



WP/05/210

IMF Working Paper

Does Performance Budgeting Work?

An Analytical Review of the Empirical Literature

Marc Robinson and Jim Brumby

IMF Working Paper

Fiscal Affairs Department

Does Performance Budgeting Work? An Analytical Review of the Empirical Literature

Prepared by Marc Robinson and Jim Brumby¹

Authorized for distribution by Jack Diamond

November 2005

Abstract

This Working Paper should not be reported as representing the views of the IMF.

The views expressed in this Working Paper are those of the author(s) and do not necessarily represent those of the IMF or IMF policy. Working Papers describe research in progress by the author(s) and are published to elicit comments and to further debate.

This paper attempts to ascertain what light the empirical literature sheds on the efficacy of performance budgeting. Performance budgeting refers to procedures or mechanisms intended to strengthen links between the funds provided to public sector entities and their outcomes and/or outputs through the use of formal performance information in resource allocation decision making. The paper seeks to identify and examine the literature on “government-wide” performance budgeting systems—that is, systems used by central budget decision makers (ministry of finance and political executive) to link the funding they provide to those agencies’ performance. Performance budgeting principles are, however, applied not only on a government wide basis, but also in funding systems applied to specific categories of government services. This paper does not attempt to review the empirical literature on all such “sectoral” performance budgeting systems. Rather, it undertakes a case study of the literature on one specific sectoral system—output-based hospital funding systems.

JEL Classification Numbers: H4, H5, I1

Keywords: performance budgeting, performance measures

Author(s) E-Mail Address: mrobinson@imf.org; jbrumby@imf.org

¹ This paper has benefited greatly from comments and advice from Phil Joyce, Jonathan Breul, Harry Hatry, Katharine Mark, Jeff Richardson, Luke Connelly, Christiane Roehler, Joel Turkewicz, Matthew Andrews, Keith McKay, Bill Dorotinsky, and a number of other participants in a seminar held by the Fiscal Affairs Department of the IMF.

| Contents | Page |
|--|------|
| I. Introduction | 3 |
| II. What is Performance Budgeting?..... | 5 |
| A. Defining Performance Budgeting | 5 |
| B. Misconceptions about Performance Budgeting..... | 9 |
| C. Why Performance Budgeting? | 12 |
| D. Performance Budgeting in the Broader Context | 14 |
| E. Key Reasons why Performance Budgeting Might not Work | 15 |
| III. The Efficacy of Government-Wide Performance Budgeting | 17 |
| A. Introduction..... | 17 |
| B. Budgetary Allocation Impacts..... | 18 |
| C. Aggregate Expenditure Impacts..... | 29 |
| D. Impacts on Productive Efficiency and Program Effectiveness..... | 30 |
| IV. Sectoral Performance Budgeting Systems: A Case Study..... | 32 |
| A. The Casemix Funding System | 32 |
| B. The Empirical Evidence | 36 |
| C. Implications for Sectoral Performance Budgeting Systems More Generally..... | 41 |
| V. Conclusions and Reflections..... | 43 |
| A. Overview of Findings..... | 43 |
| B. Behavioral Foundations of Performance Budgeting and Managing-for-Results | 46 |
| Appendices | |
| I. U. S. Studies Relevant to the Efficacy of Performance Budgeting | 52 |
| II. U. S. GASB Case Studies | 57 |
| References..... | 63 |

I. INTRODUCTION

Performance budgeting has, in one form or another, been an important theme of public expenditure management for decades. In the 1990s, however, a new wave of enthusiasm for performance budgeting began to sweep through advanced nations, later spreading through developing and transitional nations. The new performance budgeting initiatives during this period have typically been part of a broader set of reforms that are changing both the way in which the public sector is managed and also the boundary between the public and private sectors. Their inspiration has been a widespread feeling that there was, and is, a compelling need to deal with the problem of disappointing public sector performance. Performance budgeting has been closely linked with other reforms in public sector budgeting and financial management that have aimed not only at improving public sector performance, but also at ensuring fiscal sustainability. Indeed, there is a close link between the performance and sustainability objectives, because the fiscal consolidation episodes experienced by many countries since the late 1980s have underlined the importance of ensuring that limited public resources are spent on the public services that are of most benefit to the community, and that these services are produced more efficiently.

Given the resources and effort that are being committed to performance budgeting systems, it is important to be confident that these systems work. It must, however, be frankly acknowledged that many analysts believe that, as Schick (2003, p. 83) puts it, “efforts to budget on the basis of performance almost always fail.” Many have looked at the experience of performance budgeting and come to the same conclusion that the U. S. Congressional Budget Office reached in a report a decade ago—namely, that there is

... little evidence of the much-touted advances in performance-based budgeting in local, state, and foreign governments. Performance measures have a limited ability to influence the allocation of resources (CBO, 1993, p. x).

Such assessments should not be lightly dismissed, or disregarded. Is performance budgeting perhaps inherently political naïve and administratively utopian, as some critics have suggested? Is the renewed enthusiasm for performance budgeting yet another case of the human propensity to allow hope to triumph over experience?

These questions are the inspiration for this paper, which examines empirical literature on, or relevant to, the efficacy of performance budgeting, paying particular attention to the experience of the new wave of performance budgeting. Our aim has been to identify and review the empirical literature pertaining to government-wide² performance budgeting systems in the period since 1990. This literature tends to fall into two categories: quantitative analysis based on expenditure data or opinion surveys, and case studies. It is a literature that is overwhelmingly the work of political scientists.

² The distinction between government-wide and “sectoral” performance budgeting systems is discussed in Section II.

Our second aim is to examine more selectively the empirical literature on sectoral performance budgeting systems. Here the example of the output-based “casemix” hospital funding model has been chosen for particularly close attention. Economists were and are heavily involved both in the design and the evaluation of casemix funding.

It would be unreasonable to require conclusive empirical proof of the efficacy of performance budgeting, given the methodological difficulties involved. As Jones and Kettl (2003, p.1) observe more generally in respect to the global public management reform, “there is a glaring need to understand the short- and long-term outcomes of the reforms... [but] doing so is almost impossible in the short term and exceedingly difficult in the long term.” Clearly, the well-known difficulties of public sector performance measurement create major obstacles to assessing the impact of reform generally on allocative efficiency and even, albeit to a somewhat lesser degree, on productive efficiency. Beyond this, there are at least three difficulties in attempting to assess the specific impact of performance budgeting. First, because performance budgeting is part of a broader set of management reforms, there is a seemingly intractable problem in distinguishing the impact of performance budgeting from the impact of other reform strategies. Second, the success of performance budgeting and other reform strategies depends not on only the technical design of those strategies, but upon a range of other contextual factors—including the political system, political culture, and the fiscal environment—the impact of which is exceedingly hard to measure. Third, the very concept of performance budgeting has not always been well-defined.

Although we would therefore not necessarily expect the empirical literature taken in isolation to be conclusive, we believe that this literature does shed substantial light on the efficacy of performance budgeting. It is particularly relevant to establish the extent to which negative views on the efficacy of performance budgeting are in fact validated by the empirical literature. Beyond that, we would hope that the empirical literature would also shed light on the preconditions for effective performance budgeting, and might also be of assistance in the important broader task of identifying which forms of performance budgeting work best.

The structure of the paper is as follows. Section II considers the definition of performance budgeting and delineates it from the broader reforms with which it is usually linked. This section also critiques problematic definitions employed in the literature—particularly where they have served as the basis for empirical studies. Section III focuses on the empirical literature concerning the efficacy of government-wide performance budgeting systems. In Section IV, we turn to “sectoral” performance budgeting systems, and examine the empirical literature on the impact of the casemix hospital funding system. Section V identifies and considers a number of issues to which the empirical literature points, and makes some suggestions for fruitful future research. It concludes with certain reflections on the potential role that economists might play in improving their understanding of what types of performance budgeting work and under what circumstances.

II. WHAT IS PERFORMANCE BUDGETING?

A. Defining Performance Budgeting

The multiplicity of definitions of performance budgeting in the literature makes it essential that any rigorous treatment of the subject start by clearly defining the concept. In this paper, performance budgeting refers to procedures or mechanisms intended to strengthen links between the funds provided to public sector entities and their outcomes and/or outputs through the use of formal performance information in resource allocation decision-making. “Formal” performance information in this context refers to performance measures,³ measures of the *costs* to particular parties of outputs and outcomes,⁴ and assessments of the effectiveness and efficiency of expenditure obtained through the use of any of a range of analytic tools.⁵ The core objectives of performance budgeting are enhanced allocative and productive efficiency⁶ in public expenditure.

This definition is quite close to those advanced by, for example, the OECD (“a form of budgeting that relates funds allocated to measurable results” (OECD, 2003b, p. 7)) and the U. S. General Accounting Office (“the concept of linking performance information with the budget” (GAO, 1999a, p. 4)). Like these definitions, it embodies a very general notion of performance budgeting which encompasses a diverse range of specific performance budgeting systems. It includes the classic U. S. performance budgeting models—the Planning, Programming and Budgeting System (henceforth “program budgeting” for short), Zero-Base Budgeting (ZBB), and the “performance budgeting” proposed by the Hoover

³ No distinction is drawn in this paper between performance “measures” and “indicators.”

⁴ A note on terminology: consistent with predominant usage, *outputs* in this paper refer to goods or services provided to external parties (particularly to the public) by agencies, whereas *outcomes* refer to the desired changes induced in persons, social structures or the physical environment by those outputs and other factors. Thus, for example, the treatment of an appendicitis patient is an output, whereas the desired outcome is (roughly speaking) the full recovery of the patient. *Inputs* are the resources (labor, materials, equipment, buildings etc) used by agencies to produce outputs. Unfortunately, even though this represents predominant contemporary usage, terminology in this area is not completely standardized. See M. Robinson “The Output Concept and Public Services,” forthcoming in *Public Performance and Management Review*.

⁵ Ranging from conventional policy reviews based on relevant social/economic theory, through to program evaluation methods and more formal methodologies such as data envelopment analysis and cost/benefit analysis.

⁶ Productive efficiency refers to the degree of success in delivering a defined output (or carrying out a specific activity) at minimum cost while holding quality constant, given prevailing input prices. This corresponds to the term “efficiency” (as opposed to effectiveness) as used in the public administration literature. Productive efficiency encapsulates two distinct economic concepts of microeconomic efficiency. The first of these is “technical efficiency,” which refers to the choice of the least-cost combination of inputs for the production of the good or service concerned, for given input prices. The second is efficiency in the sense of the absence of waste (which economists sometimes refer to, following Liebenstein, as X-inefficiency).

