1980), and Thompson, Levacic and Mitchell (1991).

– those who deliver the service on the ground (doctors, nurses, teachers, social workers). However, the underlying assumptions concerning the motivational structure of the principal and agent and the relationships between them differ between the models.

The quasi-market model corresponds most closely to conventional principal-agent theorising. The agents are assumed to be self-interested ‘knaves’ (Le Grand 1997); that is, egoistic individuals motivated chiefly by their own financial reward. In economists’ terms, they are assumed to want to maximise their own utility function, the principal argument in which is their own income. The principal is assumed to be something closer to a public-spirited altruist – a ‘knight’ rather than a knave – who wishes to ensure the provision of an efficient and equitable public service. That is, he or she wishes to provide the service in such a way as to maximise some form of social welfare function: one that incorporates all the government’s social objectives. The relationship between the principal and agent is a contractual one, exchanging the provision of services for income. The contract has to be designed so that it will be in the agents’ self-interest to provide services in such a way as to meet the principal’s perception of the public interest.

In the command and control model, the principal again is assumed to be a knight. But for subordinates *within* the hierarchy, motivation is irrelevant. It is assumed that subordinates will simply carry out the orders and instructions that emanate from the top. Non-compliance is met with punishment. Agents are neither knaves nor knights, but something closer to pawns.²

In the network model, both principal and agent are assumed to be knights – and, moreover, knights with the same perception as to what constitutes an efficient and equitable service. Put another way, they are assumed to want to maximise identical social welfare functions. Hence there is no need for an exchange relationship of any kind between them: trust will suffice. Principals simply give agents resources and the freedom to allocate them as they wish.

Of course, none of these models correspond exactly to real modes of service delivery. In practice there are many overlapping principals and agents, within both government and social service providers. In the real world, models are mixed, with elements of all three present in most systems for service delivery. However, it is useful to distinguish between them at a conceptual level; for only thereby is it possible to examine their relative merits and demerits and thereby to ascertain which model – or which mix of models – is most appropriate for any particular service.

Which Model?

Each of these models can be – and has been – criticised on both theoretical and empirical grounds.

Command and control models have been challenged in two principal areas: their information requirements and their incentive structure. Following the work of Hayek and his followers, the information critique is well known, the incentive one less so. With respect to the latter, it is argued that the threat of external punishment drives out so-called intrinsic motivation: that is, the motivation to perform activities ‘for their own sake’ (Frey, 1999; Deci and Ryan, 1985). Staff are demotivated, morale is low, individual initiative suppressed. The quantity and quality of service provided is low, as is productivity. Systems of monitoring put in place to try to counter these effects encourage attention to be paid only to those things that are monitored (and then only sufficient to get by).

That these consequences are not simply theoretical predictions are amply demonstrated by the experiences of the Soviet Union and Eastern Europe. As one contemporary observer noted ‘the

observation that workers’ efforts and morale tend to be poor in the USSR today is familiar enough. So is the perception that the economic system itself fosters shoddy work, idleness and dishonesty’ (Hanson, 1984, p. 85).

However, this is not to say that command and control is inappropriate in all circumstances. The information problems can be overcome if the task is simple and the method of achieving it is mechanical and well known. The incentive problems can be surmounted if it is easy to monitor both the task itself and individuals’ contributions to the task – and if the punishments for non-compliance are harsh enough. It is not accidental that armies are run by command and control; in most cases they precisely fit these conditions.

Following the work of Robert Putnam (2000) and others on social capital, network models are currently fashionable; but their problems have been less well rehearsed. Their weaknesses include high set-up costs, long-term instability, the opportunities they offer for corruption and the protection they can provide for underperformance (Kirkpatrick, 1999).

The last of these – the concern about underperformance – seems particularly important in a productivity context. Again, this is part of a broader issue, relating to incentives and motivation. As noted above, the network model assumes that the providers of services are knights. But what if in fact they are knaves? In a world where the principal relationship between principal and agent is one of trust, who will protect the users of the service from opportunistic exploitation by self-interested providers? Or, even if providers are not exclusively or even largely knaves, what if they are a different kind of knight from the principal? That is, suppose they have a different conception of the public interest (as, for, instance, may be the case with faith schools)? Then, in the absence of any controlling or contractual instrument, how can the principal be sure that his or her ends will be achieved?

If motivations are indeed identical between principal and agent, however, networks do have some advantages over command and control models (and indeed over quasi-market ones) in certain situations. In particular if outputs are difficult to monitor and individual contributions difficult to assess, then both the other two models face difficulties that the network model does not.

The benefits and costs associated with the quasi-market model have been more extensively studied than the others, at least in recent years (Le Grand and Bartlett, 1993, and many others). Again it is apparent that conditions have to be fulfilled if it is to work at its most effective. These include the existence of a competitive market, with appropriate opportunities for entrance and exit, good information, low transactions costs, limited opportunities for cream-skimming, and, again, appropriate motivational and incentive structures.

With respect to the last, as with command and control, there is a question of extrinsic motivation driving out intrinsic motivation. That is, does the introduction of financial incentives motivate agents to provide more of the service concerned – or does it, by devaluing their intrinsic contribution, reduce their supply? This was an argument first put forward by Titmuss (1971) in the context of blood donation, where he argued that to replace the system of donating blood by a system of payment would result in a diminution of both the quantity and quality of blood supplied. The argument has been extensively developed and applied by Frey (1999), who has surveyed the empirical evidence on what he terms the crowding-in (extrinsic motivation reinforces intrinsic motivation) and crowding out (extrinsic motivation suppresses intrinsic effects). He concludes that both effects exist, and that, on occasion, they will dominate the more normal pattern of incentives.

² An alternative interpretation is that the agents are knaves, and the relationship is a contractual one. However, unlike in the quasi-market case, the contract is between employers and employees, where time, not a specific service, is exchanged for income. And the principal motivational tools are fear of punishment, such as job insecurity, not service-specific financial reward as in the quasi-market model: sticks rather than carrots.

The British Experiment

In the 1990s in Britain an experiment was tried, the results of which shed some light on the discussions above, especially with respect to the impact of the models on productivity. This was the replacement of models containing elements of networks and command and control by quasi-market models in health care (in 1991) and education (1989). Subsequently (post-1997) the quasi-market was partly rolled back in health, but not in education.

To begin with productivity in health care. A recent review compared changes in a (cost-weighted) activity index for hospitals with changes in the volume of resources going into the NHS (Le Grand, 2002). Dividing the activity by the resource volume index gives an indication of the units of activity achieved per unit of resource: an indicator of the efficiency of resource use. There is an interesting pattern. 1991-92, the first year of the operation of the quasi-market, showed a significant upswing in efficiency, a pattern matched in three out of the four subsequent years. But in 1996-97 the efficiency gains seemed to peter out and in the subsequent three years efficiency actually fell - at an ever-increasing rate.

Moreover, there is other evidence that the efficiency in the way resources are used within the NHS appears to be falling. A recent study by the Office of National Statistics using a more comprehensive index for the NHS than the cost weighted activity index - in that it included the activities of GPs - found that while the index had increased by 15.3% from 1995 to 2000, resources had increased by 25.3% over the same period, suggesting a significant fall in productivity.³ On a more disaggregated level, unpublished data from the University of Birmingham shows a fall in the number of Finished Consultant Episodes (FCEs) undertaken per consultant in key specialities from 1994-95 to 2000-01. They include a fall of 19% for general surgery, 10% for urology, 22% in trauma and orthopaedics, 20% in ear, nose and throat, and 8% in plastic surgery. Only in ophthalmology was there an increase and then only of 5%⁴.

Now these measures all have their problems. They do not measure all health service activities; and, more fundamentally, they only measure activity, not the 'outcomes' of that activity (such as patients' satisfaction or their health and welfare gains). Putting the latter point another way, the measures make no adjustment for quality improvements either in the process of treatment or in its outcome; and a measure which did properly take account of those (which would not be easy to find) might tell a different story. Moreover, it would be unfair either to attribute the apparent productivity gains in the early 1990s directly to the internal market, or the subsequent falls to the market's abandonment; it was still operating, after all, in 1996-7 and indeed in 1997-8. But what is clear is that some things did change for the better during the quasi-market period and that whatever efficiency gains were made in the mid-1990s are now in danger of being lost.

The story is somewhat different in education. Glennerster (2002) has used the performance of pupils in the relatively new national attainment tests (tests taken by all pupils at age 7, 11, 14 and 16) to assess the impact of the quasi-market. His conclusions are striking. First, the percentage of pupils reaching a given level of achievement

in England steadily increased from 1995 (the first full year of tests) to 2001, especially at the end of primary school. Some of the improvements were remarkable; for instance, the percentage of pupils gaining the expected level of competence in maths at the end of primary school moved from 45% in 1995 to 70% in 2001. This should be set against the fact that the best available evidence suggests that there was no improvement in the maths skills of children in the early years of secondary school for thirty years prior to 1995. Second, these improvements were not confined to good schools; in fact, the lowest performing schools in 1995 were the ones to show the greatest improvement by 2001. The same is true of schools ranked according to the wealth of the area; over the period schools in poor areas were catching up with schools from rich ones.

Again, caution has to be used when interpreting these results. There is anecdotal evidence of teaching to the test and of fraudulent manipulation of the test results. Further, recent unpublished work at the University of Durham suggests that different tests may yield rather different conclusions.⁵ However, the improvements are so great that it is unlikely that they can all be explained away by fraud of one kind or another.

A more substantive issue concerns attribution. Were these improvements due to the quasi-market, or to factors such as an increase in resources, the regulatory agency OFSTED, or some of the other changes introduced by the governments concerned (such as the Labour Government's numeracy and literacy hour).

Glennerster examines the issue of resources and points out that public expenditure per pupil in England increased only a little over the period at primary school level, and was virtually static at secondary school; public education spending as a whole actually fell as a percentage of GDP. So there were actually not only improvements in outcomes over the period but, since these were achieved with little, if any, increase in resources, there was also an increase in the efficiency with which those resources were used.

The question as to whether these improvements in both outcomes and efficiency were attributable to the operations of the quasi-market or to other policy initiatives is more difficult to resolve. However, more micro-studies of the performance of schools suggest that efficiency improvements can be directly attributed to the competitive pressures arising from the operation of the quasi-market (Bradley, Jones and Millington, 2001).

Conclusion

Although the evidence has to be treated with care, the experiments in health care and education in Britain in the 1990s suggest that, among the alternative models of public service delivery, the quasi-market may be, if not ideal, then the least worst. It seems as though competitive pressures, perhaps coupled with a judicious use of some command and control, can bring about significant improvements in public sector productivity. At the least, it would seem that further experimentation in this direction would be desirable.

3 Pritchard (2002), supplemented by personal communication.

4 Michael Harley, Inter-Authority Comparisons and Consultancy, University of Birmingham. Personal communication.

5 See <http://cem.dur.ac.uk/pips/Standards.OverTime.asp>

References

- Bradley, S, Johnes, G. and Millington, J. (2001), "School choice, competition and the efficiency of secondary schools in England" *European Journal of Operational Research*, 135: 527-544.
- Deci, E. L. and Ryan, R. M. (1985), *Intrinsic Motivation and Self-determination in Human Behaviour*, New York, Plenum Press.
- Exworthy, M., Powell, M. and Mohan, J. (1999), "The NHS: quasi-market, quasi-hierarchy, and quasi-network?" *Public Money and Management*, October-November: 15-22.
- Frey, B. (1999), *Economics as a Science of Human Behaviour*, Dordrecht, Kluwer.
- Glennerster, H. (2002), "United Kingdom Education 1997-2001" *Oxford Review of Economic Policy*, 18:120-136.
- Kirkpatrick, I. (1999), "The worst of both worlds? Public services without markets or bureaucracy" *Public Money and Management* (October-December): 7-14.
- Hanson, P. (1984), "The Novosibirsk Report: Comment" *Survey*, Spring:83-37.
- Le Grand, J. (1982), *The Strategy of Equality*, London, Allen and Unwin.
- Le Grand, J. (1991), "The theory of government failure" *British Journal of Political Science*, 21: 423-442.
- Le Grand, J. (1997), "Knights, knaves or pawns? Human behaviour and social policy" *Journal of Social Policy*, 26:149-169.
- Le Grand, J. and Bartlett, W. eds. (1993), *Quasi-Markets and Social Policy*, Houndmills, Macmillan.
- Le Grand, J. (2002), "The Labour Government and the National Health Service" *Oxford Review of Economic Policy*, 18:137-153.
- Ouchi, W. (1980) "Markets, bureaucracies and clans" *Administrative Science Quarterly*, 25:129-141.
- Pritchard, A. (2002) 'Measuring productivity change in the provision of public services' *Economic Trends*, 582: 20-32, May.
- Putnam, R. (2000) *Bowling Alone: the Collapse and Revival of American Community*, New York, Simon and Schuster.
- Thompson, G., Levacic, F. and Mitchell, J, eds (1991) *Markets, Hierarchies, and Networks: the Co-ordination of Social Life*, London, Sage.
- Titmuss, R. (1971) *The Gift Relationship*, London, Allen and Unwin.
- Williamson, O. (1983) *Markets and Hierarchies*, New York, Free Press.
- Wolf, C. (1988) *Markets or Governments: Choosing between Imperfect Alternatives*, Cambridge, MIT.

The Response of the Los Angeles Police Department to Increased Oversight

Canice Prendergast

Professor of Economics
University of Chicago

Introduction

This paper provides some data on the response of Los Angeles police officers to increased oversight since 1998, and are provided to illustrate some agency issues in public sector settings.