Commission⁷—and variants of these. It also embraces models developed more recently in other countries, such as the output-purchase budgeting systems of New Zealand and Australia, and the British Public Service Agreement (PSA) system.⁸ No distinction is, incidentally, drawn here between performance budgeting and other widely used terms such as “performance-based” budgeting.⁹

This is a definition that does not restrict the performance budgeting concept to resource allocation decisions made in the formulation of the government-wide budget. Rather, it is relevant also to the way in which agencies execute their budgets—either when allocating resources internally or when distributing funding to lower-level public sector entities which they supervise. Thus the use by a health ministry of a technique such as “program budgeting marginal analysis” (Debble, 2000, pp.17–19) to improve the allocation of the health budget, constitutes a form of performance budgeting. So also does casemix funding of public hospitals, or the use by an education ministry of a combination of output funding and outcome bonuses for the funding of public universities. Casemix and performance-based university funding are examples of what are referred to in this paper as “sectoral” funding systems—that is, systems used to allocate funding between *multiple* public producers of specific types of service.

One reason why this is important is that allocative efficiency can never be a function simply of the allocation of resources in the government-wide budget. Even though budgeting systems vary in the extent to which they *seek* to centralize resource allocation decisions, the reality is that even in the most centralized systems a great deal of allocative decision-making necessarily takes place at the discretion of agencies within the parameters of their budget authorizations.¹⁰ Information costs and bounded rationality make this inevitable. All too often, as Joyce (2003, p. 7) notes, discussions of performance budgeting implicitly assume that “resource allocation is something that occurs only (or at least mostly) in the central budget office or in the legislature,” which leads to an unduly narrow empirical focus upon allocative impacts only at the level of government-wide budget formulation.

⁷ For good outlines of these see, for example, Mikesell (1995), Diamond (2003b) and GAO (1997a).

⁸ Outlines of the New Zealand models may be found in Carlin (2003), Robinson (2000) and Ball (1994). Australian introduced in the late 1990s a system heavily influenced by the New Zealand purchaser-provider model, but which claimed to place greater emphasis upon outcomes (see Robinson, 2002a, 2002b). The U.K. PSA model is described in James (forthcoming), and in much official U.K. Treasury material. The literature on other non-U.S. models is, regrettably, much more limited.

⁹ There are, of course, a range of other terms which tend to be used more or less synonymously in the literature, such as “results-based budgeting” and “performance funding.” Joyce’s (2003, p. 6) term “performance-informed budgeting” is perhaps the one which best captures what performance budgeting is about.

¹⁰ There are, indeed, three levels of decision-making pertaining to the economist’s concept of allocative efficiency: resource allocation decisions embodied in the government-wide budget, resource allocation decisions within agencies, and service (output) design decisions which affect program effectiveness.

Different performance budgeting systems attempt to link performance information to funding decisions in different ways—with, undoubtedly, varying degrees of effectiveness. The key procedures or mechanisms used for this purpose are the following:

- Classification of expenditure into “programs” with common objectives to facilitate better prioritization and more relevant accountability. This *program classification principle* was the core idea of program budgeting and remains—albeit often with a change of terminology¹¹—an element of many contemporary performance budgeting models.
- The setting of outcome or output targets which are *linked to the level of funding provided*. The U.K. PSA system, in which primary emphasis is now placed upon *outcome* targets, is one notable example of a target-based performance budgeting system. The PSA approach was driven by a determination that significant increases in funding in areas such as health and education should yield improved outcomes, rather than being swallowed up entirely in cost increases. The most fundamental principle of the system was, consequently, that the provision of additional budget funding would in general require agreement on tougher performance targets. Many OECD countries claim to set budget-linked performance targets, although the real extent of the linkage between targets and funding is not always clear.¹²
- Formal agreements (sometimes referred to as “contracts”) intended to clearly articulate such funding-linked targets, and sometimes also to give some indication of the consequences of delivery performance.
- Formula-based budget estimation based upon outputs: in particular, estimating expenditure requirements by multiplying forecast or planned output volume by a per-unit funding amount. This idea was part of the Hoover’s Commission version of performance budgeting.¹³ School funding formulas based upon estimating student course-years, possibly adjusted for other factors, fall broadly into this category.¹⁴ An example of this in internal agency budgeting is the U.S. Working Capital Funds

¹¹ In which “programs” have been renamed in a variety of ways (e.g., output groups/classes, outcome groups, missions).

¹² OECD (2003b, p. 9) reports that of 27 OECD nations which responded to a survey, 12 asserted that they had performance targets which were set as part of the budget process, and in 9 of these the results relating to those targets were reported ex-post to the finance ministry.

¹³ The Hoover Commission approach also included the use of funding formulas based upon support services (intermediate products) and activities, and was thus not purely output-based.

¹⁴ This can be contrasted with input-based funding (based, for example, on teacher numbers) or outcome-based funding (e.g., payment of bonus funding based upon achievement of above-standard student exam results).

which base the funding of internal support services in some major government departments upon unit costs.¹⁵

- Formularized output-based supplementary funding arrangements: this is a special form of formularized output-based funding designed to cope with unanticipated extra demand for a service deemed essential. By way of example, under the “workload agreements” (and subsequent “purchasing agreements”) used in Australia during the 1990s, the arrival of an unexpectedly large number of illegal immigrants subject to Australia’s mandatory detention policies resulted in the immigration department receiving additional funding on a formularized per-detainee/day basis (DIMA/DOFA, 2001).
- “Output-purchase” budgeting: This involves payment for outputs actually delivered. Under the “casemix” funding system, public hospitals in a number of countries are funded principally on the basis of the services they deliver, with a different price set for every output type as defined by the “DRG” output classification system (see Section V). This means that hospitals make a loss (profit) if the actual cost of the service delivered exceeds (is less than) the price they receive. The hospital is only paid for the services it delivers to patients (“payment by results”), so a failure to deliver expected volumes of services results in less funds. Assuming appropriately calculated prices, the system creates a strong financial incentive for efficiency. The “output purchase” performance budgeting system developed by New Zealand in the 1990s¹⁶ was at least at its inception motivated by a desire to generalize this approach to the entire government budget.¹⁷
- Outcome/quality-based performance bonus funding: the existence of an explicit component of funding which is awarded on the basis of success in achieving outcomes and/or delivering quality outputs. Such funding may be formula-based (e.g., bonus payments to universities based upon measures such as graduate employment rates or post-graduation salary levels). An alternative approach involves a more judgmental performance rating or assessment process. Louisiana and a handful of other U.S. states have in recent years made provision for explicit performance bonus payments to state departments which are assessed as having performed well.
- Procedures for feeding usable performance information into budgetary priority-setting processes and structures. A simple example is a requirement that agencies justify their budget bids through the use of performance measures and other relevant performance

¹⁵ We are indebted to Howard Borgstrom for drawing these to our attention.

¹⁶ In the mid-1990s Singapore adopted what would seem to have been a very similar system, called “Budgeting for Results” (Jones, 1998).

¹⁷ These systems also stressed the principle that that the “prices” paid to ministries and agencies by government for their outputs would equal the cost of production by an efficient, competitive producer, thus strengthening further the incentives to improve efficiency.

information, and that finance ministry analysts consider this information in assessing the bid. The finance ministry or other central agencies might also conduct program performance rating or program evaluations for use in the budget process. A recent example of such an approach is the U.S. PART (Program Assessment Rating Tool) system, which is an important element of the “budget and performance integration” initiative launched by the Bush administration. Under PART, Office of Management and Budget (OMB) budget examiners rate in each annual budget round the performance of 20 percent of government programs, in a process which is designed to place the onus on the agency to demonstrate that its programs are effective (OMB, 2003; GAO, 2004). These ratings are expressly intended to inform resource allocation decisions in the president’s budget (Executive Office of the President/OMB 2002). A broadly similar, but less recent, example of such a process is the linkage of a formal program evaluation system introduced in Australia in the 1980s to expenditure priority reviews by the cabinet-level Expenditure Review Committee.

Some scholars have defined performance budgeting as the reporting of performance measures in public budget documentation (e.g., Jordan and Hackbart, 1999). This differs from the conception of performance budgeting employed in this paper, which focuses on there being mechanisms in place for the *use* of performance measures and other performance information in funding decisions.

In defining performance budgeting, it is important to mention one other crucial ingredient which underpins all performance budgeting systems. This is the proposition that it is essential to reduce centrally determined budgetary and other controls over the mix of inputs which agencies employ to deliver services. From the 1960s, program budgeting sought to replace traditional “line item” budgeting—in which budgets appropriated funds to agencies by detailed input categories, and agencies were largely unable to shift funding from one line-item to another—with budgetary allocations or classification by outcome/output. Program budgeting also often called for the removal of analogous input constraints such as staffing controls. Since that time, a wide consensus has emerged that to tie the hands of public sector managers on the mix of inputs used in the production process is simply to guarantee technical inefficiency and high costs.¹⁸

B. Misconceptions about Performance Budgeting

The above list of performance budgeting processes and mechanisms, although not comprehensive, underlines the diversity of performance budgeting systems. This is important to any investigation of the efficacy of performance budgeting because there is *prima facie* reason to expect that not all forms of performance budgeting will work, and that the efficacy

¹⁸ The point is actually a more general one. Thus in the context of a performance budgeting system which sets *outcome* targets for agencies, it is necessary to accord to the agency significant control over the operational management design of the services.

of specific versions of performance budgeting may depend upon the context in which they operate. In addition, it draws attention to the unduly narrow nature of many of the definitions of performance budgeting to be found in the literature, which relate too specifically to performance budgeting practice in particular countries or in particular time periods.

Definitions which imply that performance budgeting is only concerned with the link between funding and performance *measures*—and not with performance information more generally—are not uncommon.¹⁹ Such definitions are half right, because performance measures are a necessary and fundamental element of the performance information required by any performance budgeting system. But they are only half right because in many contexts performance measures can provide a starting point in making decisions, and need to be used in conjunction with other types of performance information. Thinking of performance budgeting simply in terms of the use of performance measures in budgeting has perhaps led some analysts to look for a direct or even mechanical link between measures and budget decisions which it is often not reasonable to expect.