Public officials are rarely offered the kinds of pay-for-performance mechanisms that pervade the agency theory literature. Instead, the primary way of overseeing their behaviour is by investigating the details of the cases that they handle, where their careers may be affected by the quality of their decisions. But this type of oversight is rarely random: instead, agencies typically look for signs of error (police brutality, medical malpractice, racism by immigration officials, etc.) before they intervene.

To get such relevant information, institutions often rely on the information of consumers to identify if their employees are performing adequately. (For instance, most public agencies have complaints procedures, which the public can use.) But consumers are typically not disinterested parties, and in many public sector settings, they often have opposed interests to the institution. For example, suspects in criminal cases and potential immigrants often have very different interests to society. This is because many goods allocated by the public sector are benefits (citizenship, unemployment insurance, and so on), which consumers would like to receive even if not warranted. This implies that consumers are often unwilling to point out known errors. For example, guilty suspects have no reason to point out a police officer's error in not arresting him, nor do unqualified immigrants own up when incorrectly allowed into the country.

Agency problems then arise for two reasons. First, if a consumer is mistakenly given rents, he will not complain. This implies that bureaucratic investigations are less precisely focused because consumers cannot be trusted to reveal that an error has been made. My main interest below is in a second problem, namely, the harmful incentives that this asymmetry implies. Bureaucrats are well aware that their performance is under the spotlight when complaints are made against them. Not surprisingly, this means that from the bureaucrat's perspective, all that matters is that there are not 'too many' complaints. This implies that she has an incentive to give customers what they want, even when it is not socially efficient, simply to avoid the possibility of a complaint. For example, a police officer could choose not to arrest someone to avoid the possibility of a wrongful arrest complaint or the possibility that she has used excessive force. This is the subject of the data below. Similarly, an immigration official could allow an unqualified candidate to enter the country rather than avoid the type of cases reported in the media in the US where officials were accused of racism. Finally, there has been a recent increase in oversight of the Internal Revenue Service, which has resulted in "a sharp roll-off in tax investigations as auditors, fearing for their bureaucratic lives, proceed timidly. Tax collectors are too worried about their jobs to be aggressive" (Star Tribune, 2000).

The Los Angeles Police Department

Recent changes have increased the ability of consumers to complain against Los Angeles Police Department officers. The increased

oversight of officers was partly caused by well-publicized media accounts of infractions by anti-gang squad officers during the mid-1990s. These concerns led to the Rampart Board of Inquiry, whose recommendations affected the policies of the Department. One of these policies was that the Department became more responsive after January 1998 to perceived infractions by increasing the penalties that officers face for malfeasance. In addition, all complaints against officers are now investigated by the Internal Affairs division of the police force. I provide some exploratory evidence here that officers are indeed avoiding such confrontations since external oversight increased. In summary, the changes appear to have caused less aggressive policing.

Among the first changes implemented after the scandal were procedures that allowed the public to fill out complaints online, and an order by the Chief of Police that every complaint against a police officer be forwarded to the Internal Affairs Division. Throughout the early to mid 1990s, the department typically investigated between 50% and 70% of consumer complaints. Since 1998, this has been 100%. Not surprisingly, this resulted in a huge increase in complaints. Data are not provided for all years for the entire force, but the department notes that 5339 complaints were made against officers (and civilian staff) in 1998 and 5280 in 1999, up from 1912 in 1997, which itself was down from a previous peak of 2359 in 1992. One might imagine that this increase in complaints after 1998 would involve more marginal infractions being brought forward than in prior years, where only the most serious problems would be worth pursuing. If this were so, a lower fraction of these new complaints should ultimately be sustained against the officer after the changes. Despite this, the fraction of complaints sustained (against the officer) increased from 50% in 1997 to 53% in 1998, which suggests not only that consumer complaints were now easier to file and investigate, but also that they were more likely to harm the officer involved.

These changes have not been well received by the officers, who now perceive themselves more at risk of suspension and losing their jobs. For example, between 1992 and 1997, an average of 13 officers per annum were removed from the force for malfeasance. In 1998, 55 officers were fired, with 44 fired in 1999. In total these changes have resulted in over 800 officers being disciplined,¹ 113 terminated, and many who have left the force rather than be investigated.

There are a variety of ways in which police behaviour could change. First, officers may simply cut down on arrests, as suspects do not complain when set free. Second, many of the disciplinary cases involve the use of excessive force. As a result, officers may become more circumspect about physically confronting suspects and in their use of force, where officers are now more likely to be disciplined for going beyond acceptable levels of restraint.

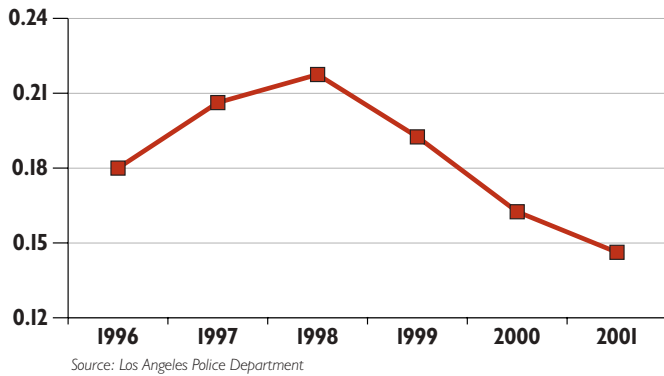
Arrest Rates

First consider the arrest record of officers. Figure 1 plots the ratio of adult arrests to crimes for the more serious Part 1 crimes² from 1996 through until 2001.

¹ This represents about 10% of all officers.

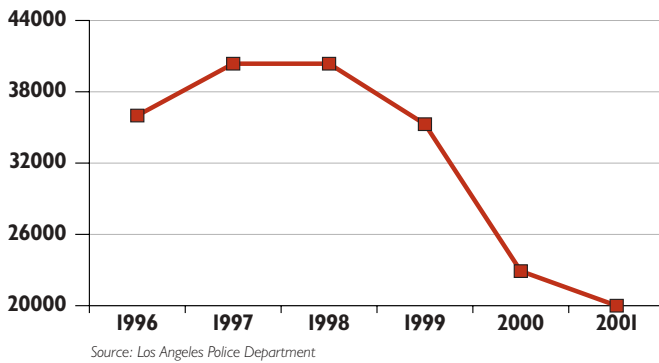
² Part 1 crimes consist of homicide, rape, robbery, aggravated assault, burglary, larceny, and auto theft.

Figure 1: Arrest rate per Part I crime



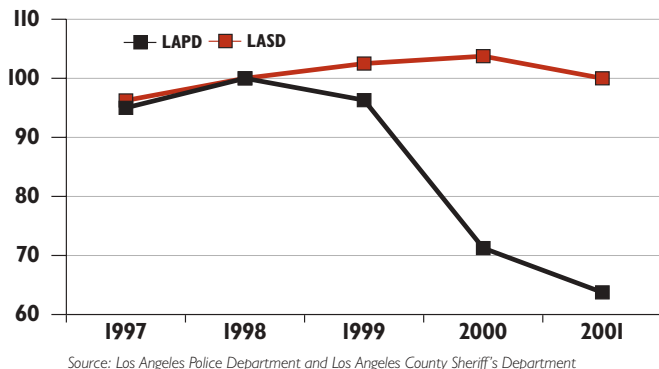
If police officers change their behaviour to avoid attention, they are especially likely to do so in cases where the victims of crime are diffuse ("victimless crimes"), and are unlikely to complain about a particular officer's actions. One instance where this is likely to arise is failure to solve narcotics crimes, where the victims are often unaware of the actions of the officer. The time series of narcotics arrests is given in Figure 2, where I obviously do not normalize by total crimes, as this number is unknown. The number of narcotics arrests shows a drop of 50% over the last three years, once again suggesting a change in how policing is carried out after the changes in oversight.

Figure 2: Narcotics Arrests



The Los Angeles Police Department (LAPD) is not the only law enforcement agency in the city of Los Angeles. The other major enforcement agency is the Los Angeles County Sheriff's Department, who have similar jurisdiction to the Police Department for LA County, but have not had the same change in oversight. I now compare the arrest records of LAPD officers with those of the Los Angeles Sheriff's Department (LASD), to identify whether the changes in arrest patterns above are specific to those officers employed by the Police Department.

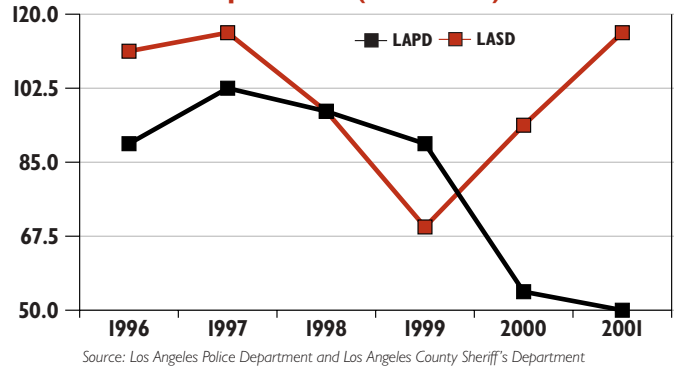
Figure 3: Arrest rates per crime – LAPD and Sheriff's Department (1998=100)



3 Another obvious measure to consider would be reports of use-of-force by police officers. The use of force per arrest fell by almost 20% in 1998, the year of the change in policy. However, it is difficult to claim that this was necessarily caused by the change in oversight, as there had been a negative trend in the propensity to use force since 1994. While it may be a response, the claim seems unproven. Unfortunately, no data on use-of-force are available after 1998.

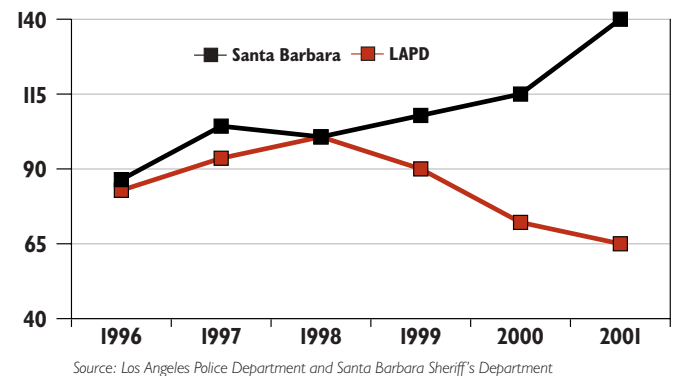
As can be seen, there has been no change in the arresting patterns of officers in the Sheriff's Department; instead, it is solely those employed by the Police Department who are arresting less. Figure 4 provides further confirming evidence on this by examining the primary victimless crime; arrests for narcotics offences. As can be seen, the arrest numbers for the Sheriff's Department are now higher than in 1998, in marked contrast to the Police Department.

Figure 4: Narcotics Arrests – LAPD and Sheriff's Department (1998=100)



As another indicator that these results are not merely reflective of changes in the crime environment in Southern California, consider how the arrest to crime rate differs between the LAPD and the Santa Barbara Sheriff's Department in Figure 5. Once again, the responses seem specific to the officers of the LAPD.

Figure 5: Arrest rates per Part I crime – LAPD and Santa Barbara Sheriff's Department (1998=100)



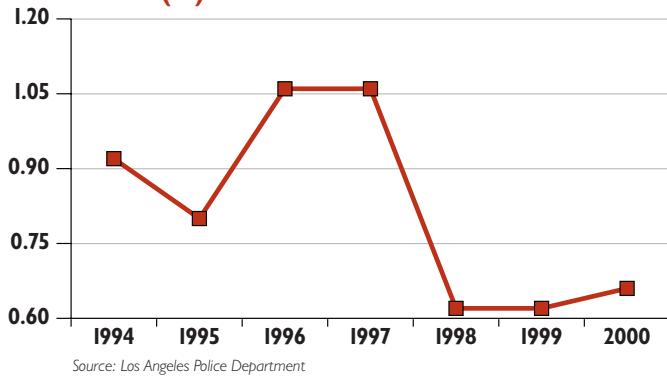
Direct Measures of Police Behaviour

As I mentioned above, many of the disciplinary cases taken against police officers involve physical altercations with suspects.³ To see how this has changed officer behaviour, first consider evidence on the frequency of officer-involved shootings: again, the paper predicts that increased oversight of officers will result in fewer shootings. As can be seen from Figure 6, shootings declined in 1998 by 45%, and remained at this level over the succeeding two years.

The premise of the paper is that increased oversight will result in officers being less likely to confront suspects. One measure, which reflects the aggressiveness of policing, is the number of assaults on officers. Consider data on the number of assaults on police officers, which reflect a more aggressive form of policing. Notably, by the end of 2001, assaults on police officers were down 30% from 1997.