As always, context matters. In some sectoral performance budgeting systems—such as the casemix funding of hospitals—a very direct, formula-based relationship between measures and funding can work. Government-wide performance budgeting system, however, not only have to contend with more severe performance measurement problems, but are centrally concerned with the allocative task. The efficient budgetary allocation of expenditure between competing purposes can never be determined simply by reference to performance measures, no matter how good. Even if we were able to measure the marginal social benefit/dollar of existing programs, we would still need to consider whether a poorly performing program should be redesigned before deciding to cut expenditure. And because there are in fact such great limits to our capacity to measure the performance of many public programs, extensive reliance has often to be made of relevant theory and other analytic tools not only to interpret performance measures but also to show us the way in areas where measurement is unable to shed light. In short, good budgeting should be based on good policy making, and performance measurement is only one element of good policy making. This is a point which classic forms of performance budgeting such as program budgeting clearly recognized, by emphasizing not just measures by other analytic tools.

Another unduly narrow conception of performance budgeting is that it aims to create links only from *past* performance to present funding. Andrews and Hill (2003, p. 138), for example, suggest that performance budgeting necessarily “introduces a ‘rule’ that budgetary decisions should be made on the strength of previous performance, not accepted process.”²⁰ Equivalently, some analysts see performance budgeting as necessarily and by definition

¹⁹ For example, the Florida Office of Program Policy Analysis and Government Accountability (OPPAGA, 1997, p.13) refers to “performance-based budgeting” as a system in which “resource decisions are to be made on changes demonstrated by performance measures.”

²⁰ See also, along the same lines, Layzell’s (1998:107) definition of “performance-based budgeting.”

concerned with the creation of budgetary incentives for performance.²¹ It is quite true that some (not all) performance budgeting systems do seek to build specifically *budgetary* incentives for agency performance, so that, for example, strong performance by an agency results in additional funding or retained resources (thus creating a stronger *ex post* linkage between results and funding). Such agency-level financial incentives have, arguably, had proven success in some sectoral performance budgeting systems (such as the casemix funding model considered in Section IV). The idea is also influential in some government-wide performance budgeting models. Thus a key element of the thinking behind the Bush administration's "budget and performance integration" initiative has been the desire to change a situation where there is "little reward, in budgets or in compensation, for running programs efficiently" (Executive Office of the President/OMB 2002, pp. 27–28).

The fact that *some* performance budgeting systems stress *ex post* links between results and funding does not, however, mean that all performance budgeting systems have this property. It is, in fact, links between funding and expected *future* performance which are the focus of some performance budgeting systems. There are, moreover, a number of forms of performance budgeting which do not incorporate any notion of budgetary rewards for good past performance. Such an idea is, for example, quite foreign to program budgeting.²² Logically, moreover, it is quite possible to have a performance budgeting system which aims to link funding to expected future results through, say, targets, but in which sanctions for management failures are of an exclusively non-budgetary kind. Indeed, there are significant unresolved issues about the role of financial performance incentives within government, as a consequence of which it is "[not] always clear what should be the budget implications of poor performance" (Diamond, 2003a, p. 6). The potential role and modus operandi of such incentives in government-wide performance budgeting systems seems to be particularly unclear at present. As a 1997 Florida study of such systems in other U.S. states observed, "in all systems we studied, understanding of how to apply incentives and disincentives to change performance in desired ways has been difficult" (OPPAGA, 1997, p. 23).

²¹ Thus Robert D. Behn (1994) critiques "the popular currently budget reform known as performance budgeting," which he characterizes as follows: "Measure the performance of different agencies, and then adjust agency budgets accordingly. Agencies that do a good job should have their budgets increased. Agencies that do a poor job should have their budgets cut." See also, along similar lines, the Memorandum by the Institute of Directors (paragraph 3.3) to the recent enquiry into the PSA system by the British House of Commons Select Committee on Public Administration (<http://www.parliament.the-stationery-office.co.uk/pa/cm200203/cmselect/cmpubadm/62/62ap18.htm>).

²² A more subtle difficulty is the ambiguity of references to past poor performance. One might distinguish loosely between poor performance which reflects the inherent incapacity of a service or program to deliver outcomes sufficient to justify its cost, and poor performance which reflects remediable management weaknesses in the organization or design of a potentially useful service. All performance budgeting systems concerned with allocative efficiency by definition agree that poor past performance in the first of these senses should lead to funding being cut. There is, however, no similar consensus that funding cuts are an appropriate response to poor performance arising from remediable management or design problems.

One further misconception about performance budgeting is that it necessarily adopts a central planning approach to the task of maximizing allocative efficiency—seeking, in other words, to concentrate allocative choices entirely in the hands of central budget decision-makers. There are, as discussed in Section III, some performance budgeting systems for which this was a description with some validity—including classic models such as program budgeting. These days, it is widely recognized that information asymmetries mean that central budget decision-makers should specialize their focus away from matters of operational detail and instead focus on the policy and administrative issues which will determine sectoral expenditure allocations (World Bank 1998; Campos and Pradhan, 1999). Many—although not all—contemporary performance budgeting systems recognize this. As part of this, agency budget appropriations often take a highly aggregated form, giving agencies freedom to shift money within budgetary programs, and with agreement, between budgetary programs. This is an additional reason why it is inappropriate to employ a definition of performance budgeting which excludes *internal* agency budgeting. Another way of expressing this point is that performance budgeting is relevant to both *budget formulation* and *budget execution*.

C. Why Performance Budgeting?

As noted above, the core objectives of performance budgeting is enhanced allocative and productive efficiency in public expenditure.

In respect to allocative efficiency, performance budgeting reformers have been driven by a belief that expenditure allocation in the public sector tends to be insufficiently responsive to changing social needs and priorities. The perception is that money can keep flowing year after year to ineffective programs because of a lack of accountability for results linked to the budget process. As the recent *President's Management Agenda* (2002, p. 27) put it, “once money is allocated to a program, there is no requirement to revisit the question of whether the results obtained are solving problems the ...people care about.” Many program budgeting proponents took the view that a key part of the problem was excessive budgetary “incrementalism”—that is, a tendency for the “base” funding of established agencies and programs to be unthinkingly renewed in each budget. This becomes an obvious problem particularly when government priorities change significantly, or when significant new public policy challenges emerge. Increasing the responsiveness of the budgetary allocation of resources to government priorities has, therefore, been an express objective of some performance budgeting systems. In New Zealand, for example, the financial management reforms of the late 1980s and 1990s were designed partly to “assist the government to translate its strategy into action” more effectively (New Zealand Treasury, 1996, p. 7). Another aspect of the allocative efficiency objectives of some performance budgeting systems has been a desire to increase the responsiveness of resource allocation to short-run unexpected fluctuation in needs/demand for services in some crucial areas.

More recently, performance budgeting has been increasingly influenced by the view that public sector agencies tend to perform disappointingly. Reflecting this, most models of

performance budgeting developed since 1990 tend to view the budget process as a means of increasing the pressure upon agencies to lift their performance.²³ In almost all cases, they seek to use budget-linked targets, prices and performance agreements to greatly increased clarity about the results which agencies are expected to deliver with funding provided to them (i.e., to strengthen the *ex ante* linkage between results and funding).

This reflects a major change in attitudes to public sector performance. In the period after the second world war, when the modern welfare state was constructed and the role of the state extended greatly in both nonmarket and market production, optimistic expectation about the superior performance potential of public sector entities were widespread—even amongst economists (Shleifer, 1998, p. 133–35; Tanzi and Schuknecht, 2000, p.10–13). The reality turned out to be much more sobering. Disappointing public sector performance is widely attributed, and by no means only by economists, to the absence of market performance disciplines. Specifically, the view is that the absence of competition, of a market for corporate control, and of the strong financial incentives for individual performance—leaves the general government sector chronically prone to shirking, empire-building and other behaviors likely to undermine both technical and allocative efficiency. The traditional instruments of democratic political accountability—important as they are—are now often regarded as insufficient to fully compensate for this performance deficit.

Although productive efficiency and efficiency in the allocation of public expenditure are the “core” objectives of performance budgeting, there is also an important link between performance budgeting and macro-level goals pertaining to the level of aggregate expenditure and to fiscal sustainability. Clearly, to the extent that performance budgeting succeeds in boosting productive efficiency, aggregate public expenditure should, *ceteris paribus*, be smaller than it would be otherwise. Expressed differently, there is a “productivity dividend”, which can be used either to keep the tax burden down, or to fund new service priorities.²⁴ Moreover, in the case of a government seeking consciously to achieve reductions in aggregate government expenditure—whether for allocative efficiency or fiscal consolidation purposes—improved expenditure prioritization is also potentially of the greatest importance. The principle is that, by making expenditure reductions more discriminating and thereby reducing reliance upon crude techniques such as across-the-board cuts, spending cuts become somewhat less difficult to accomplish.

²³ A more detailed discussion of the new performance budgeting will be found in M. Robinson “Tightening the Results/Funding Link in Performance Budgeting Systems” in F. Mellemvik and A. Bourmistrov (ed), *International Trends and Experiences in Government Accounting*, 2005.

²⁴ Note that, for example, the stated aim of efficiency targets introduced in the first round of British Public Service Agreements in 1998 was “to ensure more resources go direct to front line services.”

D. Performance Budgeting in the Broader Context

Changed approaches to performance budgeting have, of course, been only one element of a much wider response to disappointing public sector performance. The boundaries between the public and private sector—in both production and provision—have been reconsidered, and the consequence has been widespread privatization and an increased private production of publicly-funded services.

In terms of public sector management, performance budgeting has in recent decades commonly been adopted as part of a broader set of management and budgetary reforms designed to improve the efficiency and effectiveness of the public sector and/or to facilitate the achievement of fiscal sustainability. Many of these reforms fall into the category of what is commonly referred to as *managing-for-results*, while others introduce increased consumer choice and competition. Managing-for-results can be defined as the use of formal performance information to improve public sector performance. Its fundamental starting point is maximum clarity about the outcomes which government is attempting to achieve, and about the relationship of outputs and activities to those desired outcomes. Often, this is linked with broader strategic planning models incorporating significant elements of private-sector corporate planning practices. Managing-for-results also tends to emphasize the *ex ante* stipulation of performance expectations for agencies, work units and individuals through the use of performance targets and standards. A standard element of the “strategic human resources management” component of managing-for-results is the introduction of stronger performance-based extrinsic incentives (rewards and sanctions) for public officials. The other crucial element is the call to “let the managers manage”—to strip away procedural controls which are seen as having encumbered management in the past.