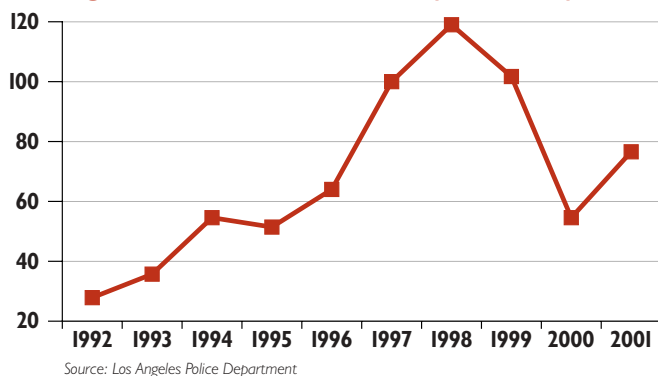
These data suggest some response by police officers, but say little about whether there is a price to pay for such reactions. Perhaps the

Figure 6: Officer-involved shootings per officer (%)



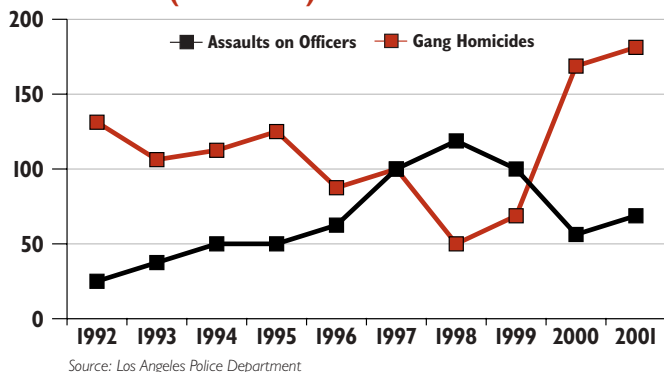
most interesting and relevant evidence comes from changes in gang activity in Los Angeles. After many years of decline, gang-related violence in Los Angeles increased significantly over the last two years; for instance, by December 2000, the homicide rate was 143% higher than at a comparable time the previous year, and drive-by shootings were up 69%. Figure 8 provides recent data on this, where I compare homicide rates to attacks on officers, from 1992 through the end of 2001. Both are normalized to 100 in the year preceding the change, 1997.

Figure 7: Attacks on officers (1997=100)



These data suggest a negative relationship between attacks on officers and gang-related homicide rates. Until 1998, gang related homicides fell as attacks on officers rose. However, since the change in oversight in 1998, assaults on officers have declined while the homicide rate has soared to a rate not seen since the mid-1990s.⁴ This change in crime rates is not specific to homicides. More generally, the violent crime rate fell almost 50% from 1992 to 1999, but has increased by almost 20% in the last two years.

Figure 8: Gang-related homicides and officer attacks (1997=100)



⁴ This change in crime rates is not specific to homicides. More generally, the violent crime rate fell almost 50% from 1992 to 1999, but has increased by almost 20% in the last two years.

Conclusion

This paper has provided some data on the response of Los Angeles police officers to increased oversight since 1998, and has described some agency issues in the public sector. It notes that although pay-for-performance mechanisms pervade the agency theory literature, in practice the primary way of overseeing public officials' behaviour is by investigating the details of the cases they handle and amending their career prospects accordingly. But this type of oversight is rarely random - agencies typically look for signs of error before they intervene. Bureaucrats consequently have an incentive to give customers what they want, even when it is not socially efficient, simply to avoid the possibility of a complaint. The data above supports this insight by indicating that, in response to increased oversight from 1998, police officers in the LAPD altered their behaviour to reduce the risk of complaints for wrongful arrest or the use of excessive force.

Is Performance Related Pay for Teachers Possible?

Peter Dolton

Professor of Economics
University of Newcastle

Introduction

Everybody knows that there are good teachers and bad teachers. Many argue that it would be desirable to reward the good ones with higher pay and induce the bad ones to either perform better or leave the profession. It has been argued that a performance related pay (PRP) system could facilitate this. The problem of course is that the output of school teachers, i.e. the education of their pupils, is multidimensional and not easy to measure in terms of either quantity or quality, nor is it easy to observe which teachers are responsible for enhanced performance. In addition, there are inherent weaknesses in any system of measurement and reward since it may induce undesirable side effects.

Stated simply, any government would be interested in introducing PRP for teachers since there is good evidence from the private sector to suggest that if workers' wages are geared explicitly to their output they put in more effort and perform more efficiently. In certain situations, with incomplete labour contracts, there is the possibility of enhancing the performance of the worker by the use of PRP, which matches rewards with outputs (see Lazear, 1995). Such schemes are efficient in the sense of providing appropriate incentives for employees but at the same time avoid the agency problems, which accompany situations in which the output or effort of the employee is not straightforward to observe.

The literature (Dixit, 2000; Burgess and Metcalf, 2000) suggests that using PRP in the public sector could induce dysfunctional behaviour in the sense that employees could direct their effort:

- (i) on some aspects of their work to the detriment of other aspects; and
- (ii) in a counterproductive way when team work or the cooperation of colleagues is involved in the production of any output.

This short position paper will not consider: the size of the present teacher shortage in the UK, the pattern of teacher pay or the detailed workings of the government's new Pay and Management Reforms. Nor will it discuss how to evaluate the empirical evidence on PRP schemes. These issues are fully discussed in Dolton et al (2002).

Why is PRP for Teachers Difficult to Implement Effectively?

(i) Education has Multiple Goals and Teaching is a Multitasking Occupation

Dixit (2000) lists the multiple goals of public education as:

- Imparting basic skills of literacy, mathematics and science for communication, reasoning, and calculation.
- Fostering the emotional and physical growth of children.
- Preparing students for work, by teaching them vocational skills and attitudes suitable for employment.
- Preparing them for life, by teaching them skills of health and financial management.

- Preparing them for society, by instilling ideals of citizenship and responsibility.
- Helping them to overcome disadvantageous circumstances at home, including in many cases poor nutrition and poor study environments.
- Providing an environment free from drugs and violence.

Dixit suggests that although these goals are not mutually contradictory they do compete for resources and teacher attention. To this degree they are alternative outputs in the educational production process and teacher effort put into one of these objectives may detract wholly, or in part, from one or more of the other goals.

A teacher's output is multidimensional. Kerr (1975) contains many examples of situations in which pay schemes have been set up in an attempt to provide an incentive for one aspect of a job and it has resulted in other areas of the job being ignored. It is well known and observed that if we reward teachers (or schools) based on the performance of the median child then this focuses effort on that child and away from the less able or most gifted. Holmstrom and Milgrom (1991) develop a model that explains the way incentive pay may not be appropriate even when accurate performance measures are available. They extend the standard principal-agent model to one in which there are several dimensions to effort. The general result is that the agent will have an incentive to divert effort away from the less accurately measured task. Hence it is shown that if the principal wishes the agent to allocate effort towards a task that is not easily measured then incentives on the measurable tasks must be weakened.

In a multitasking environment incentive contracts are likely to produce dysfunctional behaviour where workers neglect tasks, which are not rewarded. Prendergast (1999) argues that in these complex jobs PRP will normally be absent as explicit contracts are inefficient. He suggests that they will often be replaced by subjective performance evaluation that may allow for a more 'holistic' assessment of performance.

(ii) The Education System and Teachers have Multiple Principals

The actions of any individual teacher (agent) could be affected by many other people (principals) who are in a position of influence. Most specifically the wishes of parents, head teachers, governors and even pupils may influence the actions and decisions of individual teachers. Dixit (2000) describes how multiple principals would clearly fare better if they got together in advance to offer a scheme that best furthered their joint interest and bargained to split the gains between them with an agreed formula. However, Dixit (1997) shows (under regularity conditions) that the existence of several principals makes the overall incentives for the agent much weaker. This weakening of incentives occurs because each principal will seek to divert the agent's effort to his most preferred dimension. Obviously the more principals that are involved with competing interests the more diluted will be the incentive structure for the agent.

We can see how these arguments will conflict in the work of an individual teacher if a head teacher primarily wishes to see the proportion of pupils who achieve a given level on a Key Stage examination maximised, parents wish to see the potential attainment of their own child maximised, and at the same time the governor, who is a local businessman, wishes to see the emphasis on the attainment of basic literacy and numeracy skills.

At a wider level, the UK education system has numerous 'stakeholders' who can act as external principals in the agency relationship. These include:

- Pupils or students who are the recipients of the teaching/learning.
- Parents or guardians who are responsible for the welfare of the pupils.
- Individual teachers who teach in schools.
- Teacher Unions who represent the interests of their teacher members in the sense that they want higher pay and better working conditions for their members.
- Local Authorities wish to balance their spending on education with other local services and amenities in order to satisfy local residents.
- Local Education Authorities wish to be able to manage the schools in their area by being able to recruit teachers and deliver the appropriate educational outputs for a given catchment area.
- Taxpayers at the local and national level want lower public costs to reflect in a lower tax bill. This could involve the balancing of higher educational expenditure with the social problems and costs caused by low educational expenditure.
- Potential employers of students wish to see them taught the necessary skills to make them productive employees. This often means a disproportionate emphasis on problem solving, numerical and literacy skills as well as the ability to work in teams and respect authority.
- Educational, religious and ethnic pressure groups who are concerned with the content of the curriculum.
- Private schools that are in direct competition with state schools for pupils and often for government and external funding.
- Political parties with different ideological stances on the organisation and politics of educational provision and finance.
- Society as a whole wants good education to produce responsible citizens who grow up to pay their taxes and participate in society in the fullest sense of the term.

Each of these principals acts to divert the course of state education to their own ends.

(iii) Schools are not Firms: The Resourcing and Organisational Context.

Schools are not like firms in the sense that they do not exist to make financial profit. Although the 1988 Education Act introduced financial delegation to schools and this involved the introduction of 'formula funding' in which school income is based directly on pupil numbers. Popular schools were allowed to expand without limit,

and conversely unpopular schools, mostly in inner cities, to contract or even close. This approach was designed to provide teachers and schools with appropriate incentives when the education system is operating in a quasi-market. Although a school has to live within its budget this is still not the same as employees knowing that their efforts contribute to the profit 'bottom line' of a firm.

(iv) The Inputs to the Education Production Process and the Outputs of Teachers are Difficult to Observe.

The raw material, or input, a teacher works with is highly variable. It is well known that the problems of teaching the same material to children from poor homes in deprived areas is more difficult than teaching to motivated children from middle class homes. Even if one tries to measure 'value-added' in terms of improvement of exam scores these can be a distortion of the improvement in attainment as such a calculation assumes that other factors and their influence are fixed over time. There is often a huge variation in the terms and conditions of work of some teachers compared to others. Some schools are poorly resourced in terms of buildings and equipment with less efficient administrative support.

A teacher's output is difficult to observe since the educational process is complex. The implied assumption of PRP is that the educational production process is a functional relationship based on the inputs of teacher time and effort, which are easy to observe and hence reward accordingly. There are numerous problems with this:

- A teacher's output and value may not be apparent until years after they put in their performance. Often the value of what is learnt by the pupil is not apparent, used or tested until several years after they learn it.
- It assumes that the output of individual teachers can be calculated and/or that those teachers responsible for pupil improvement can be identified.
- It assumes that no other inspiring teacher (or adult) has had an outside influence on the pupil as an externality, i.e. that there is no other contribution to value added, e.g. private tutors, attendance rate, poverty, parental input.
- It is not impossible that teachers (or principals other than government) may view their output differently from the government. Teachers may want to promote curiosity, induce creative thinking, provide pastoral care and develop a wider curriculum. The government may prefer to structure the curriculum, standardise teaching methods, meet minimum standards on basic skills and maximize performance on SAT test scores. It is easy to see how a conflict could arise.

(v) 'Career Considerations' may not Operate Effectively in Teaching.

In repeated or long run relationships it may be unnecessary to induce extra effort (via a PRP) in the early stages. This is because the prospect of indirect incentives in the form of a better prospect of future rewards can suffice to induce effort on the part of workers. Hence in many organisations, career considerations alone may be effective at generating effort from young workers and those with short tenure, simply because they are working not only for pay now, but the prospects of continued employment, pay enhancement and promotion in years to come (see Dewatripont *et al*, 1999).

The implication of career considerations in teaching is that they could provide a logical alternative to a PRP or merit pay structure for teacher remuneration. This could be arranged by having the teaching profession organised as a career ladder with logical steps up the teaching hierarchy governed by promotion through internal school and external peer review. In such a structure each teacher has an incentive to take on more responsibility or seek promotion

as part of a career goal of achieving a higher salary and position within the school.

The logical flaw with such a 'career structure approach' to thinking about the teaching profession is that many teachers do not seek to progress up the teaching/school hierarchy by taking on increasingly administrative jobs. Indeed, many teachers will argue that they became teachers because they wanted to teach children and assiduously avoid the administration that would be involved with becoming a senior teacher, deputy head or head teacher. In this sense the incentive to work hard to climb the teacher career ladder is therefore not relevant for a substantial proportion of teachers.

(vi) Peer and Subjective Performance Evaluation Schemes are Difficult to Operate for Teachers.

Often judgements concerning which teacher is to gain a pay rise or promotion may be based on subjective evaluation (since there are no unambiguous measures of performance). There are numerous problems with pay schemes that are based on peer review and/or subjective evaluation:

- The process could be prone to favouritism and unfair practises.
- It may compress ratings and induce 'leniency' and 'centrality' biases.
- It could induce workers to misallocate their effort to create a favourable impression of assessors.
- Such a system would require costly evaluation, monitoring and regulation.

(vii) Teaching is a Team Activity with Group Incentives

Teachers often have to work in teams, which means in close proximity and collaboration with colleagues, hence an individual's contribution may be difficult to discern. Individual incentive pay in a team teaching environment will give incentives for:

- The allocation of difficult or low ability children to specific classes and all teachers will wish to avoid them.
- The reduction of co-operation between colleagues.
- The concentration of energy on marginal pupils to get them across national acceptable thresholds.

(viii) Teachers are Motivated Agents who are Subject to a Code of Professional Ethics

The general problem of providing incentives for effort assume that the agent gets utility solely from the money income the principal pays them, and disutility from the effort he exerts on behalf of the principal. In reality the agent may get utility directly from some aspects of the work itself. This can be the case in teaching. Many teachers do the work because they enjoy the job and derive satisfaction from doing it. In this context, it is clear that the principal can pay smaller bonus payments and still get the same level of effort, as they would do from a worker who is not so motivated and is rewarded handsomely with bonuses for extra effort. In teaching, many teachers share some idealistic, altruistic or ethical purpose that serves the agency in the sense that the main goal is the general education of the pupils and all personal and individual agenda are submerged by this goal.