Contemporary forms of performance budgeting represent a component of the broader managing-for-results package. However, the distinction between performance budgeting and the broader managing-for-results package remains important. Not all managing-for-results strategies are part of performance budgeting, because there are a range of other nonbudgetary uses of performance information. These include, for example, strategic planning, nonfinancial rewards and sanctions for measured agency performance,²⁵ and many aspects of the use of performance targets in human resource management. Similarly, although the input de-control theme of performance budgeting is one element of the “let the managers manage” theme, the later is about considerably more than greater flexibility in the choice of inputs. There are, naturally, crucial synergies between performance budgeting, other elements of managing-for-results, and broader elements of contemporary public sector reform (Diamond, 2003a, 2003b; Poocharoen and Ingraham, 2003, p. 181). For example, the motivational effect of agency-level output or outcome targets is likely to depend partly upon how well linked

²⁵ For example, in the British Comprehensive Performance Assessment system of performance monitoring of local governments by national government, local governments with strong performance ratings are “rewarded” with reduced central controls. Those with poor ratings may be subject to sanction ranging from intensive reviews through to imposition of new management by the central authorities.

agency-level targets are to performance targets set for individuals and work groups within the agency. It follows that the efficacy of performance budgeting is likely to depend in significant measure upon the broader reforms which accompany it.

Despite the close relationship between contemporary approaches to performance budgeting and the broader managing-for-results package, performance budgeting can only be sensibly discussed if we keep in mind that it represents a distinct element within the broader picture, the defining characteristic of which is that it is concerned with the *budgetary* use of performance information—that is, its use in budget formulation and execution. Viewed in this light, the definition used by Melkers and Willoughby as the basis of their empirical work on performance budgeting is problematic. Melkers and Willoughby define performance budgeting as “requiring strategic planning regarding agency mission, goals and objectives, and a process that requests quantifiable data that provides meaningful information about program outcomes” (2003, p. 54). This is, however, a definition which does not pertain directly to budgeting,²⁶ and which effectively equates performance budgeting with the particular approach to managing-for-results reform which has dominated the United States in recent years, particularly the Government Performance and Results Act (GPRA) (see, e.g., Radin, 1998).

E. Key Reasons why Performance Budgeting Might not Work

As outlined above, one way of viewing performance budgeting is to distinguish between two dimensions. The first is concerned with allocative efficiency in the allocation of resources between competing purposes at the level of the central budget. The other is concerned with agency performance improvement—both in productive efficiency, and also in allocative efficiency insofar as allocative choice is delegated to the agency on program level. Critics and other analysts have pointed to problems which potentially impact upon the capacity of performance budgeting to achieve its goals in respect to both of these dimensions.

In respect to allocative efficiency, performance budgeting unabashedly aims for greater rationality in expenditure planning, with the aim of allocating limited funds more effectively to the areas where they will be of greatest social benefit. Since at least the time of program budgeting, however, there have been many who have questioned the realism of this goal. In its most extreme form, the argument is that budgeting is inherently political rather than rational, and that politics will always win out, thus making the aim of rational expenditure prioritization largely an illusion. This is a view which, implicitly or explicitly, regards politics and rationality as largely antithetical. Closely linked to this is the proposition, mentioned earlier, that budgetary expenditure allocations tend to be characterized—as a

²⁶ The U.S. Government Performance and Results Acts, the key requirements of which the Melkers and Willoughby definition of performance budgeting more or less reflects, did not address the issue of the use of performance information in the budget process (although some of the legislative sponsors of the Act did emphasize links with the budget).

result of the nature of the political process and of bureaucracy—by great inertia. This, roughly speaking, is the “incrementalism” thesis.

A more common contemporary critique casts doubt on the use of performance measures and ex ante specification of desired outcomes and outcomes in order to increase pressure on agencies to perform. It constitutes, in fact, a critique of the whole venture of managing-for-results. One point emphasized is that, because of uncertainty amongst other factors, it is not always possible in the public sector to clearly specify intended outcomes and their relationship with outputs and activities. Another, more central, point is that performance measures are inherently imperfect, and that targets and “price” incentives which are linked to imperfect performance measures can potentially lead to significant adverse behavioral distortions (Smith, 1995). Thus, for example, because output measures tend not to capture the quality dimension—and are not intended to capture outcomes—a focus upon outputs in performance budgeting and management may lead to the erosion of quality and outcomes. More generally, “when the output [i.e., outcome or output] is difficult to measure, as is true in most government bureaucracies ... installation of specific goals may focus effort but may send the bureaucrats marching in the wrong direction” (Heckman, Heinrich, and Smith, 1997, p. 394). This is an argument which appears to have considerable basis not only in the academic literature (whether from public administration or from organizational economics), but also to reflect the lessons of history. A favorite theme of this school of critics is references to the Soviet central planning experience, which abounds with delightfully amusing stories about the dysfunctional ways in which planning targets were filled (e.g., the nail factory which found that it was easier to achieve its annual production target—which was specified in weight terms—by producing absurdly large and unusable nails (Nove, 1984)). It is in this vein, for example, that critics of the British target-based PSA system have attacked the Blair government for “neo-Stalinism” (Keaney, 2001).

Another, related concern, is that to link funding and, more specifically, financial incentives—at either the agency or individual level—directly to imperfect performance measures can be expected to “crowd out” the ethical/altruistic motivations which are felt by some to be crucial to good public sector performance.

These critiques in turn raise further important issues. Both the critics and the proponents of contemporary forms of performance budgeting assume that individual public officials will be motivated to score well on measures and targets. However, performance budgeting models rarely make clear just how it is that performance budgeting is supposed to impact upon the work motivation of individuals. Precisely this ambiguity was, for example, identified as a missing link in the PSA system in an examination of that system by the British National Audit Office (NAO, 2001, p. 34). Although, as pointed out above, some (certainly not all) contemporary versions of performance budgeting emphasize the notion of funding “rewards” for good agency performance, even then it is not made clear how or whether these agency-level funding rewards are supposed to motivate individuals to perform better. If the answer is through a formal link between individual performance pay and performance budgeting, then the need to adequately address concerns about behavioral distortions and the possible crowding out of “intrinsic” motivation obviously becomes particularly compelling. If the link

is not to be through “extrinsic” incentives, then exactly what is the motivational force underpinning performance budgeting?

These issues will be further discussed in the concluding section, in the light of the review of the empirical literature. They emphasize, however, the desirability in a review of the empirical literature not only of attempting to establish what evidence there is of desired behavioral impacts occurring, but also of ascertaining what the evidence is of perverse and dysfunctional effects.

III. THE EFFICACY OF GOVERNMENT-WIDE PERFORMANCE BUDGETING

A. Introduction

Assessing the efficacy of government-wide performance budgeting systems is a methodologically formidable task. Allocative efficiency, program effectiveness and productive efficiency are all subject to the influence of a multiplicity of variables, of which performance budgeting is merely one. Changes in the configuration of political forces and preferences may, for example, bring about substantial expenditure re-allocations which have nothing to do with effectiveness and efficiency. Nonbudgetary forms of managing-for-results—such as, for example, strategic human relations management involving the redesign of incentive structures facing individual public servants—may powerfully impact upon productive efficiency and program effectiveness. Features of systems of government and of budgetary institutions may greatly affect the capacity to make use of performance information to rationalize prioritize expenditure. Interactions between performance budgeting and some of these variables are so great as to make it exceedingly difficult to isolate the specific impact of performance budgeting. The measurement problems which arise are, in many cases, severe. The existence of a variety of forms of performance budgeting means, moreover, that what one should be testing is not the efficacy of “performance budgeting” in general, but that of specific forms of performance budgeting.

The magnitude of these methodological problems means that there are great limits to what can be established empirically about the efficacy of performance budgeting. The problem is, however, not an unusual one in the social sciences, where causality is complex and experimental methods generally impossible. This is, indeed, a key reason why so much reliance has to be placed on theory and policy logic to distinguish the good from the bad in public policy and management. Nevertheless, even in the face of such methodological difficulties, it is crucially important that public policy and management are to the maximum possible degree “evidence-based.” Hence the importance not only of establishing what the empirical evidence tells us about the efficacy of performance budgeting, but also of identifying ways in which empirical research in this area might be developed further in the future.

In this section the empirical evidence on the efficacy of government-wide performance budgeting systems is considered under three headings: budgetary allocation; aggregate expenditure; and productive efficiency and program effectiveness.

B. Budgetary Allocation Impacts

To what extent has performance budgeting proven itself capable of improving the efficiency of the allocation of public expenditure between competing purposes at the broad functional and program levels? Does experience show, as is sometimes suggested, that efforts to make expenditure prioritization more “rational” through performance budgeting techniques have been a waste of time? To what extent has experience validated a priori expectations that performance budgeting is more likely to be effective in improving the allocation of resources if accompanied by certain other types of budgeting or results-oriented management reforms? What has been learnt about other pre-conditions or circumstances which affect the allocative efficacy of performance budgeting? These are key questions upon which one would hope the empirical literature sheds light.

A threshold methodological problem facing empirical work on these questions is the absence of a practicable way of measuring the degree of allocative (in) efficiency in any given budgetary expenditure allocation. It is, however, possible to measure changes in the allocation of public expenditure. One might, therefore, expect to find a body of empirical work which takes observed expenditure reallocations as its starting point, and which then attempts to assess the contribution which performance budgeting to those reallocation. There is, however, almost no literature of this type. The primary focus of the literature has, instead, been upon *subjective* assessments of allocative impacts. There is, it is true, a separate body of literature on budgetary “incrementalism” which does analyze changes in expenditure allocations. However, the incrementalism literature does not attempt to assess the impact of performance budgeting on expenditure allocations. It is a concerned, rather, with testing a proposition the validity of which might lead one to conclude that performance budgeting cannot work to improve expenditure allocation, but the falseness of which would shed no light whatsoever on the efficacy of performance budgeting.

Subjective assessments of allocative impacts

Subjective assessments of allocative impacts are the focus of a body of survey and case studies literature—almost all of it American—which studies the views of public officials on the impact of performance measures on budget appropriations and on resource allocation generally.