Teaching as a profession is informally governed by a code of professional and ethical behaviour. The presence of asymmetric information and significant positive externalities can provide a rationale for the widespread compliance of individuals to a code of professional ethics. The individual's incentives for compliance may

be enhanced by rewards to the school of enhanced devolved or discretionary budgets as a result of good performance. These positive externalities to the school induced by a code of professional behaviour may be enough of an incentive for most teachers to work hard. Indeed the professional nature of the teaching occupation may be a substitute for an incomplete labour contract in the sense of ensuring the appropriate amount of effort from each teacher. It is suggested that a code of professional behaviour may illicit effort and appropriate behaviour through 'peer esteem' a 'common service ideal' and self-regulation and hence is a reasonable substitute for a PRP. (See Matthews (1991) and Eshel *et al* (1998)). Although complying with a professional code of ethics imposes costs on teachers, they may view this as an investment in the provision of a local (reciprocal) public good for their colleagues.

(ix) Teaching is a Heavily Unionised Activity

The position of the teacher unions in the UK is rather unique. They are simultaneously the organisation that represents the employees in pay negotiations (and conditions of work) and they are also partially responsible for the maintenance of the professional code of behaviour. This responsibility is taken differently by the different teacher unions. Some of them, or their members, will often refuse to strike or take any industrial action that may affect the pupils.

In the UK a substantial fraction of teachers are members of a teacher union but this is often not because the unions are effective at winning pay awards. Rather it is because membership of the union offers some insurance against litigation by a pupil or parents and the union will defend the member against any claim of unfair dismissal.

(x) The State Education System has a Lack of Competition

State schools in the UK, in many areas, operate essentially as a monopoly provider. Only around 7% of school children in the UK attend independent schools. Due to the scale of their fees, for most parents, these independent schools do not present a realistic alternative to state schools. It was this lack of competition that prompted the last Conservative government to publish school league tables and insist on the right of parents to send their children to any school they wished. The central idea behind the creation of a quasi-market in state education is the theory that the introduction of competition would provide the appropriate incentives to schools to become more efficient. Theoretically this, in turn, may provide incentives for teachers to improve their performance. However, this naive faith in the power of market forces must be tempered by the reality that multiple tasks and multiple agents will weaken the power of such incentive structures.

Conclusion

To summarize, the provision of state education is a multi-task, multi-principal operation. Teachers' effort is difficult to verify and reward with appropriate incentives. Most teachers are union members who work in co-operative teams in which competitive behaviour would be counter-productive. In addition teaching is, to a large extent, an occupation chosen by motivated agents who act according to a strong code of professional behaviour. In these circumstances economic theory would suggest that any PRP scheme will have only weak incentive effects.

The problem of the government is therefore to devise a payment system which:

- (1) Induces effort and increases productivity
- (2) Minimizes the need for expensive monitoring
- (3) Avoids distortionary counterproductive incentives

- (4) Encourages team work and does not induce unnecessary competitive behaviour between teachers which may result in sabotage
- (5) Retains the appropriate balance between all the various activities a teacher is expected to perform
- (6) Encourages good teachers to stay in the job and poor teachers to seek alternative careers
- (7) Provides adequate incentives to encourage new recruits to the profession and discourage early retirement
- (8) Is flexible enough to allow extra financial incentives to encourage recruits in shortage subjects in which the 'outside option' is highest without alienating other subject teachers or the teacher unions.

This is a difficult task.

References

Burgess, S. and Metcalfe, P. (1999), "Incentives in the Public Sector: A Survey of the Evidence" *CMPO Discussion Paper 99/016*, University of Bristol.

Dewatripont, M., Jewitt, I. and Tirole, J. (1999), "The economics of career concerns: Part I - Comparing information structures" *Review of Economic Studies*, 66:183-198.

Dixit, A. (1997), "Power of Incentives in Public versus Private Organisations" *American Economic Review*, 87(2):378-382.

Dixit, A. (2000), "Incentives and Organisations in the Public Sector: An Interpretative Review" mimeo, University of Princeton.

Dolton, P., McIntosh, S. and Chevalier, A. (2002), *Teacher Pay and Performance: A Review of the Literature*, Bedford Way Papers.

Eshel, I., Samuelson, L. and Shaked, A. (1998), "Altruists, Egotists and Hooligans in a Local Interaction Model" *American Economic Review*, 88(1):157-179.

Holmstrom, B. and Milgrom, P. (1991), "Multi-tasking principal-agent analyses: Linear contracts, asset ownership and job design" *Journal of Law, Economics and Organisation*, 7:24-52.

Kerr, S. (1975), "On the folly of rewarding A while hoping for B" *Academy of Management Science Journal*, 18:769-783.

Lazear, E. (1995) *Personnel Economics*, Cambridge, MIT Press.

Matthews, R. (1991), "The Economics of Professional Ethics: Should the Professions be more like Business?" *Economic Journal*, 101:737-50.

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

Value for Money, Educational Resourcing and Pupil Attainment

David Mayston

Professor of Public Sector Economics, Finance and Accountancy
University of York

Introduction

Considerable attention has been given in recent years to the concept and measurement of *value added* in primary and secondary education (see, for example, Jesson, 2002). Value added measures seek to overcome one of the main deficiencies of existing school league tables of examination results that have become a regular feature of competition between schools in a quasi-market for pupils and resources, but which fail to take into account differences in the characteristics of the pupil intake between different schools. In contrast, value added measures seek to obtain a more accurate measure of *relative school effectiveness* by systematically adjusting for variations in *pupil prior attainment* at earlier Key Stages in the educational process. However, rather less attention has been given until recently to the assessment of *value for money* and *school productivity*, in terms of the role which resources play in influencing educational outcomes. This paper outlines some of the reasons why greater attention needs to be given to the issues which are involved in this area.

Educational Resource Allocation and Resource Management

The first reason for being concerned with issues of value for money and the productivity of resources in influencing educational outcomes is that productivity and value for money are two-way concepts. If productivity, as measured by a suitable ratio between educational outputs and resource inputs, is a finite number even for fully efficient schools, then inputs are indeed required to produce outputs. Such an inter-relationship can be embodied more formally through the concept of an *educational production function* that incorporates the relationship between resource inputs and characteristics of the pupil intake and the attainable levels of educational outcomes which efficient schools could achieve with these resource and pupil inputs. The microeconomic theory of duality enables us to associate with such an educational production function an *educational cost function* specifying the minimum expenditure which is required to achieve a given level of educational outcomes when facing different input prices and different characteristics of the pupil intake. Knowing what this cost function is itself becomes of some importance in the context of *decision-making* in the major educational *resource allocation* processes that regularly take place within central and local government.

At the peak of the decision-making pyramid in Figure 1 below, knowledge of this cost function could help to inform the key policy process of balancing the Government's desire for greater national levels of educational achievement with the Exchequer's willingness, and assessment of the need, to pay for improved educational outcomes. The now central decision-making, and indeed *quasi-contractual*, process of Comprehensive Spending Reviews (CSR) combined with Public Service Agreements (PSAs) between HM Treasury and individual Departments, such as the Department for Education and Skills (DfES), depends upon making such an accurate two-way assessment of the resources which are required to deliver target levels of educational outcomes, and the target levels of educational outcomes which can be achieved from the resources which are to be made available. HM Treasury's 2002 Spending Review, and indeed the more general *Modernising Government*

agenda (Cabinet Office, 1999a,b), require that such decision-making and assessments be increasingly *evidence-based*.

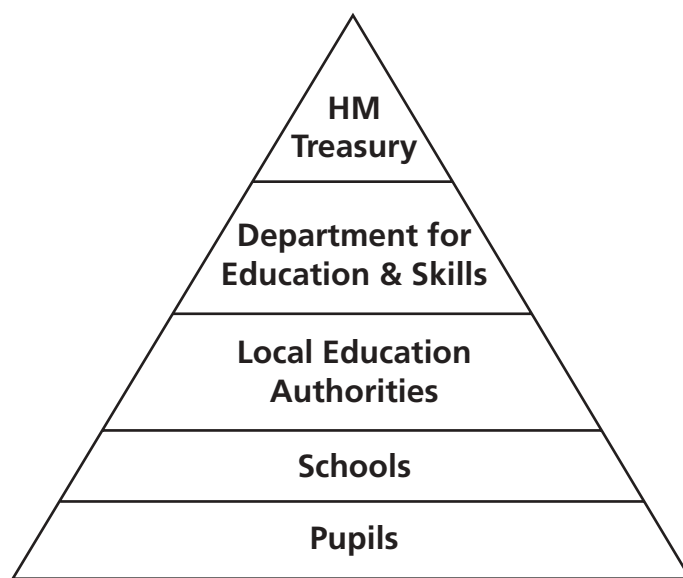


Figure 1

In the next stage of the educational resource allocation process, the devolution of resources to individual schools and pupils passes through the complexities of local government finance, in which the resourcing formulae of the educational components of the Standard Spending Assessment (SSA) of individual local authorities play a major part in determining the block grants received by local authorities from central government. However, despite the increased emphasis in recent years on educational targets for and within local education authorities (LEAs), there is no explicit consideration given in these resourcing formulae to any evidence-based relationship of a link between resources, attainable levels of educational achievement and characteristics of the local pupil intake. Under the current target that LEAs devolve 87 per cent of their Local Schools Budgets to their individual schools, and the statutory requirement that 75 per cent of these Individual School Budgets be determined on the basis of Age-Weighted Pupil Numbers (AWPN), the Standard Spending Assessments of individual local authorities are in turn likely to have a major impact upon the funding of all of its individual schools, in ways which do not necessarily relate well together, resources, target levels of attainment and local characteristics of the pupil intake. This will be particularly so when there is a significant degree of variation in the nature of the pupil intake across the LEA. Thus schools in the more affluent parts of a city such as Sheffield, which has large areas of socio-economic deprivation, are likely to receive more funding per pupil than they would *with an identical pupil intake* in a more uniformly affluent city.

The funding of individual schools within an LEA also depends critically upon the LEA's choice of the relative *age weights* which it applies across individual primary and secondary school year groups in its computation of Age Weighted Pupil Numbers. However, there is no substantial body of research evidence on the relative

educational effectiveness of different values of these key parameters on which to base this choice, or indeed on the empirical connection between the relative age weights in the school funding formula and how each school subsequently allocates its resources across the different year groups.

Decision-making on resource allocation and resource management is clearly also required *within* individual schools. This includes decisions on the extent of teaching time devoted to different individual pupil groups, on non-contact time, and on other expenditure items within school budgets, which themselves may amount to several million pounds for an individual secondary school. The main rationale for the *Local Management of Schools* initiative in 1990 and its successor *Fair Funding* regime (that have been central parts of successive Governmental educational policy) has indeed been to devolve as much decision-making as possible on resourcing and budgeting down to school level. However, as Dennison (1990) noted, "recent research activity on resource management in education has been much less than the topic's importance demands. In particular it has left institutions badly placed to cope with growing requirements upon them resulting from increased financial autonomy".

One key resourcing issue for both schools and the educational system as a whole is that of *class size* and its impact on educational performance. Several US studies based upon the Tennessee Student/Teacher Achievement Ratio (STAR) project (see e.g. Krueger, 1999) have found a positive impact on pupil performance of reduced class sizes in the early years from ages 5 to 8, although other US studies (e.g. Hoxby, 2000) claim to find no such significant link. In the UK, Blatchford and Mortimore (1994) have concluded that: "It is very unsatisfactory – given the importance of class size, and how it determines vast costs in education – how little investment we have put into research on class sizes and teaching groups, on the effects on teachers and pupils of different class sizes, and on the opportunities that might be provided".

Complementing the requirements for improved decision-making on resource allocation and resource management in education are the information requirements for improved *accountability* of schools, LEAs and indeed the DfES, in their use of resources. Requirements for local authorities to seek *best value* in the use of their resources are indeed a key part of their current regime (see e.g. DETR, 1998). *Value for money* in the use of resources by individual schools and LEAs is also a key item for examination in OFSTED and Audit Commission inspections (see OFSTED, 1995). Greater knowledge of the frontier of what was educationally achievable from given resources and pupil inputs, combined with statistical studies of the extent to which individual schools were capable of productivity increases, would add to the ability to carry out such value for money monitoring. In addition, it would enable schools themselves to carry out their own *diagnosis* of their current relative performance, given their resources and pupil characteristics, and to take steps to improve the effectiveness of their resource allocation, with less need for costly external inspections. Knowledge of the production possibility frontier would also enable decision-makers to distinguish between fat and bone, and have greater information on the resources required to achieve sustainable long-run performance targets.

Data Needs

The patchy nature of existing data sources on the deployment of resources within schools has been discussed in detail in Mayston and Jesson (1999). Thus whilst there have been earlier studies of schools' management of their finances by the Audit Commission (1993-96), these have been based upon relatively small samples of schools, in a way which does not permit a detailed matching with examination success and the many different characteristics of the pupil intake. The Headteacher HP and HS forms which schools

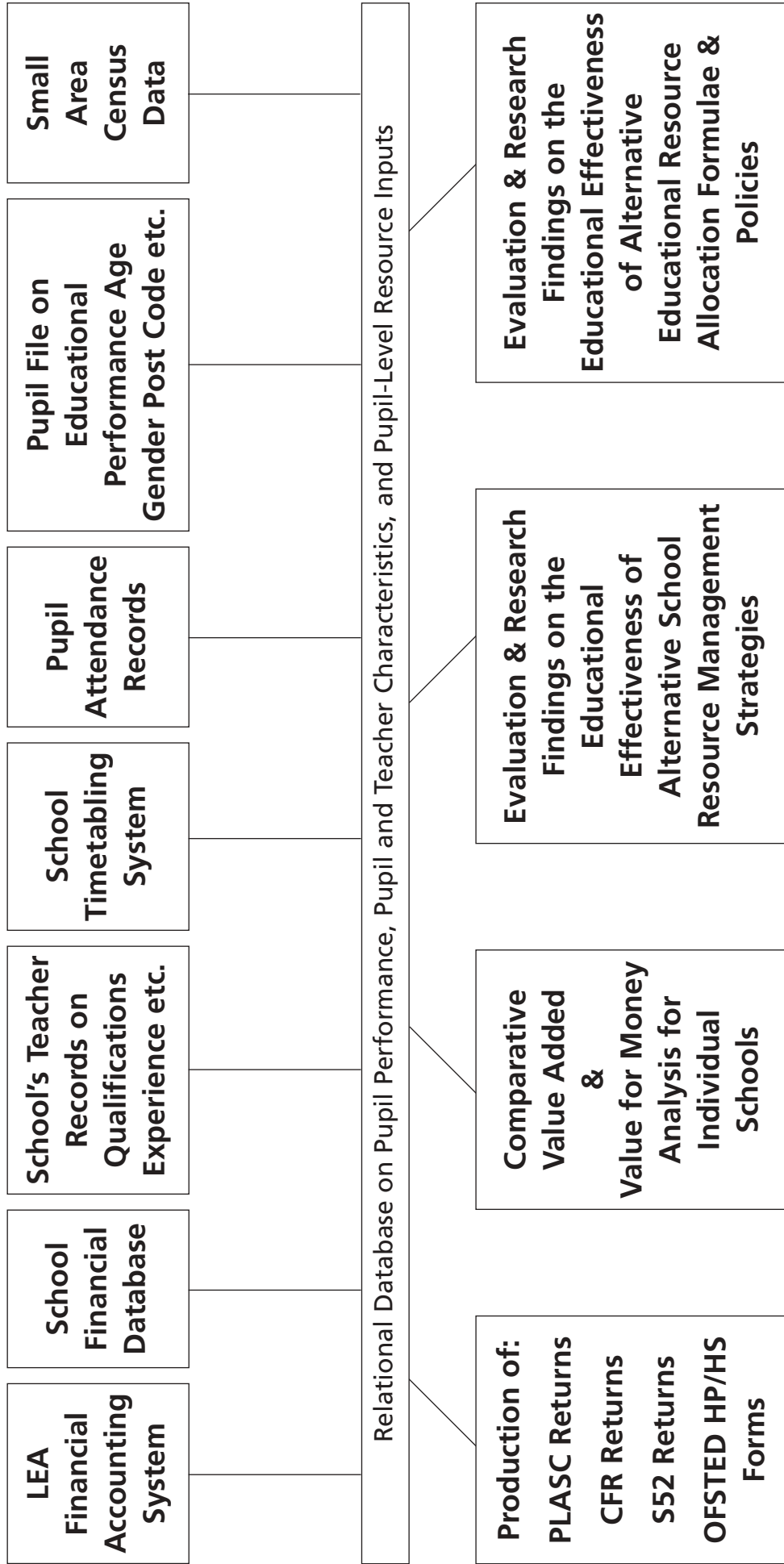
complete before OFSTED inspections have contained much detailed information on school resourcing for a large sample of schools, but unfortunately have not been available in electronic form. Data from S42/S122/S52 returns on LEA school resourcing formula and income devolved to individual schools have not for many years been available centrally or in electronic form. In recent years, the Standards Fund has distributed large sums of public money in a more targeted way than the complexities of local government finance permit. However, systematic data on which individual schools have received how much of this income, and the impact of this funding on their educational performance, given their pupil intake, have not been collected centrally. Whilst the Funding Agency for Schools (1997, 1999) carried out several useful comparative studies of resource deployment within Grant Maintained schools, based upon their consistent Rainbow Pack (FAS, 1998) for financial reporting, both the FAS and its Rainbow Pack disappeared with the later abolition of Grant Maintained schools.

Fortunately, the subsequent establishment of the Value for Money Unit inside the DfES has enabled some of this useful work to be continued. A framework for Consistent Financial Reporting (CFR) of school-level expenditures has now been established, though it is likely to take some years before all schools and LEAs are in a position to provide this information on a common basis. Section 52 returns on total school-level income and expenditure are also now becoming available electronically. The new Pupil Level Annual School Census (PLASC) has now come on stream electronically from January 2002. This will provide extensive data on school teaching staff, pupil characteristics, and class size on a selected 'snapshot' census time on the third Thursday in January each year, though curiously not on which subject pupils are studying at that time in each class. PLASC has evolved historically from the manual Form 7 Annual School Census that provided a database for answering parliamentary questions on issues such as the largest primary class size in the country. There is a strong need now to ensure that a rich database is generated which can help to meet the above wider needs for improved evidence-based decision-making and accountability, supported by high quality research and analysis of the available data. At the same time, schools are themselves increasingly making use of electronic systems of timetabling, pupil attendance records and achievement files. There now exists the potential to establish powerful *relational databases*, as in Figure 2 below, that can link together electronic data from the school's timetabling system on which pupils were receiving how much teaching time from which teachers in each different subject, linked also to administrative files on pupil performance and attendance records, teacher qualifications and experience, and through knowledge of pupil post codes to small area census data on pupil background characteristics. If well-designed, relational databases can provide, as *low-cost by-products* of routine administrative systems, both good *local management information* and inputs into *national comparative databases*. Similarly, rather than facing the burden of many *ad hoc* requests for information from schools, such school-level databases would enable schools to generate well-specified routine returns, such as those for CFR and PLASC, at low marginal cost. The piloting and implementation of a relational database of this kind for schools would indeed be a practical example of achieving the goals of both '*joined-up government*' and '*information age government*' set out in 1999 in the *Modernising Government* White Paper (Cabinet Office, 1999a).

Endogeneity Bias

A second main reason for less current information being available on the *supply-side* production frontier between attainable pupil outcomes, resource and pupil inputs is that the resource patterns and educational performance levels we actually observe result not

Figure 2: Relational Database



only from this supply-side relationship but also from demand-side factors (Mayston, 1996, 2000, 2002). These may include considerations of governmental *willingness to pay* considerations for greater educational performance that may direct more resources to those pupils and schools which have relatively low levels of educational performance. Similarly, *competition in the housing market* between parents seeking to live in the catchment areas of well-performing schools may establish an additional inter-relationship between the socio-economic background of pupils and school performance beyond that which is involved simply in the educational production function. Competition between schools in the *labour market* for teachers may also mean that better performing schools are able to recruit higher quality teachers, in a way which makes the relationship between the underlying variables not simply the input-output relationship embodied in the educational production function.

As discussed in Mayston (2000), these additional inter-relationships are likely to *bias downwards* the estimated link between resources and educational attainment when standard Ordinary Least Squares (OLS) methods of regression analysis are used to estimate the educational production function. Little reliance can then be placed upon earlier findings by Hanushek (1986) and others of no apparent significant relationship between resources and educational outcomes, based upon OLS regressions. Such endogeneity bias will also affect the reliability of the conclusions of studies of the educational production frontier based upon Data Envelopment Analysis (DEA) (see Mayston and Jesson, 1988; Orme and Smith, 1996) or upon Stochastic Frontier Analysis (Aigner et al, 1977). In order to overcome endogeneity bias, more extensive microeconomic modelling is required, as discussed in detail in Mayston (2002) and in line with the recommendations of the Performance and Innovation Unit's *Adding It Up* report (PIU, 2000).

Allocative Efficiency

Whilst educational productivity has been for many years been the subject of DEA studies (see Worthington, 2001), these have focussed on the issue of *technical efficiency* that is concerned with the extent to which efficiency gains can be achieved through *equal proportionate changes* in all controllable inputs or in all outputs. In doing so, they have neglected the assessment of price or 'allocative' efficiency, that is concerned with achieving an efficient balance in the mix of inputs and of outputs. This is despite the fact that one major long-run economic and financial trend which faces education, as a *labour intensive* public service, is that of the *relative price effect*. Rising consumer real incomes over time, through the operation of competition in both the housing market and the labour market, will make the recruitment and retention of teachers increasingly difficult unless teachers' pay increases broadly in line with the real increases achieved elsewhere. The real cost of education will then rise compared to that of less labour-intensively produced goods and services, unless substitution with other inputs is possible. This in turn raises issues that need to be explored more systematically, of the extent to which more efficient combinations of teachers, classroom assistants and ICT systems can help to produce enhanced educational outcomes out of given total school budgets.

DEA studies also tend to neglect issues of the long-run price of stress and effort in the labour market. Greater pressures on teachers to achieve higher target levels of educational output potentially reduce the *non-pecuniary benefits* of teaching and making more difficult the retention and recruitment of teachers, unless large compensatory pay increases are made. Again the interaction between stress, effort, teaching loads and support from classroom assistants and ICT equipment needs to be carefully examined if more efficient ways of producing improvements in educational attainment are to be developed which can support rising teacher pay levels within limited educational budgets.

A further important aspect of allocative efficiency which has received insufficient attention in recent years is that of the detailed mix of *educational outputs*. The use of ill-designed summary measures of aggregate educational output by which schools and the educational system as a whole are judged may indeed lead to important distortions in the nature of this output mix, and in associated school resource management strategies. One summary measure which has been prominent for several years, both in school league tables and in the national Public Service Agreement for secondary education, is that of the percentage of pupils attaining 5 or more GCSE grades A* to C. Unlike the alternative point score approach, this measure fails to satisfy the desirable design property of *positive responsiveness*, i.e. that the measure should increase in value whenever individual pupils improve their performance, such as by attaining grade As rather than grade Cs, or six GCSEs rather than five. Pressure to maximise this performance measure, such as from its use in school league tables or performance related pay, is likely to encourage a strategy of '*managing the margins*' (Mayston, 1992), whereby resources are concentrated heavily upon those pupils who are on the margin of attaining grade Cs, to the neglect of other pupils. Evidence that some schools are now engaging in such a resourcing strategy is provided in a study by Gillborn and Youdell (1999).

Whilst basic levels of numeracy and literacy have fortunately received greater attention in recent years, educational performance at higher levels in key subjects, such as mathematics, merits further investigation. The above performance measure itself encourages the use of 'restricted grade' examinations in mathematics with a lower technical content, in which the maximum feasible grade is a B but in which a grade C is perceived as more achievable by marginal candidates than under the full syllabus. At Advanced-level, the use of aggregate point scores as a school performance measure encourages a form of '*subject indifference*' in which each subject is given the same weight. Pupils of less than exceptional ability may then be discouraged from taking subjects, such as mathematics, physics and economics, that have a high technical content and in which it may be more difficult for them to achieve the same grades as in less technical subjects. At the same time, the availability of well-qualified mathematics and other specialist teachers is constrained both by the supply of university graduates in these subjects and by competing demand from elsewhere in the economy for their scarce skills.

One important feature of educational change in recent years has indeed been substantial relative reductions in the numbers of pupils taking A-levels in mathematics (which has dropped from 11.6% of A-level entries in 1990 to 8.8% in 2001), in economics (which has dropped from 6.6% of A-level entries in 1990 to 2.2% in 2001) and in physics (which has dropped from 6.6% of A-level entries in 1990 to 4.1% in 2001), at the same time as substantial increases in the number of pupils taking A-levels in less technical and less mathematical subjects, such as business studies (which has risen from 1.8% in 1990 to 4.9% in 2001 – see AQA, 2002). Concern over falling levels of technical knowledge of students who have received the same A-level grades in mathematics over the past decade has also been expressed in a long-term comparative study of the mathematical abilities of entrants into a university electronics department (Todd, 2000).

Difficulties in recruiting students with good mathematics training is already impinging on the ability of university computer science, electronics, mathematics, physical science, and economics departments to deliver well-trained graduates into the labour market, with numerous university physics and chemistry departments, and some mathematics departments, already having closed due to student recruitment problems. Higher education departments find themselves in a *demand-driven market* in which their financial survival is directly related to their ability to recruit students with acceptable A-level grades. If schools and pupils are in

turn driven by short-term pressures rather than longer-term analyses of *labour market needs*, questions of the efficiency of the detailed output of educational institutions in meeting future labour market needs are again raised. The long-term productivity of educational investments in human capital in this context is likely to depend in large part upon the availability of high quality individual careers guidance.

A detailed assessment of labour market needs is fortunately now becoming more readily available in the context of the National Skills Agenda proposed by the Skills Task Force (2001), supported by more detailed *sectoral and geographical analyses* of labour market needs (LSC, 2001). These in turn can help to inform the resourcing strategy of the new national and local Learning and Skills Councils, that from April 2002 are responsible for funding all post-16 education, other than higher education. However, the interface between such analyses and the proposed major expansion of higher education, and its own staff recruitment problems in subjects with high demand elsewhere in the economy, merits greater attention.

Conclusion

Improved decision-making and monitoring to improve the productivity with which large sums of public money are spent in primary and secondary education, whether in the context of quasi-markets and contracts or under a more command and control regime, depend crucially upon improvements in the *organisation and availability of information*. The scope for further improvements in this direction is outlined above. However, if richer data sources are to be translated into useful information and evidence for improved decision-making and monitoring, greater sophistication in *modelling and analysing* the underlying inter-relationships between pupil attainment, resource deployment and characteristics of the pupil intake is required. If the investment in education is to be fully productive, greater attention must also be given to issues of both *input and output mix*, and to the interface between existing *educational performance measures* and future *labour market needs*.