Appendix I summarizes in tabular form the results of U.S. surveys of this type conducted since 1990. In four of these surveys, respondents were questioned about the degree to which performance measures have led to changes in budget allocations/appropriations in the government-wide budget (or its city/county equivalent).²⁷ In two of these, the majority of responses were negative. The most apparently negative impression is given by a 1999–2000

²⁷ To be more precise, one of these four (Jordan and Hackbart) asked about impacts on appropriation recommendations in the budgets proposed by Governors to State legislatures.

survey conducted by Melkers, Willoughby, and others on behalf of the GASB which found, amongst other things, that only 7.5 percent of state budget office officials (executive and legislative²⁸) believed that performance measures were “very effective” or “effective” in changing budget appropriations. On the other hand, Jordan and Hackbart’s 1997 survey of state executive budget offices gave more positive results, with officials representing 29 out of 45 respondent states agreeing that “achievement of performance standards affects budget recommendations in the Governor’s Executive Budget,” and 33 out of 46 claiming that performance affects funding in the next fiscal year. At the local government level, Poister and Streib (1999) found, in a relatively large-scale survey, that 60 percent of city managers from jurisdictions which had “centralized, citywide performance measurement systems that incorporate most departments and programs” believed that the use of performance measures had brought about either moderate or substantial changes in city budget allocations.

There are eight surveys which asked about perceptions of the *use* (as opposed to *impact*) of performance measures in decisions about budget allocations or about resource allocation generally. It should be noted that “resource allocation” decisions can cover internal budgeting as well as the allocation/appropriation of resources in the government-wide budget. Budget appropriations/allocations tend to be increasingly aggregated,²⁹ increasing the importance of internal resource allocation for allocative efficiency. Four³⁰ of these eight surveys provide information on the perceived use of performance measures in *government-wide* budget allocation decisions. The most apparently negative of these was again the 1999–2000 Melker, Willoughby et al. survey, which found that only 26.3 percent of state executive budget offices considered output and outcome measures to be either “important” or “very important” in budget appropriation decisions. At the more positive end of the spectrum were, once again, Jordan and Hackbart, and Poister and Streib. The former reported that 23 of 41 respondent state executive budget offices believed that performance indicators were an important tool in budget allocation decisions, and the latter that almost two-thirds of city managers from the type of jurisdiction described above were of a similar opinion. In a survey conducted by Lee in 1995, 30 percent and 18 percent of state budget offices asserted that “substantial use” was made of output or outcome measures respectively in the formulation of the executive budget.³¹

²⁸ In U. S. parlance, the “executive” budget office refers to the central budget office of the state executive government, and legislative budget offices refer to the staff who manage the appropriations and other budget processes of the legislatures.

²⁹ In some cases, this may take the form of single “global” appropriations to agencies, with the breakdown of expenditure by output/outcome groups (“programs”) used for planning and information rather than as a control variable.

³⁰ Possibly five—the nature of Wang’s survey is, as noted in Appendix I, not entirely clear.

³¹ However, Lee also reported some decline in the use of such information in budgeting from 1990 to 1995.

Five of the eight surveys which address the use of performance measures report results which appear relevant to their use in *internal* resource allocation within agencies. In a large-scale GAO survey of federal government managers in 2000, 43 percent of respondents reported using performance measures to a “great” or “very great” extent in making resource allocation decisions. Poister and Streib’s survey reported that almost two-thirds of respondents from the jurisdictions with “citywide performance measurement systems” believed performance measures were either important or very important for budgeting purposes. An earlier 1996 GASB survey also directed by Melkers and Willoughby was apparently more negative. It reported that of respondents from state departments which claimed to have developed performance measures, only 44.7 percent and 41.7 percent claimed that output and outcome measures respectively were used for resource allocation purposes. For municipalities, the corresponding figures were 20.1 percent and 17.7 percent, and for counties, 33.6 percent and 28.1 percent. Results of earlier surveys by the GAO in 1996–97 and by Wang in 1996 were also on the negative end of the spectrum.

Taken as a whole, the results of surveys pertaining to internal resource allocation uses of performance information appear possibly to be more positive than those which relate to resource allocation at the level of the government-wide budget, although the data is not such as to permit strong conclusions to this effect.

There appears to be very little survey or case study work of this type outside the United States. A survey of local government in the state of Victoria, Australia, by Kluvers (2001, p. 40) indicated that of those respondent entities with program budgeting systems, approximately half believed that it influenced the allocation of resources.³² Similar results were obtained in an earlier survey of the same group (Bellamy and Kluvers, 1995, p. 52).

What do these surveys tell us about the efficacy of performance budgeting? Jordan and Hackbart (1999, p. 85) suggest that the lesson is that “performance budgeting may impact the appearance and preparation of the budget document, but the outcome in terms of funding is not significantly (if at all) affected.” This conclusion does not, however, appear to be justified either by their specific results or by the results of the other surveys. Negative perceptions on the use and effects of performance measurement in budgeting reported in these surveys *might* reflect the very early state of development, or even the absence, of performance measurement and performance budgeting systems in many jurisdictions. After all, the development of good performance measures—including output and outcome measures—is a necessary precondition of performance budgeting and this is something which takes both time and serious effort. And even a formal decision by a jurisdiction to implement performance budgeting—of the type a number of U.S. states made in the mid and

³² This survey also provides results on the *use* of performance information for resource allocation purposes (Kluvers, 2001, p. 38–9), but these results are not easy to understand. A problem with these two surveys is their focus exclusively on program budgeting, although particularly at the time of the later survey other versions of performance budgeting (such as output-purchase models) were influential in Victoria.

late 1990s—does not instantly create effective links between performance information and the budget process. It is known that, notwithstanding the major resurgence of interest and effort in the development of performance measurement in the U.S. public sector during the 1990s, even as at the late 1990s many states and even more local governments lacked well-developed performance measurement systems.³³ We do not, moreover, know how many U. S. states—let alone local governments—have made explicit decisions to introduce procedures and mechanisms for the use of performance information in the budget process.³⁴ It is true that Melkers and Willoughby (1998) have reported elsewhere that by the end of the 1990s all but a handful of U.S. states had adopted performance budgeting. However, their definition of performance budgeting is, as previously discussed, problematic because it does not pertain to the budget process (budget formulation). What they were actually reporting was the adoption—in many cases late in the 1990s—of GPRA-style performance measurement and strategic planning systems.

The views of officials from jurisdictions which both (a) had introduced performance budgeting—that is, had explicitly introduced procedures or mechanisms for the use of performance information in budgeting, and (b) already had reasonably well developed performance measurement systems may be the opinions of most value. The Poister and Streib survey of local government usefully directs its question on resource allocation uses and effects to officials from those jurisdictions which claimed to have “centralized, citywide performance measurement systems that incorporate most departments and programs.” It is also the survey which reports the most positive views.³⁵

The results of a series of 17 standardized GASB case studies of the use and effects of performance measures—including the use and effects in resource allocation—in U.S. states, cities, and counties are also useful here.³⁶ The relevant findings of these case studies are

³³ It has been suggested by one reviewer of this paper that this is partly due to the hold-up in the issue of performance measurement reporting requirements by the Government Accounting Standards Review Board, with many jurisdictions taking a “wait and see” approach.

³⁴ Andrews and Hill (2003, p. 142) suggest that “although many states have adopted the PBB idea, few have moved from performance measurement to actual use of performance data in changing budgetary behavior.”

³⁵ The 1996 GASB survey directed by Melkers and Willoughby, referred to above, is also worth reconsidering in this context. The survey reported respondents who believed that output and outcome measures were used for resource allocation purposes as a proportion of those from entities which claimed to have developed performance measures. However, the survey also asked about the development specifically of output and outcome measures, and many of the entities which claimed to have developed performance measures indicated that these measures did not (yet?) include output and outcome measures. As indicated in Appendix I, if one recalculates the percentage so as to relate positive responses on the use of output and outcome measures for budgeting purposes to entities which claim to have developed output and outcome measures, the results look similar to, or better than, Poister and Streib’s.

³⁶ Although it should be pointed out that it is not clear on what basis the jurisdictions covered by the case studies were chosen, so that no claims could be made as to how representative the results are.

summarized in Appendix II. In only 6 out of the 17 jurisdictions did it appear that performance measures were significantly used in, and impacted upon, resource allocation decisions—a result not inconsistent with the more negative survey results. However, a majority of the 17 jurisdictions had apparently only quite recently (within the previous couple of years) initiated moves to use measures in the budget process, and most of these appear to have started from a weak performance measurement base. Of the eight jurisdictions which were not newcomers to performance budgeting (i.e., which had introduced it no later than the mid-1990s), six claimed that it had a substantial effect on budget allocations.^{37 38}

It is unclear how much weight one should place on these results. Opinion surveys, and to lesser extent case studies, are prone to self-reporting bias. Commenting on survey data in its 1993 report on *Performance Measures in the Federal Budget Process*, the Congressional Budget Office (CBO) suggested that “a self-reported survey is limited in its ability to provide detailed and verifiable information....surveys may be poorly suited to generating useful information in a complex, hard-to-define area such as performance measurement” (CBO, 1993, p. xi). Given that performance budgeting, and performance-based management generally, are widely regarded as leading-edge public management techniques, there is reason to fear that in these particular surveys self-reporting bias could be considerable. The GASB case studies probably reduce this self-reporting bias to some extent by asking for concrete examples of uses and impacts, but some degree of bias is nevertheless quite likely to remain.

Another concern in relation to the surveys and the GASB case studies is the focus upon the impact upon budgeting of performance *measures*, rather than of performance information more generally. It is noteworthy that the GASB case studies of two of the jurisdictions claiming to possess effective performance budgeting systems highlight the lack of a simple relationship between measures and funding. In Oregon, for example, there had initially been an attempt to establish rather mechanical linkages between *outcomes* and funding. This is an effort which theory would suggest is doomed to failure,³⁹ and it did in fact fail. However, the case study reports, “surprisingly, the use of performance measures for budget decision making did not fade away.” Instead, “the use of performance measures in the budget process seems to have evolved into one of providing elected officials with information that will assist

³⁷ These were Oregon, Texas, Iowa, Winston-Salem, Prince William County and Sunnyvale. One of the remaining two was Tucson, where it was said that measures were used in internal budgeting by departments, but not in the city-wide budget. Resource allocation use and effects in the eighth jurisdiction (Multnomah County) are not made clear by the case study.

³⁸ Moreover, of the three jurisdictions which, while only very recently having introduced performance budgeting, claimed to invested substantial effort into performance measurement over a longer period of time, two (Louisiana and the City of Austin) asserted that measures were already being used in, and were affecting, resource allocation.