References

Aigner, D., Lovell, A.K. and Schmidt, P. (1977), "Formulation and estimation of stochastic frontier production function models" *Journal of Econometrics*, 6:21-37.

Assessment and Qualifications Alliance (2002), *GCE Interboard Statistics – Advanced Level 1990-2001*, Guildford, AQA.

Audit Commission (1993-1996), *Adding up the Sums 1-4*, London, TSO.

Blatchford, P. and Mortimore, P. (1994), "The Issue of Class Size for Young Children in Schools: What Can We Learn From Research?" *Oxford Review of Education*, 20:411-428.

Cabinet Office (1999a), *Modernising Government*, London, TSO.

Cabinet Office (1999b), *Professional Policy Making for the 21st Century*, London, Cabinet Office.

Dennison, W. (1990), "The Management of Resources: Limited Research Effort on Major Issues" in Sara, R. and Trafford, V., eds, *Research in Education Management and Policy: Retrospect and Prospect*, London, Falmer Press.

Department for the Environment, Transport and the Regions (1998), *Modern Local Government: In Touch with the People*, London, DETR.

Funding Agency for Schools (1997), *Cost and Performance Comparisons for Grant-Maintained Schools; Expenditure and Income 1995/96*, York, FAS.

Funding Agency for Schools (1998), *The Rainbow Pack*, York, FAS.

Funding Agency for Schools (1999), *Energy and Water Supplies Management; Developing a Cost Effective Approach to Energy and Water Use*, York, FAS.

Gillborn, D. and Youdell, D. (1999), *Rationing Education: Policy, Practice, Reform and Equity*, Buckingham, Open University Press.

Hanushek, E. (1986), "The economics of schooling: production and efficiency in public schools" *Journal of Economic Literature* 24:1141-1177.

Hoxby, C. (2000), "The Effects of Class Size on Student Achievement: New Evidence from Population Variation" *Quarterly Journal of Economics*, 115:1239-1285.

Jesson, D. (2002), *Value Added and the Benefits of Specialism*, London, Technology Colleges Trust.

Krueger, A. (1999), "Experimental Estimates of Education Production Functions" *Quarterly Journal of Economics*, 114:497-532.

Learning and Skills Council (2001), *Meeting Sector Skill Needs: Some Emerging LSC Approaches*, LSC Discussion Paper, London, LSC.

Mayston, D.J. (1992), *School Performance Indicators and Performance Related Pay*, London, Assistant Masters and Mistresses Association.

Mayston, D.J. (1996), "Educational attainment and resource use: mystery or econometric misspecification?" *Education Economics*, 4:127-142.

Mayston, D.J. (2000), "The demand for education and the production of local public goods", *Discussion Papers in Economics*, 2000/50, York, University of York.

Mayston, D.J. (2002), *Tackling the Endogeneity Problem when Estimating the Relationship Between School Spending and Pupil Outcomes*, DfES Research Report RR328, London, Department for Education and Skills.

Mayston, D.J. and Jesson, D. (1988), "Developing models of educational accountability" *Oxford Review of Education*, 14:321-340.

Mayston, D.J. and Jesson, D. (1999), *Linking Educational Resourcing with Enhanced Educational Outcomes*, DfEE Research Report RR179, London, Department for Education and Employment.

Office for Standards in Education (1995), *The OFSTED Handbook – Guidance on the Inspection of Secondary Schools*, London, TSO.

Orme, C. and Smith, P. (1996), "The potential for endogeneity bias in data envelopment analysis" *Journal of the Operational Research Society*, 47:73-83.

Performance and Innovation Unit (2000), *Adding It Up*, London, Cabinet Office.

Skills Task Force (2001), *Skills for All: Proposals for a National Skills Agenda, Final Report of the National Skills Task Force*, London, Department for Education and Skills.

Todd, K. (2000), *An historical study of the correlation between GCE Advanced level grades and the subsequent academic performance of well-qualified students in a university engineering department*, York, University of York.

Worthington, A. (2001), "An empirical survey of frontier efficiency measurement techniques in education" *Education Economics*, 9:245-268.

Public Purchasing of Health Services

Jim Malcomson

Professor of Economics
University of Oxford

Introduction

This paper is concerned with purchasing arrangements for the provision of public sector health services. Its starting point is the concern to get service suppliers to deliver appropriate care without excessive cost. Two issues, in particular, make that problematic:

- (1) Contractual arrangements for the provision of health services are incomplete in the technical economic sense that it is impractical to specify everything of concern when arrangements are made for provision.
- (2) The cost of providing any given level of service is to some extent under the control of the supplier.

There are, of course, many aspects of the provision of health services for which measures of performance have been devised and there is a lot of valuable research being done in this area. But it is almost certainly expecting too much to hope that these will ever cover all aspects of provision for even very specific health services. There are ultimately so many aspects of performance that there are always going to be some that cannot be measured satisfactorily. It makes sense to recognise, and take account of, that in making purchasing arrangements. For convenience in what follows, I refer to aspects of performance that are measured as *quantities*, those not measured as *qualities*.

The history of contracting in the British National Health Service (NHS) provides an interesting observation on this. In discussions with people negotiating on behalf of health authorities and NHS Trusts I was told that, in the early days, contracts were often extremely detailed. But, over time, there has been a move to less detailed contractual arrangements because it was recognised that many of the specifications in the early contracts could not be monitored effectively.

A common terminology in the literature is that the cost of providing a given level of service depends on the supplier's *effort to keep costs down*. The term comes from models in which the supplier is an individual person. But the concept is much more widely applicable. One can think of this effort as whatever determines the relationship between the value of outputs and the cost of inputs in any organisation – in other terms, its efficiency.

The discussion here is restricted to health services for which the whole cost is met from the public purse. Much of it also applies to patients with full private health insurance. But there are some differences that arise because private health insurers can typically be expected to have different objectives and to operate within the context of a market for insurance. I also do not discuss the issue of whether the suppliers should be privately or publicly owned. On that, see Hart, Shleifer and Vishny (1997).

The focus of the discussion here is on two issues:

- (1) The provision of health services themselves (from, for example, hospital trusts and GPs in their role as suppliers of services);
- (2) The control of access to services (via, for example, GPs in their role as gatekeepers who refer patients to specialist consultants).

For a fuller discussion of the issues in public purchasing of health services, see Chalkley and Malcomson (2000).

Service Provision

An essential issue in service provision under the conditions of concern here is a potential conflict between high quality and low costs that goes deeper than the obvious one that higher quality requires more inputs that must be paid for. This potential conflict is illustrated well by two forms of purchasing arrangements that have been widely used for health services:

- (1) *Cost reimbursement* under which the purchaser reimburses the supplier for all the costs the supplier actually incurs in treating a patient;
- (2) *Prospective payment*, the essence of which is that payment is related to the anticipated cost of treating a patient but independent of the cost actually incurred.

With cost reimbursement, there is in theory no problem in getting a supplier to provide high quality services. The experience with traditional full insurance in the US bears this out in practice. But the supplier has no interest in keeping costs down because they are all reimbursed, so it is tempting to provide what the patient wants whenever it is possibly beneficial, even if it is not cost effective. Again, US experience bears this prediction out. With prospective payment, in contrast, the supplier retains all cost savings and so has strong incentives to keep costs down. But the temptation here is to skimp on any aspect of quality that is not being monitored because that also saves on costs.

The literature has focused on a number of mechanisms to mitigate this conflict. One of these is the use of suppliers that are *not-for-profit institutions*. Not-for-profit institutions are widespread in the provision of health services. An obvious reason for an institution to have not-for-profit status is to commit itself to not distributing profit and hence to reduce the incentives to skimp on quality for financial reasons, see Glaeser and Shleifer (2001). The empirical evidence from hospitals (mostly for the US) has not, however, consistently found that not-for-profit hospitals provide higher quality services, though it is not clear that many of the studies adequately control for why particular institutions should be not-for-profit in the first place. For a review, see Sloan (2000). In any case, there are neither good theoretical nor empirical reasons to think that not-for-profit status is sufficient to overcome concerns about the quality-cost conflict.

A second mechanism discussed in the literature for mitigating the conflict is *clinical freedom of consultants*. Many hospitals give control of medical decisions to consultants who may be less concerned with costs, and more concerned with patient welfare, than hospital administrators, see Harris (1977). There has been little formal analysis of this and I am also unaware of empirical evidence on it. But it is certainly conceivable that this mechanism acts to provide a commitment that quality of service will not be skimmed in the interests of reducing costs. If that is correct, it suggests caution in acting on the view that consultants in the NHS have too much power and should have some of it taken away.

A third mechanism discussed in the literature for mitigating the quality/cost conflict is based on the *demand for services* from a supplier. If patients (and/or those from whom take advice, for example GPs) are aware of the quality of service provided by a supplier, then quality of service may affect the number of patients that supplier gets to treat. If the supplier also receives payment on the basis of the number of patients treated, it then has reason to provide good quality services to attract patients. Note that this mechanism does not necessarily require *competition* between suppliers – some patients may choose not to be treated at all for certain conditions if quality of service at their monopoly local hospital is sufficiently awful. There is a substantial literature showing that, under these circumstances, prospective payment with an appropriate payment per patient can maintain any quality level desired by the purchaser at minimum cost. This does not even require that patients and their advisers perceive quality of service correctly as long as their perceptions are correlated with quality. Moreover, the mechanism can work even for suppliers operating at full capacity (and hence unable to treat the extra patients the higher quality would generate) provided the purchaser can make payment depend on the number of patients added to the waiting list.

But one should not be too starry-eyed about this mechanism. It is unlikely to work in areas of, for example, accident & emergency where speed of attention is all-important. Moreover, there is substantial evidence (from the US) that patients misperceive *relative* dimensions to quality and then the mechanism works less well, see the summary in Chalkley and Malcomson (2000). Finally, prospective payment may be unnecessarily expensive if there is substantial variation in costs between suppliers or patients with the same basic diagnosis that the purchaser has less good information about than the supplier.

Asymmetric information of this sort about costs has led to many calls, particularly in connection with the Medicare system in the US, to replace prospective payment by some form of *partial cost reimbursement* (at a rate less than 100%) despite the adverse effects that may have on cost-reducing effort. The essential point is that, with prospective payment, the payment must be sufficient to cover the cost of a high-cost supplier and/or a high-cost patient. In contrast, with partial cost reimbursement, less can be paid to lower cost suppliers and for lower cost patients. There is a practical problem, however, that the precise nature of the optimal reimbursement is very sensitive to circumstances, so it is hard to come up with any more precise conclusions useful to policy makers. Moreover, many policy makers are sceptical that the cost savings would be substantial, particularly when account is taken of the expense of monitoring costs closely.

In an attempt to see whether the cost savings could conceivably be large, Chalkley and Malcomson (2002) investigated the potential for cost savings using data from the US Medicare system. Because optimal reimbursement is so sensitive to circumstances that it is hard for an outside investigator to know about, the approach used was to take the assumptions most favourable to large cost savings and see how big the numbers would be if they applied. Those assumptions are that all observed cost differences between patients in each diagnosis-related group (DRG) are forecastable to the supplier in advance of treatment, that suppliers are in a position to avoid treating patients at a loss, and that the purchaser is in a position to make “take it or leave it” offers to suppliers. We had rather hoped the analysis would show that, even under these most favourable assumptions, cost savings were unlikely to be large so that the issue would cease to be of concern. Unfortunately, it did not turn out that way. Cost savings may, of course, not actually be as large as those assumptions would indicate. But, for the moment at least, one cannot conclude that cost savings from partial cost reimbursement would be too small to be worth worrying about.

Gatekeepers

An important element in UK health services is the use of GPs to act as gatekeepers who control access to specialist services through their decisions about which patients to refer. There is a certain amount of evidence that health care systems with gatekeepers have lower costs than those without, see Gerdtham and Jönsson (2000, p.46). One of the reasons for setting up the GP fund holding system was to influence the behaviour of GPs in their role as gatekeepers. There has not, however, been much rigorous research on appropriate frameworks for analysing the gatekeeping function.

I myself have recently made a start on this, see Malcomson (2001). That research has come up with some reasonably robust conclusions:

- (1) Giving gatekeepers a budget from which they must meet the cost of any treatments for which they refer a patient (as with fund holding) gives too strong incentives not to refer patients in cases that are not clear-cut. An implication for a private insurance system is that it is better to have a third party act as the gatekeeper rather than the insurance company or health maintenance organisation (HMO) that provides the insurance.
- (2) Giving NHS patients the choice between a fund holding GP and a non-fundholding GP is likely to prove an unstable arrangement, certainly in the long run. Other things equal, patients will move towards non-fundholders. (The same conclusion would not necessarily apply to patients who pay for their own private insurance.)

Conclusion

I conclude with some general notes of caution. First, a simple point but one that is often overlooked. Health service provision is almost always an instance of what in the theoretical literature is called *multi-task agency*. That is, health service professionals are involved in lots of different activities. An important lesson from that literature is that providing the agent with increased incentives to perform one task well may result in other tasks being performed less well, especially those that cannot be easily monitored. This indicates caution about developing measures of more and more dimensions of performance and rewarding either improvements in these or achievement of specified targets. Unless one can devise measures of *all-important* dimensions of performance, it may simply result in the neglect of some other important aspect of performance. In the discussion about paying GPs for achieving targets for the proportions of children immunised against certain infectious diseases, for example, very little attention has been paid to what GPs will *not* do instead.