³⁹ The relationship between outcomes and resources is subject to major indeterminacy due in particular to the impact of uncontrollable so-called “contextual factors.”

them in decision making and in understanding the role of various state programs in achieving the results deemed important by the legislature and the governor, not in *directly* helping establish funding levels” (Fountain, 2000, pp. 11–12).⁴⁰

Measuring the impact of performance budgeting on expenditure allocations

Remarkably, it has been possible to identify only one paper which examines the impact of performance budgeting on actual expenditure allocations. This paper, by Reddick (2003a), sets out (amongst other things) to quantitatively test the impact of “rational” budgeting upon expenditure levels in each of eight broad functional sectors (corrections, education, health etc) in the U.S. states. “Rational budgeting” is defined by Reddick as either program budgeting, zero base budgeting (two forms of performance budgeting) or biennial budgeting (a separate, albeit related, type of budget reform). The problem with this paper is that what it tests is not the relationship of “rational budgeting” to *reallocations* of expenditure between these functional areas, but the impact on the *level* of expenditure in each. The finding is that rational budgeting has been associated with the restraint of spending levels in only a minority of functional spending sectors. Reddick interpreted this as evidence of “the uneven spread of rational techniques throughout the different areas of state finance.” This appears, however, to be based upon an assumption that it is an objective of performance budgeting to restrain spending levels within each broad functional sector. As discussed in Section II, the expectation of a relationship between performance budgeting and *aggregate* levels of public spending may be a reasonable one (and Reddick’s paper, as discussed below, also addresses this issue). But it is not at all clear why one would expect performance budgeting to restrain spending in each disaggregated sector.

A paper by Brumby, Edmonds and Honeyfield (1996) examines the allocative impact of “financial management reform”—essentially performance budgeting plus a broader package of managing-for-results and related reforms—on the New Zealand public sector. The paper analyzes changes in broad functional expenditure allocations and shows that, in the period after the introduction of these reforms from the late 1980s, there were “bold reprioritization decisions” that were consistent with “quite clearly described [government] budgetary preferences.” The paper does not, however, seek to disentangle the allocative contribution of the financial management reforms (let alone of performance budgeting more specifically) from that of other crucial factors such as the political commitment to a re-thinking of the role of the state in New Zealand—and in the light of this concludes that “there is insufficient evidence, at this stage, to form a judgment concerning financial management reform’s contribution to improved prioritization.”

⁴⁰ The Texas case study, on the other hand, suggests that experience there had shown that: “It is difficult to link actual dollars allocated to specific performance measures and convincingly argue that X dollars were appropriated because of Y performance. However, there is evidence that budget decisions include discussions about agency and program performance” (Tucker, 2000c, p. 2).

Budgetary incrementalism and the allocative flexibility of budgets

Budgetary incrementalism was perhaps the most influential positive theory of public budgeting from the 1960s to the late 1980s. It was advanced by Wildavsky and others as in opposition to the classic performance budgeting models of that time, which these critics labeled “comprehensive-rational.” It remains commonplace today for incrementalism to be viewed as the antithesis of performance budgeting.⁴¹

The most common version of budgetary incrementalism was the proposition that budgeting is characterized by “inattentiveness to the (budgetary) base”—in other words, that budgetary decision-makers take the budgetary base more or less for granted as the starting point in budget formulation, and focus their attention on the size of the increment (or, occasionally, decrement) in agency or program budgets (Berry, 1990). Empirically, budgetary incrementalism has usually been interpreted as meaning that this year’s budgetary allocation to any given agency or functional spending category is strongly predicted by the previous year’s expenditure/budgetary base.

A *strong* form of incrementalism might be defined as one in which the present budget allocations are so heavily determined by past allocations that the budget is highly inflexible. Since performance budgeting is an instrument for improving budgetary prioritization, evidence of a strong form of budgetary incrementalism in jurisdictions claiming to have a system of performance budgeting would imply either that performance budgeting was ineffective, or that it was not being seriously implemented. At least to this extent, the empirical literature on budgetary incrementalism is clearly of relevance to performance budgeting.

Systematic surveys of the pre-1990 empirical literature on budgetary incrementalism, the main focus of which was U.S. federal budgeting, may be found in Tucker (1982) and Berry (1990). Prior to the 1980s, this literature came to mixed results as to the extent of incrementalism. In the 1980s, however, “most studies ...concluded that the United States federal budgetary process is no longer incremental,” and that there were in fact significant periodic shifts in budgetary expenditure allocations (Berry, 1990, p. 168).⁴² It was indeed partly for this reason that the theory of budgetary incrementalism fell out of favor, and political scientists turned their attention to explaining the causes of periodic shifts in budget allocations.

⁴¹ Andrews and Hill (2003, p. 145), for example, argue that “organizational systems ...may measure outputs or outcomes but cannot use such information within an established incremental budgeting process, because the emphasis on process has undermined the development of a decision-making process in which performance information could be considered.”

⁴² There is also a (much smaller) literature on incrementalism in British local government, almost the entirety of which again appears to have preceded 1990, and which again produced mixed results (see the review in Boyne, Ashworth, and Powell, 2000).

The post-1990 statistical literature on budgetary incrementalism is very limited. Dezhbakhsh, Tohamy and Aranson (2003) analyze U. S. federal nondefense expenditure from 1946 to 1996, firmly rejecting incrementalism. They also find statistical evidence of a number of factors which tend to trigger major shifts in budgetary expenditure allocations, including certain types of shift in political power, and public pressure over the size of the deficit. Their study is also noteworthy for its strong statistical methodology, and its persuasive demonstration that some of the classic studies by proponents of the theory of budgetary incrementalism were methodologically flawed.

By contrast, Reddick (2003b), in a study of U.S. federal expenditure over the period 1968 to 1999, found that “most of the [broad functional] expenditure categories follow a process where present expenditure is an extrapolation of past spending decision(s).” In a further study, he compared the U.S. federal budgets with those of the national governments of the U.K. and Canada over the period 1950 to 2000⁴³ (Reddick, 2002a). He concluded that, overall, “the evidence found for budgetary incrementalism demonstrated no overwhelming support.” Nevertheless, he did find a greater degree of incrementalism in the United States than in the other two countries. He suggested that this “may be attributed to the different institutions of the budget process in a parliamentary system compared to a presidential system”—a point discussed further below.

The few other recent studies of incrementalism relate to other nations or levels of government. Boyne, Ashworth and Powell (2000) test for incrementalism in 403 English local governments over the period 1982 to 1996. Their results “cast considerable doubt on incrementalism as a theory of local spending decisions.” Reddick examines U.S. *state* budgets over the period 1960–96, concluding that incrementalism had “relatively low explanatory power” in relation to the broad functional categories of expenditure (Reddick, 2003a, p. 336). However, in two other papers in which he investigates Canadian provincial budgeting, Reddick finds evidence of some degree of incrementalism (Reddick, 2002b; 2003c).

Taken as a whole, this literature does not provide the grounds for a belief in a strong form of budgetary incrementalism. Nor does the fact that there is some research supporting a milder form of incrementalism appear to be relevant to the efficacy of performance budgeting. A capacity to change expenditure allocations in response to the evidence about program effectiveness, to changing social needs, to new political priorities, or in response to a need to cut aggregate expenditure for fiscal policy reasons, does not imply that budgets need be in a state of continuous allocative volatility over time. Moreover, the more rational past policy-making and budgeting has been, the more reasonable this year’s budget will be as a starting

⁴³ 1961 to 2000 in the case of Canada, for data availability reasons.

point for formulating next year's budget. Arguably, therefore, effective performance budgeting is not incompatible with some degree of incrementalism in budgeting.⁴⁴

The view that incrementalism is the sworn enemy of performance budgeting owes much to the nature of the classic performance budgeting models critiqued by the early proponents of incrementalist theory. These tended to be "comprehensive" in the sense that they aimed at the continuous comprehensive and centralized review of public expenditure. Zero-Base Budgeting (Pyhrr, 1973), with its proposition that every program ought to be reviewed from the bottom up in each budget round, is an extreme example of this. Critics like Wildavsky (1984) pointed out, quite validly, that this was quite impractical, and that there were great limits to the allocative planning capacity of central budget decision-makers. These days, with our knowledge of the information-based failure of central planning, and with developments in the economics of information, the point seems almost an obvious one. Many contemporary models of performance budgeting have, however, moved a long way from the "comprehensive" approach, and recognize the impracticability of attempting to keep all or most expenditure under continuous review. Instead, they take a strategically selective and/or periodic approach to the review of expenditure priorities. Given this, it no longer appears justifiable to regard any element of incrementalism as incompatible with performance budgeting.

The incrementalism literature raises a number of broader methodological issues. One is that any literature which seeks to assess the degree of allocative flexibility or rigidity faces the challenge of distinguishing endogenous from exogenous changes in expenditure allocations, where "endogenous" refers to the shifts in budget allocations which would occur when expenditure policies remain stable. A pertinent contemporary example is the projected growth in pension and other age-related expenditures in many advanced countries due to demographic factors. Although there is some degree of recognition of this issue in the incrementalism literature, and in certain cases limited attempts are made to deal with it, it does not appear in general to be well handled. Recent developments in medium and longer-term fiscal projections in a number of countries—which aim precisely to quantify the magnitude of the endogenous ("constant policy") fiscal impacts of demographic and other factors—are of potential relevance to future research in this area.

A second methodological point is that the incrementalism literature focuses on expenditure allocations at a quite highly aggregated level—usually at the level of the broadest categories in the SNA classification of the functions of government, or something similar. Generally, this is for compelling reasons of data availability. However, to focus on expenditure allocations at such a high level of aggregation does potentially miss out on allocative flexibility at a less aggregated level—including in the context of internal resource allocation

⁴⁴ To express the point differently, performance budgeting is entirely compatible with a form of what some political scientists call "punctuated" incrementalism.

within agencies. As discussed below, this may be particularly problematic in the context of the U.S. political system.

Factors impacting on the allocative efficacy of performance budgeting

As discussed above, both the survey and the case study literature bear out, unsurprisingly, the importance of a developed performance measurement system for the success of performance budgeting. What else does the empirical literature tell us about factors which favor the success of the performance budgeting mission to improve allocative efficiency?

Reference was made above to Reddick's finding on the degree of incrementalism in the United States relative to the United Kingdom and Canada, and to his suggestion that this result was related to differences in the budget process between the parliamentary system and the presidential system. In a similar vein, a 1993 U.S. Congressional Budget Office report drew from a series of case studies the conclusion that "performance management systems seem to have worked best for two forms of government: on the local level, the council/manager forms ...and on the national level, governments with parliamentary systems...In both cases, the development of goals and the setting of objectives crucial to the generation of meaningful measures are encouraged by a concentration of political power in only one branch of government" (CBO, 1993, pp. 11–12).

It is regrettable that there has not been closer empirical consideration of this question—and also more empirical research on the efficacy of performance budgeting systems outside the United States—because it may be that the distribution of budgetary powers in U. S. federal and state governments makes it particularly difficult for performance budgeting to improve the budgetary allocation of resource. A distinctive feature of the U. S. system of government, at federal and state levels, is the very substantial budgetary power accorded to the legislature. This stands in complete contrast to parliamentary systems where budgetary power tends to be much more heavily concentrated within executive government. In these parliamentary systems, if party discipline is strong, the political executive drawn from the parliament (the "Cabinet") can commonly be guaranteed that its budget will receive parliamentary endorsement with little or no amendment.⁴⁵ By contrast, in the U. S. system the "executive budget" put forward by the president or state governor may be, and frequently is, changed very considerably. Legislative appropriations committees in particular exert enormous influence over budgetary allocations. In many states, agencies put forward separate budget bids to the governor and to the legislature. There are even some U. S. states where there is in fact *no* executive budget, and the budget process is entirely structured around the bids presented by agencies directly to the legislature (Rubin, 1997).

⁴⁵ Of course, this depends upon other structural features of these systems, including the extent of any expenditure power which resides with the upper house of the parliament. In many parliamentary systems, convention or explicit constitutional provisions eliminate or greatly reduce the power of the upper house to amend the budget legislation.

An obvious implication of this is that, whereas in a parliamentary system, the Cabinet is at least in principle in a position to impose a single set of priorities upon public expenditure, in the U.S. system neither the legislature nor the president/governor is in a position to do this. Performance budgeting carried out solely within the executive branch of government is likely to be largely disregarded in legislative budget decision-making. This is widely seen as a key reason why performance budgeting reform prior to GPRA failed.⁴⁶ A distinctive feature of GPRA is that it aims explicitly to deal with this problem by building consensus between legislative and executive arms on expenditure priorities by participatory processes including joint involvement in strategic planning. However, achieving such consensus is clearly an extraordinarily difficult task⁴⁷—since, as the GAO notes, consultative processes alone “cannot be expected to eliminate conflict inherent in the political process of resource allocation” (GAO, 1997a, p. 22).

A further implication of the U.S. system of government is, arguably, the way in which politics influences budgets. In any type of democratic system, of course, budgeting is necessarily and essentially political. Cabinets in parliamentary systems, being composed of politicians anxious to retain majority control of the parliament, are obviously greatly influenced by electoral considerations when making public expenditure decisions. Nevertheless, substantial *legislative* budgetary power tends arguably to place significant budgetary power in the hands of individual legislators who can be particularly susceptible to the influence of special interest groups. It may become significantly more difficult in such systems, for example, to eliminate or scale back programs which are ineffective or inefficient, but which have significant local interest group support.

These observations appear, *prima facie*, to be consistent with experience in Australia, New Zealand, and the United Kingdom over the past two decades. In each of these countries, major expenditure redirections were achieved in processes which appeared well designed to make good use of performance information. In Australia’s case, the process focused on the cabinet Expenditure Review Committee (Campos and Pradhan, 1999, p. 256), which benefited from information provided by a formal system of program evaluation. New Zealand, as referred to above, had a somewhat similar experience to Australia. In the United Kingdom, the approach has been different, centering on a Treasury-dominated Comprehensive Spending Review process implemented at a time when performance measurement and targeting was being heavily emphasized through the vehicle of the Public Service Agreements.

It was noted earlier that one impression given by the U.S. survey data was that performance information may have more impact on resource allocation decisions taken internally within agencies than upon resource allocation decisions at the level of the government-wide budget.

⁴⁶ As the U.S. General Accounting Office put it “past efforts failed to link executive branch performance planning and measurement with congressional resource allocation processes” (GAO, 1997a, p. 2).

⁴⁷ And, in the view of one well-positioned reviewer of this paper, neither Congress nor the agencies have made meaningful use of the consultation requirements.

This is an impression reinforced by a selection of recent selection of case studies of U.S. federal programs by Hatry, Morely, Rossman and Wholey (2003).⁴⁸ If true, this would not be surprising, as the aspects of the U.S. political system discussed above could be expected to have their main impact upon resource allocation at the government-wide budget level. It is precisely this which led Joyce (2003) to recently suggest that those researchers looking for resource allocation effects of performance budgeting at the government-wide budget level in the United States were looking in the wrong place. Joyce suggests that U.S. researchers need to focus much more attention in empirical work on internal resource allocation impacts.

C. Aggregate Expenditure Impacts

However limited the empirical literature on the expenditure allocation impacts of performance budgeting may be, there is much less literature on other aspects of the performance budgeting track record. The impact of performance budgeting on aggregate spending appears to be an almost completely neglected topic.

Reference has already been made to Reddick's paper (2003a) on the U.S. states, in which he tested the impact of performance budgeting on levels of state expenditure in broad functional categories. This paper also tested the impact of performance budgeting on *aggregate* expenditure, and came to the conclusion that "rational [budgeting] reforms have been successful in reducing total expenditures" (2003a, p. 336). In other words, the results suggested that states with program budgeting, ZBB or biennial budgeting were more likely, after controlling for a set of other relevant determinants of the level of aggregate expenditure, to have restrained aggregate expenditure. A potential problem with this conclusion is, however, that, in the absence of a test for the direction of causality in the relationship between "rational" budgeting and aggregate expenditure levels, the statistical evidence presented in the paper is consistent with alternative explanations (for example, that governments facing fiscal pressure are more likely adopt a form of performance budgeting and take it seriously).⁴⁹

The previously-mentioned paper by Brumby, Edmonds, and Honeyfield (1996) also considers this issue. It points to the success of the New Zealand government, in the period after the financial and public management reforms were introduced, in achieving its aim of reducing central government expenditure as a proportion of GDP. This contrasts with the seemingly inexorable growth of the public sector prior to that time. The author's conclude

⁴⁸ No claim is made by the authors of this report about the representativeness of their 16 case studies.

⁴⁹ Another of Reddick's papers purports to test a "rational choice" model, which although linked by Reddick with "rational and comprehensive approaches—such as program budgeting" (2002a, p.1, 24) is defined in such a way as to have no obvious association with any form of performance budgeting. Essentially, "rational choice" budgeting as defined by Reddick as to a form of fiscal smoothing in which governments permit the budget surplus/deficit to vary from year to year consistent with a farsighted rational expectations commitment to respecting the intertemporal budget constraint while disregarding short-run fluctuations in fiscal variables.

that “the evidence ...is consistent with, although it does not conclusively establish, FMR [financial management reform] having made it easier to control public expenditure” (1996, p. 2).

D. Impacts on Productive Efficiency and Program Effectiveness

It might be thought that a starting point for the assessment of the impact of performance budgeting upon productive efficiency and program effectiveness would be measures of improvements, if any, in output costs and outcomes achieved by jurisdictions with performance budgeting systems. The attraction of working with concrete data would appear considerable, even if one would still need to contend with the difficult task of somehow assessing the extent to which such improvements were attributable to performance budgeting, rather than to a range of other relevant factors. Remarkably, however, there appears to be almost no literature of this type. A partial exception is the Brumby, Edmonds, and Honeyfield (1996) paper, which looks at unit cost data for a sample of four significant “process” type outputs from three ministries. The data for three of these four outputs suggested significant productivity gains. The authors were appropriately cautious about what this evidence showed, explicitly recognizing the problem of a very limited sample size, and the possibility that productivity gains may have been unrepresentatively larger for more measurable outputs precisely because what is measured is more likely to be managed (Brumby, Edmonds, and Honeyfield, 1996, pp. 18–19). Once again, however, they were in no position to comment on the extent to which these productivity gains, assuming they were representative, could be attributed the overall package of New Zealand financial management reforms or more specifically to the performance budgeting component of that package.⁵⁰ This highlights some of the measurement problems confronting empirical work on the productive efficiency and program effectiveness impacts of performance budgeting. The point is not merely that there are, as is well known, significant limits on the measurability of some aspects of performance (particularly output quality and outcomes). Even where measurability is not a problem—such as in respect to the unit costs of relatively standardized “process” outputs—one needs good “before and after” data if one is interested in the impact of performance budgeting. If, as is so commonly the case, only relatively few robust performance measures existed prior to the implementation of performance budgeting, the “before” data will often be missing.

These difficulties go some way to explaining the dominance in the relevant empirical literature, once again, of work gauging *subjective* assessment of impacts. The U. S. survey literature discussed in the previous section, and some other quite similar literature, falls into this category. Importantly, however, this literature does not enquire about impact of performance *budgeting* on efficiency and program effectiveness. Rather, its focus is upon the

⁵⁰ As they noted, “the analysis of unit cost data cannot show how much of a change in trend can be attributed to change in the public management system: ministers may have been more determined to get results, or technological changes may have offered greater gain” (Brumby, Edmonds, and Honeyfield, 1996, p. 24).

use and effects of performance *measures*. Although this is not the subject of this paper, some of the more salient results of studies are worth mentioning.

The 1997 Poster and Streiber (1999, p. 333) survey of city managers reported that 46.4 percent of those from cities with citywide performance measurement systems believed that performance measures had a moderate or substantial impact in reducing the costs of services. When asked about impacts on the quality of services, the corresponding figure was 71.6 percent. These results are broadly comparable with the 1999–2000 GASB survey, which concluded that “more than half of all respondents agree that the implementation of performance measures has increased both the efficiency and effectiveness of their various governmental programs” (Melkers et al., 2002, p. 32).⁵¹ They are also consistent with the flavor of the GASB case studies of the use and effect of performance measures referred to earlier. These case studies—which include quite a few concrete examples of specific applications of performance measurement to improve efficiency and effectiveness—give the impression that, in those jurisdictions which were well advanced on the performance measurement path, uses and impacts for improving effectiveness, quality and efficiency were considerable.⁵²

Poocharoen and Ingraham (2003, p. 174) suggest that it is unfortunate that so much of the managing-for-results literature “focuses heavily on the performance measures component of managing-for-results,” because “while performance measurement is very important, it is not the only element of managing-for-results.” Given the complementarities between performance measures and other types of performance information and the managing-for-results strategies for using performance information generally, it would appear inappropriate to attempt to assess the impact of performance measures in isolation. One survey which did in essence ask respondents their views on the impact of managing-for-results as a whole was a GAO 1996-97 survey of federal managers’ view on GPRA. This found that 42 percent of respondents indicated that they believed that GPRA had improved programs to a “moderate or greater extent”⁵³ (GAO, 1997b, p. 86). GAO accompanied this survey with follow-up case studies of GPRA pilot agencies, which identified concrete examples of improved performance resulting from implementation of GPRA framework (GAO, 1997b, pp. 8–9,

⁵¹ The quantitative results of this survey in this area are not susceptible to quick summary, but may be found principally on pp. 20–21, read in the context of p. 11.