Second, economists do not have a very good understanding of professional ethics. In consulting work I did with others for the Department of Health on consultant contracts, Binmore, Malcomson and Ulph (2001), we could make consultant behaviour on private practice fit with a standard model only by building in a considerable element of what we called *NHS loyalty*, that is, a preference for doing NHS work over private work entering as a direct element in consultants' utility functions. Because the data we were given was not rich enough to do anything other than work with a short-term model, it is possible that this NHS loyalty is the result of some long-term reputation effect that the data would not allow us to capture. But there remains a concern that it may actually be the result of a genuine sense of loyalty. And if it is that, there is a concern that it may be dissipated if contractual arrangements are made too formal.

References

Binmore, K., Malcomson, J. M. and Ulph, D. (2001), *Consultants' contracts and the incentives to reduce NHS waiting times*, Report to the Department of Health.

Chalkley, M. and Malcomson, J. M. (2000), "Government purchasing of health services" in Culyer A. J. and Newhouse J. P., eds, *Handbook of Health Economics*, Volume 1A, Amsterdam, Elsevier Science.

Chalkley, M. and Malcomson, J. M. (2002), "Cost sharing in health service provision: An empirical assessment of cost savings", *Journal of Public Economics*, 84(2):219-249.

Gerdtham, U.G. and Jönsson, B. (2000), "International comparisons of health expenditure: Theory, data and econometric analysis" in Culyer A. J. and Newhouse J. P., eds, *Handbook of Health Economics*, Volume 1A, Amsterdam, Elsevier Science.

Glaeser, E. L. and Shleifer, A. (2001), "Not-for-profit entrepreneurs" *Journal of Public Economics* 81(1):99-115.

Harris, J. E. (1977), "The internal organization of hospitals: Some economic implications" *Bell Journal of Economics*, 8(2): 467-482.

Hart, O., Shleifer, A. and Vishny, R. W. (1997), "The proper scope of government: Theory and an application to prisons", *Quarterly Journal of Economics*, 112(4):1127-1161.

Malcomson, J. M. (2001), *Health service gatekeepers*, Department of Economics, University of Oxford.

Sloan, F. A. (2000), "Not-for-profit ownership and hospital behaviour" in Culyer A. J. and Newhouse J. P., eds, *Handbook of Health Economics*, Volume 1A, Amsterdam, Elsevier Science.

Performance Management in Health Care: Information, Incentives and Culture

Peter Smith, Russell Mannion and Maria Goddard*

Centre for Health Economics
University of York

Introduction

Health care presents some of the most serious challenges for measuring and promoting public service productivity. The report for the Treasury by Derek Wanless (2002) discussed many of the complex issues involved in seeking to measure and understand the concept, and a recent OECD conference highlighted the universal international policy concern with issues of health care performance and productivity (Smith, 2002a).

Amongst developed nations, the performance of the UK National Health Service has been subjected to an unparalleled level of scrutiny, and there has been a relentless series of organizational innovations. In the early 1990s these took the form of an internal market in health care, and experiments with general practitioner fund holding. Few of the market-based innovations were formally evaluated, and there is doubt amongst commentators as to whether – as implemented – they had a great deal of impact on NHS performance (Le Grand *et al*, 1998).

Since 1997 the reliance on the internal market has diminished, and instead a sometimes bewildering panoply of performance management instruments has been put in place (Smith, 2002b). The purpose of this paper is to summarize the findings to date of a programme of empirical research at the Centre for Health Economics that is examining the influence of these organizational instruments on the performance of the NHS. By the generic term “performance”, we refer to issues of both system effectiveness and system productivity. For the purposes of brevity, we assume that the objectives of the NHS are uncontested, although this is far from the case in practice. The paper can only sketch our findings in the briefest detail, and the interested reader is directed to the references for further details.

Performance Information

The NHS has a long history of publishing measures of the performance of its organizations, dating back to the first set of performance indicators in 1983. However, the reliability, timeliness and scope of UK performance measures has always been seriously hampered by lack of relevant data and the very poor NHS IT infrastructure. Furthermore, comparing even rudimentary aspects of health system performance, such as death rates after surgery, is remarkably complex, given the great range of procedures undertaken and the heterogeneity of patients (Iezzoni, 1997).

Compared with other health systems, one of the principal distinguishing features of the NHS has been its emphasis on cost control and efficiency (Mossialos and Le Grand, 1999). Measures of the relative efficiency of NHS organisations have therefore featured prominently amongst performance indicators. In recent years, one of the most important sources has been the set of hospital “reference costs” produced by the Department of Health. These report for each hospital estimates of the average costs for over 500 individual procedures, known as Healthcare Resource Groups (HRGs). The prime intention is to help NHS institutions and auditors identify areas of inefficiency and scope for improvement.

Dawson and Street (2000) examine the methodology underlying NHS reference costs, and highlight implausibly large variations in the reported costs for many procedures (see Table 1). They attribute these variations to questionable data quality and the arbitrary accounting choices that must be made when calculating such detailed cost data. They conclude that it is very difficult to identify the extent to which inefficiency is causing variation in the reference costs, and question the usefulness of such disaggregate cost data.

Table 1: NHS Healthcare Resource Groups costs (1998)

Code	Description	Number	Mean (£)	Min (£)	Max (£)
F47	General abdominal disorders <70 w/o cc	67,032	546	114	2,039
D20	Chronic obstructive pulmonary disease or bronchitis	63,543	1,184	443	8,403
E12	Acute myocardial infarction <70 w/o cc	49,060	1,076	248	4,226
E36	Chest pain <70 w/o cc	44,989	476	167	2,602
S16	Poisoning, toxic effects or overdoses	44,846	321	94	1,046

Source: Dawson and Street (2000)

It is nevertheless possible to use the detailed reference costs to construct an index of whole hospital efficiency. This can be defined as the ratio of observed costs to expected costs, where expected costs are calculated as the national average reference cost, given the mix of HRGs reported by the hospital. In 1999 the Department of Health published a series of five such cost indices, entailing different levels of methodological complexity. The most refined of these indices sought to take account of factors such as case complexity, research activity and hospital configuration as well as the mix of HRGs (Söderland and van der Merwe, 1999). In broad terms, the indices make different assumptions about what sources of cost variation lie within the control of the hospital.

Dawson, Goddard and Street (2001) examine the change in ranking amongst 213 acute hospitals implied by the use of the different indices. They find enormous shifts in rank, implying that – depending on what index is used – very different signals of hospital efficiency will emerge. Street (1999) and Street and Jacobs (forthcoming) construct 95% confidence intervals around the estimates of relative costs for one year, and find very little evidence of statistically significant differences in comparative costs between the hospitals. This is not to say that such differences do not exist – it merely indicates that the chosen instrument is not sufficiently sensitive to detect them reliably.

When Jacobs (2002) extends the analysis to include four years’ data and incorporates some quality information, it becomes easier to discriminate between good and poor performers. Importantly, she finds evidence of a trade-off between cost and quality,

* The authors are supported by the health economics research programme funded by the Department of Health. They thank colleagues Diane Dawson, Katharina Hauck, Rowena Jacobs, Stephen Martin, Nigel Rice and Andrew Street for material included in this paper.

Table 2: Pearson correlation coefficients of results for 232 hospitals from five DEA specifications and their SFA counterparts

	DEA-1	DEA-2	DEA-3	DEA-4	DEA-5	SFA-1	SFA-2	SFA-3	SFA-4	SFA-5
DEA-1	1									
DEA-2	0.2298	1								
DEA-3	0.3729	0.6340	1							
DEA-4	0.7575	0.3513	0.5372	1						
DEA-5	0.4722	0.6062	0.8352	0.6149	1					
SFA-1	0.4274	0.4667	0.5946	0.5166	0.5756	1				
SFA-2	0.0957	0.6209	0.4231	0.1831	0.4038	0.6354	1			
SFA-3	0.2154	0.4318	0.5975	0.3165	0.4852	0.8297	0.6917	1		
SFA-4	0.4192	0.4835	0.6583	0.5543	0.5998	0.8763	0.6815	0.8065	1	
SFA-5	0.3399	0.5195	0.6557	0.4633	0.6343	0.9496	0.6535	0.8731	0.8217	1

Source: Jacobs (2001)

suggesting that a satisfactory measure of NHS productivity should indeed incorporate measures of clinical quality.

Other methodologies for estimating comparative efficiency have been explored. Jacobs (2001) deploys Stochastic Frontier Analysis (SFA) and Data Envelopment Analysis (DEA) to examine whether more elaborate modeling can lead to more secure estimates of comparative hospital efficiency. She uses a single input (one of the cost indices) and five different sets of outputs. The rankings of hospitals obtained using these five plausible specifications using both SFA and DEA are then compared, using the rank correlation coefficients reported in Table 2. They indicate that there is a great deal of instability in the rankings implied by different model specifications, even when DEA and SFA are compared using identical outputs. Similar results are reported for primary care by Giuffrida and Gravelle (2001).

Hauck, Rice and Smith (forthcoming) develop an alternative approach to measuring performance. They examine fourteen diverse indicators of health authority performance taken from the NHS Performance Assessment Framework and disaggregate the data to a small area level (populations approximately 10,000). Using multi-level statistical techniques, they estimate the extent to which variations in small area performance are attributable to the health authority. They find great variations between the fourteen indicators (see Table 3). At the extremes, 80% of variation in waiting times is attributable to the health authorities, whilst only 10% of the variation in mortality is so attributable. Although such findings must be treated with some caution, they suggest that some aspects of health system performance are much more directly under the control of NHS managers than others. Therefore, unless incentives are designed carefully, managers may give low priority to important objectives of the NHS, such as the reduction of health

inequalities that imply long time horizons or liaison with other agencies.

We have been able to allude to only a few of the numerous initiatives for measuring performance of NHS organizations. The most recent has been a set of "star ratings" for acute hospitals that seeks to form a composite index of organizational attainment. This will shortly be extended to all NHS organizations. We are currently examining the theory and empirical properties of such composite indicators of performance (Smith, 2002a). The biggest challenge is to integrate measures of clinical quality into the measurement regime, an undertaking that is in its infancy, but – as the Wanless report notes – is crucial if we are to secure more meaningful measures of NHS productivity.

Performance Incentives

The most refined measurement of performance will have little impact unless it is undertaken in association with appropriate incentives. Two types of incentive can be considered: designed incentives and accidental incentives. The NHS has traditionally paid little attention to designed incentives, although that is now changing. At an institutional level, a system of "earned autonomy" has been put in place, under which organizations deemed to be performing well will be given increased freedom from control and inspection. We are currently undertaking a study of the impact of this incentive, and plan to report results in 2003. At the micro level, new contracts for general practitioners and hospital specialists will incorporate much greater attention to performance incentives than hitherto, and there have been experiments with team bonuses.

However, the predominant NHS approach has been to rely on public release of performance data and exhortation, rather than

Table 3: Proportion of variance in small area performance explained by region (RHA) and district (DHA) level

Code	Description	RHA	DHA
Smr064	Standardised mortality ratio for ages 0-64	0.04	0.10
Smr674	Standardised mortality ratio for ages 65-74	0.03	0.05
Sir074	Standardised limiting long standing illness rates for ages 0-74	0.11	0.19
Wtsurg	Waiting time for routine surgery (age, sex, speciality standardised)	0.15	0.61
Wtradio	Waiting time for radiotherapy (age, sex standardised)	0.08	0.62
Gpaccs	Accessibility to general practitioners (GPs)	0.08	0.25
Electeps	Number of elective surgery episodes (age, sex, speciality standardised)	0.15	0.41
Dcrate	Day cases as proportion of procedures (age, sex, speciality standardised)	0.14	0.57
Bedays	Length of stay (age, sex, speciality standardised)	0.10	0.11
Matcost	Maternity costs (age standardised)	0.25	0.32
Psycost	Psychiatry costs (age, sex standardised)	0.08	0.28
Wtlong	Percentage of those on waiting list waiting for 12 months (age, sex, speciality standardised)	0.12	0.58
Emold	Emergency admissions rate for people >65 (age, sex, speciality standardised)	0.23	0.14
Deaths	20 day death rate following hospital surgery (age, sex, speciality standardised)	0.11	0.15

Source: Hauck, Rice and Smith (forthcoming)

Box 1: Some unintended consequences of NHS performance measurement

- (1) **Tunnel vision 'Concentration on areas included in the performance indicator scheme to the exclusion of other important unmeasured areas'**
Examples: The focus on waiting lists has preoccupied NHS managers for years given their prime place in the performance measurement system. Other aspects of performance which are not easily or currently measured may have suffered as a consequence.
- (2) **Sub-optimisation 'The pursuit of narrow local objectives by staff at the expense of the objectives of the organisation as a whole'**
Example: Targets set for the hospital sector, such as higher rates of day case surgery or shorter lengths of stay, do not acknowledge the increased burden implied for primary care or social services.
- (3) **Measure fixation 'Pursuit of success as measured rather than as intended'**
Example: The 5 minute waiting time target for dealing with A&E patients led, in some hospitals, to the employment of the so-called 'hello' nurse who merely made contact with the patient within the first 5 minutes in order to meet the target.
- (4) **Myopia 'Concentration on short-term issues to the exclusion of long-term considerations which may only show up in performance measures in many years time'**
Example: Curative services (as measured by short-term process) may be given higher priority than preventive services (as measured by long-term outcome).
- (5) **Complacency 'Lack of ambition for improvement brought about by adequate comparative performance'**
Example: An apparently middling performance when judged against others on surgical survival can appear satisfactory and inhibit attempts to strive for further clinical improvement.
- (6) **Misrepresentation 'The deliberate manipulation of data by staff – ranging from 'creative' accounting to fraud – so that reported behaviour differs from actual behaviour'**
Example: Adverse patient satisfaction reports may be unaccountably lost, waiting lists manipulated or activity figures artificially boosted by various methods.
- (7) **Gaming 'Altering behaviour in order to obtain strategic advantage'**
Example: Trying to 'go easy' on achieving targets set in one year in order to avoid demanding targets for future years. If targets are based on year-on-year improvement there is a danger of creating a 'ratchet effect' where good performance in one year is punished with higher future targets. Some managers acknowledged that gaming occurs on the NHS efficiency targets.
- (8) **Ossification 'Organisational paralysis due to an excessively rigid system of measurement'**
Example: Using day case rates as an indicator of performance in gynaecology may inhibit the adoption of latest techniques for treating cases on an outpatient basis.

Source: Goddard, Mannion and Smith (2000)

designed incentives, as the main spurs to managerial effort. Reliance on informal incentives, such as professional and managerial prestige, career advancement and intrinsic satisfaction in good performance, may of course be perfectly satisfactory. There is however a danger that less benign accidental incentives also become influential, and we have found that the NHS has enjoyed mixed results.

In Scotland, the Clinical Resource and Audit Group (CRAG) pioneered the publication of clinical outcome indicators, such as five year survival of women with breast cancer, that are now being published throughout the UK. The intention of CRAG was to offer a supportive instrument for improving clinical performance. Using a qualitative methodology, Mannion and Goddard (2001) examined the influence of such publication on clinical practice. They found generally limited impact, and attribute this to a number of causes, including poor credibility, relevance and timeliness of the data, a lack of awareness and expertise on the part of clinical staff, and a lack of incentives and effective external scrutiny.

The unintended side effects of accidental incentives can be serious. Smith (1995) sets out a framework for analysing such side-effects, and argues that they arise from inevitable imperfections in the instruments for measuring, attributing and rewarding performance. Goddard, Mannion and Smith (2000) use this framework to study the impact of performance indicators in eight English hospitals, and find considerable evidence of unintended adverse effects. Some examples are given in Box 1. We would most

emphatically reject the claim that such phenomena call into question the wisdom of seeking to measure performance. But mitigating them may require careful attention to system design.

Historically, one of the most important aspects of NHS performance has been the waiting time for elective inpatient surgery. This has for many years been given a very high political profile, and – in contrast to many other performance measures – managerial careers may depend on satisfactory performance. Martin and Smith (1999) develop a theoretical model in which waiting time acts as a signal to both the demand side (in which a long waiting time serves to suppress demand for NHS surgery) and the supply side (in which a long waiting time serves to spur increased activity and efficiency). Empirical data from over four thousand small areas across England are then used to estimate econometric demand and supply equations. A small but negative demand effect of waiting time is detected, whilst the supply effect is (as expected) large and positive. These results are extraordinarily stable across time and speciality, and suggest that, given sufficiently high level of external scrutiny and focus, performance measures can induce the desired responses from NHS organizations (Gravelle *et al*, forthcoming; Martin and Smith, forthcoming).

In contrast, Dawson and Jacobs (2001) examine the impact of the system of efficiency targets used in the English NHS since 1992 to exert downward pressure on NHS unit costs. The efficiency targets are intended to reflect at an institutional level the Public Service Agreement efficiency objectives set by the Treasury for the NHS as

a whole. These have sought to secure an annual “efficiency” gain from the NHS of between 2 and 3 percent, where efficiency is defined in terms of costs per unit of throughput (that is, without reference to clinical quality). Over the period studied, regional offices of the NHS set a target for reductions in unit costs for each of their constituent health authorities. These were in turn implicitly passed on to the hospitals with which the health authority negotiated contracts. Although terminology has changed, this system remains largely in place.

Econometric analysis of six years’ hospital data could find no relationship between changes in hospital unit costs and the severity of the hospital efficiency targets set. Indeed, hospitals with stable unit costs appeared to face almost random assignments of targets. Interviews with key NHS staff established that there was little or no central monitoring of the efficiency gains required of individual hospitals, and little systematic recording of performance in relation to targets. The NHS continued to make the required “efficiency” gains, but this appears to have been a result of other pressures, such as waiting time requirements. In short, given the strict expenditure control exerted by the NHS, the efficiency targets regime appears to be redundant, and it is difficult to escape the conclusion that the setting of such targets has been an unproductive ritual.

A Performance Culture

Assuming NHS objectives are agreed, there are three fundamental components of the principal-agent model: the information base on which both principal and agent base their decisions; the incentives offered to the agent; and the degree of dissonance between the objectives of principal and agent. In common with most economic discourse, our discussion has focused on the first two of these. However, the most recent performance management innovations in the NHS have turned towards the third aspect of the agency problem: the objectives of the agent, in particular the clinician. Traditional economic thinking takes the agent’s objectives as given. Yet organizational sociologists have long argued that motivational issues are far more subtle than this naïve view, and the “new” institutional economics is beginning to recognize this lacuna (Kasper and Streit, 2000). There is considerable evidence that the preferences of public service employees are to some extent formed by the system in which they are asked to work (Le Grand, 1997). In particular, Victor Fuchs argues that professional norms play a crucial role in ameliorating the inevitable imperfections in the medical agency model.

There are signs that the NHS performance management regime now recognizes the importance of an environment in which clinical professionals themselves promote a culture of patient safety and continuous improvement. The NHS Plan (2000) and the Bristol Royal Infirmary Enquiry (2001) talked of the need for a “cultural transformation” of the NHS, and the policy document *Shifting the balance of power in the NHS* states that “the balance of power must be shifted towards frontline staff who understand patients’ needs and concerns. Frontline staff need to be in charge of frontline services and have the power to manage to meet the local communities needs – always within the context of clear national standards and a strong accountability framework” (Department of Health, 2001).

The natural way for the clinician to assess her own performance is with reference to others in the same discipline, and not the broad brush performance indicators described above. This clinical interest in the actions and performance of professional peers can be exploited in powerful ways in the form of professional networks and collaboratives (Ferlie and Pettigrew, 1996). Two indications of successful national NHS networks are the intensive care case mix programme and the risk stratified data on cardiac surgery outcomes (Keogh and Kinsman, 1999; Rowan and Black, 2000).

The key challenge is to integrate the traditional top-down instruments of external performance scrutiny (with their legitimate concern with standards, priorities and accountability) with a system of clinical networks (reflecting a professional concern with clinical quality, patient focus, peer review and continuous improvement). In this respect the role of senior managers is crucial. Under a concept known as clinical governance, they have a responsibility to assure health care quality, and one of their core roles should be to nurture an appropriate organizational culture. We are currently undertaking a study of the concept, measurement, and impact on performance of organizational culture in health care (Davies et al, 2000; Scott et al, forthcoming).

Conclusions

This paper has sought to shed light on some empirical aspects of NHS performance management. It is important to recognize that the management of health care is an extraordinarily complex undertaking, and that the UK has led the way in pioneering many managerial innovations. It is easy to lose sight of many successes, such as the tight expenditure control secured by UK health care. However, the results presented here do suggest that many of the performance management instruments have had only modest impact on NHS performance, and some may even have led to adverse outcomes.

The research nevertheless offers important positive messages for evidence-based system design. Our research is continuing, and the NHS performance management regime is still under development. However, at this stage, we would summarize our policy conclusions as follows:

- (1) NHS performance information has often been of poor quality, out of date, and of questionable relevance to clinicians. This is likely to change as the IT capacity of the NHS improves, but there will always be serious weaknesses in health care performance data, particularly in the crucial area of clinical quality. The use of broad aggregate measures of performance will rarely be of much interest to front-line staff.
- (2) There appears to be a serious lack of capacity to analyse and act on performance data within the NHS. Institutions such as the Audit Commission, the Commission for Health Improvement and the NHS Modernisation Agency provide a valuable national role. At a speciality level, the professions have a key role to play in developing clinical networks, and should have access to appropriate resources and expertise to that end.
- (3) There are important gaps in the scrutiny of NHS organizations. If aspects of performance are considered important, then non-executive directors, strategic health authorities and other relevant stakeholders should ensure that local managers are held to account.
- (4) Different aspects of the performance management regime are not yet fully articulated. For example, the NHS objectives expressed in the Public Service Agreement do not always filter down to the local level, through instruments such as the star ratings. A particular weakness is the lack of attention to clinical quality.
- (5) There is some redundancy in the performance management regime. All important aspects of the performance management should be subject to proper evaluation, and the system as a whole should be periodically reviewed.
- (6) The complexity of health care is reflected in the enormous volume of performance measures and targets confronting

managers. They quite reasonably complain of "target fatigue", and need clear guidance on priorities.

- (7) Attention to designed incentives is at an embryonic stage. There must be some doubt as to whether "earned autonomy" is a sufficiently powerful instrument. There is some resistance within the NHS to the use of personal or team financial incentives, but we suspect that these should form an important component of any successful performance management process.
- (8) There is considerable scope for adverse outcomes arising from the accidental incentives associated with performance management in health care. These need to be monitored carefully, and considered in any evaluation of performance management instruments.
- (9) Given the enduring information asymmetries that exist in health care, the engagement of clinicians in the performance management process is crucial. Development of an appropriate organizational culture appears to be an important means of addressing at least part of the agency problem.
- (10) Successful performance management is likely to require a judicious portfolio of instruments that addresses all three of the key elements of the agency problem: information, incentives and culture.

References

Bristol Royal Infirmary Inquiry (2001), *Learning from Bristol: the report of the public enquiry into children's heart surgery at the Bristol Royal Infirmary 1984-1995*, London, The Stationery Office.

Davies, H., Nutley, S. and Mannion, R. (2000), "Organisational culture and quality of health care" *Quality in health care*, 9:111-119.

Dawson, D. and Street, A. (2000), "Reference costs and the pursuit of efficiency", in P. Smith, *Reforming markets in health care: an economic perspective*, Buckingham, Open University Press.

Dawson, D. and Jacobs, R. (2001), *Hospital efficiency targets, A report for the Department of Health*, York, Centre for Health Economics, University of York.

Dawson, D., Goddard, M. and Street, A. (2001), "Improving performance in public hospitals: a role for comparative costs?" *Health Policy*, 57: 235-248.

Department of Health (2000), *The NHS Plan: a plan for investment, a plan for reform*, London, TSO.

Department of Health (2001), *Shifting the balance of power within the NHS: securing delivery*, London, Department of Health.

Ferlie, E. and Pettigrew, A. (1996), "Managing through networks: some issues and implications for the NHS" *British Journal of Management*, 7: S81-S99.

Giuffrida, A. and Gravelle, H. (2001), "Measuring performance in primary care: econometric analysis and DEA" *Applied Economics*, 33:163-175.

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*, 9:95-107.

Gravelle, H., Smith, P. and Xavier, A. (forthcoming), "Performance signals in the public sector: the case of health care" *Oxford Economic Papers*.

Hauck, K., Rice, N. and Smith, P. (forthcoming), "The influence of health care organizations on indicators of health system performance" *Journal of Health Services Research and Policy*.

Iezzoni, L. (1997), *Risk Adjustment for Measuring Healthcare Outcomes, Second Edition*, Chicago, Health Administration Press.

Jacobs, R. (2001), "Alternative methods to examine hospital efficiency: data envelopment analysis and stochastic frontier analysis" *Health Care Management Science*, 4:103-116.

Jacobs, R. (2002), *Does hospital efficiency come at the expense of quality?* York, Centre for Health Economics, University of York.

Kasper, W. and Streit, M. (2000), *Institutional economics: social order and public policy*, The Locke Institute.

Keogh, B. and Kinsman, R. (1999), *National Adult Cardiac Surgical Database Report, 1998*, London, The Society of Cardiothoracic Surgeons of Great Britain and Ireland.

Le Grand, J. (1997), "Knights, knaves or pawns? Human behaviour and social policy" *Journal of Social Policy*, 26: 149-169.

Le Grand, J., Mays, N. and Mulligan, J., eds, (1998), *Learning from the NHS internal market*, London: King's Fund Institute.

Mannion, R. and Goddard, M. (2001), "Impact of published clinical outcomes data: case study in NHS hospital trusts" *British Medical Journal*, 323, 260-263.

Martin, S. and Smith, P. (1999), "Rationing by waiting lists: an empirical investigation" *Journal of Public Economics*, 71:141-164.

Martin, S. and Smith, P. (forthcoming), "Using panel methods to model waiting times for NHS surgery" *Journal of the Royal Statistical Society, Series A*.

Mossialos, E. and Le Grand, J., eds, (1999), *Health care and cost containment in the European Union*, Aldershot, Ashgate.

Rowan, K. and Black, N. (2000), "A bottom-up approach to performance indicators through clinical networks" *Health Care UK*, Spring: 42-46.

Scott, T., Mannion, R., Davies, H. and Marshall, M. (forthcoming), "The quantitative measurement of organizational culture in health care: a review of available instruments" *Journal of Health Services Research and Policy*.

Smith, P. (1995), "On the unintended consequences of publishing performance data in the public sector", *International Journal of Public Administration*, 18: 277-310.

Smith, P., ed. (2002), *Measuring up: improving health system performance in OECD countries*, Paris, OECD.

Smith, P. (2002), "Performance management in British health care: will it deliver?" *Health Affairs*, 21, 103-115.

Söderland, N. and van der Merwe, R. (1999), *Hospital benchmarking analysis and the derivation of cost indices*, York, Discussion Paper 174, Centre for Health Economics, University of York.

Street, A. (1999), *Interpreting the NHS cost indices for acute trusts*, York, Discussion Paper 175, Centre for Health Economics, University of York.

Street, A. and Jacobs, R. (forthcoming), "Relative performance evaluation of the English acute hospital sector" *Applied Economics*.

Wanless, D. (2002), *Securing Our Future Health: Taking A Long-Term View*, London, HM Treasury.



HM TREASURY

www.hm-treasury.gov.uk