⁵² On an impressionistic count, in 7 out of 17 jurisdictions covered in the case studies the impacts of performance measures on efficiency and program effectiveness are either unclear or apparently negligible (this included two of the states—Wisconsin, Maine—and all of the city or local governments except Sunnyvale and Multnomah County). In most cases, these were jurisdictions which were at a relatively early stage in the development of measures.

⁵³ This surveyed a sample of managers and supervisors in 24 agencies of the federal government, obtaining useable responses from about 72 percent of the 1,300 persons surveyed. 41 percent of these managers expressed an opinion on the extent to which their agencies’ efforts to implement GPRA to date had improved their agencies’ programs, and it was 42 percent of these who responded as reported above.

44-46). Also relevant is the 1997 survey of U.S. state budgeters conducted by Willoughby and Melkers (2000, p. 114) which in effect⁵⁴ asked respondents to rate, on a four point scale from 1 (not effective at all) to 4 (very effective), the effectiveness of a GPRA-type package in a number of areas relevant to efficiency and program effectiveness. The mean ratings were 2.17 for “improving effectiveness of agency programs,” 1.75 for “affecting cost savings” and 1.79 for “reducing duplicative services.”

Overall, these survey results suggest a significant impact of performance measures and managing-for-results strategies generally on effectiveness and program efficiency. However, this survey data is subject to the risks of self-reporting bias already discussed.

Gaps in the literature on productive efficiency and program effectiveness impacts

There appear to be at least two notable gaps in the empirical literature on government-wide performance budgeting systems, and managing-for-results strategies more generally, on productive efficiency and program effectiveness impacts. One pertains to the issue of managerial flexibility in the use of inputs: there does not appear to be any empirical literature testing the importance of this for the realization of the intended aims of performance budgeting.

The other notable gap relates to information about unintended and perverse effects. There appears, for example, to be little evidence beyond anecdotes of the extent to which output-focused *government-wide* performance budgeting systems have actually had the quality and outcome-eroding effects which have so often been predicted. Measurement difficulties are no doubt an important part of the explanation for this. However, not even the survey and case study literature—which could circumvent the problem by obtaining practitioner impressions—appears to have addressed this important issue.

Much better evidence on these types of effects—and indeed also on efficiency impacts—is, however, available in relation to at least one sectoral performance budgeting system. This evidence is examined in the next section.

IV. SECTORAL PERFORMANCE BUDGETING SYSTEMS: A CASE STUDY

A. The Casemix Funding System

This section is devoted to a case study of a sectoral performance budgeting system—the casemix system of hospital funding. This example was chosen because casemix funding represents one of the most long-established and well-researched forms of sectoral performance-based funding. But it is also relevant to note that the casemix funding system

⁵⁴ See the comments in Section II on the definition of performance budgeting employed by Melkers and Willoughby.

has had a considerably wider influence on public sector reform. As a formularized output-based “purchaser-provider” funding arrangement incorporating the principle of “payment for results,” it was an important factor in stimulating broader interest in “internal market” models, and indeed helped to inspire the government-wide purchaser-provider budgeting model introduced in New Zealand in the mid 1990s (and a somewhat similar model introduced into Australia in the late 1990s).

Health is one of the sectors of public expenditure that has seen the most substantial long-term efforts to develop performance-based management and funding. Performance measurement and evaluation have a particularly long and distinguished history in this sector. In the 1970s pioneers of health care measurement made a particularly concerted effort to collect, categorize and interpret measures of the services provided by health practitioners (Coffey, 1999). One of the most important measurement instruments developed was the DRG (diagnostic related groups) output classification system for acute in-patient hospital treatments. DRGs are categories of patient treatment episodes which are “relatively homogenous in respect of the resources used” (Palmer and Reid, 2001)—and which are, therefore, often referred to as “iso-cost” output classes. By providing valuable information about the cost of providing the “same” type of product in different hospitals, DRGs facilitate the identification of significant inefficiency problems. As such, they were initially used, along with other performance information, as the basis for improved performance management—that is, for non-budgetary forms of managing-for-results.

From the late 1970s, DRGs began also to be used as the basis of hospital funding systems. The crucial move came when, following on from what were perceived as successful state experiments, the U.S. federal government in 1984 introduced the so-called Prospective Payment Systems (PPS) for Medicare payments to hospitals (subsequently extended to Medicaid also). Under this system, hospitals are paid fixed prices per unit of output actually delivered, with specific prices for each DRG output type. Any difference between the actual cost of treatment and the DRG price represents either a loss or profit to the hospital. The setting of the DRG price is important. For, example, if prices are set on the basis of the DRG costs of hospitals on average—approximately what happened when PPS was introduced in the United States—then hospitals which are more-than-averagely inefficient (efficient) are penalized (rewarded). To the extent that the tougher approach is taken of setting prices closer to the costs of the most efficient hospitals, the systems penalizes even average levels of inefficiency.

Since the way was paved by the U.S., output-based funding based upon DRG or similar output classifications—which will be generically referred to here as *casemix funding*⁵⁵—has

⁵⁵ The U.S. hospitals system is predominately private. These hospitals therefore did not, prior to PPS, receive *budgets* from the government. Rather, they had billed the federal government for Medicare and Medicaid services on the basis of activities delivered to the patient—essentially a cost reimbursement (retrospective) payment system. Hence the term *prospective* payments system used to describe casemix funding in the United States.

been adopted for hospital funding in many other parts of the world, amongst them Portugal (1990), Australia (from 1993), Norway (1997), Singapore (1997), and the United Kingdom (2004). There has also been increasing interest in the use of the DRG-type methodology in health service areas beyond hospitals.

The U. S. PPS system largely funds private (albeit often nonprofit) hospitals and is therefore not, properly speaking, a performance budgeting system. By contrast, many of the countries which have subsequently adopted casemix funding—including the four just mentioned—have used it as a basis for funding public hospitals, putting it squarely in the category of a sectoral performance budgeting system.

The operation of casemix funding systems in a public hospital context differs somewhat from the simplicity of the U. S. PPS system described above. Whereas the PPS provides open-ended funding—in the sense that treating additional patients always leads to the hospital being paid the appropriate price for those additional outputs—casemix funding systems for public hospitals in many cases operate in a context where fixed annual “global budgets” are paid to each hospital, with DRG output measures being used as a key determinant of those global budgets. As a consequence, public hospital casemix funding systems are in many cases a little more complex than the U. S. system, in ways which it is not necessary to explore here because they do not change the fundamental point—that casemix creates significant financial penalties and rewards for hospitals based on their success or failure in delivering outputs at a cost below the prevailing DRG “price.” This contrasts with the position prior to the introduction of casemix funding in these countries, where public hospitals typically received annual line-item budget allocations from government in a traditional public sector budget process in which there was very little relationship between funding and performance. As in traditional public budgeting generally, line-item (input category) controls in pre-casemix hospital budgeting were seen as an important instrument of overall expenditure control, notwithstanding the inefficiency to which they tended to give rise. For this reason, casemix funding of public hospitals has commonly been seen by its protagonists not only as a means of promoting efficiency through financial incentives, but as a means of permitting managers to manage more efficiently by “freeing ...hospitals from bureaucratic control and promoting autonomy by reducing restrictions on how hospital outputs should be produced” (Steering Committee on Government Service Provision, 1997, p. 61).

Throughout the world, the introduction of casemix funding was spurred by the desire to improve efficiency in order to contain the rate of growth of costs or, equivalently, to increase treatment volumes and reducing waiting times within given funding constraints. In the United States, for example, PPS was introduced to address very high rates of expenditure growth associated with the pre-PPS cost reimbursement payment system. Indeed, at the time PPS was introduced, it was projected that if rampant expenditure growth could not be reigned in, the Medicare Trust Fund would become insolvent within less than a decade. There was, nevertheless, considerable controversy at the outset as to whether PPS would achieve this goal. Although casemix funding clearly created substantial financial incentives for cost reduction, there were those who doubted the impact which financial incentives would have in

a system characterized by considerable medical practitioner power and the widespread presence of nonprofit hospitals. The degree of success of the casemix funding system in enhancing efficiency was, and is, therefore an empirical question of considerable interest.

Other fears loomed even larger at the time PPS was introduced. Many entertained a reasonable fear that although financial incentives would indeed have a major impact, the impact would be of the wrong kind. The concern was that, as a result of the imperfect nature of the output measures upon which casemix payments are based, hospitals would respond not only through improved efficiency, but also by reducing quality and thus compromising health outcomes for patients. In the first place, because casemix funding pays hospitals on the basis of output quantity, with no quality dimension to the output quantity measure, and no supplementary payments for output quality or outcomes, hospitals could benefit financially through:

- Under-provision of services to patients in general,
- Underinvestment in technology which improves outcomes and quality *but raises costs* (as opposed to cost-saving technology, which is encouraged by PPS), and
- Reduction in research with the longer-term potential to improve outcomes.

In the second place, the presence of varying degrees of “heterogeneity” in the DRG groups gives rise to a further set of quality concerns.⁵⁶ Even though the fundamental principle of DRG classification is to define “iso cost” treatment/patient groups, the variability of patient conditions and characteristics means that, within most DRG group, there inevitably remain significant variations in the appropriate cost of patient treatment. Thus, for example, in the early DRG classifications, each treatment of a hip fracture patient was classified to the same DRG group—it was, in other words, treated as the same type of output as all other hip fracture patient treatments. However, elderly hip fracture patients can be expected on average to cost significantly more to treat than younger patients. If the DRG price is set at or below the cost of treating the average fracture, hospitals could expect to lose money when treating such relatively high-cost patients within the DRG group. Conversely, the younger patients would tend to be profitable. These types of cost and profitability variations within the DRG classifications would create financial incentives for hospitals to engage in the following types of dysfunctional behavior (Allen and Gertler, 1991; Ellis, 1998):

- “Skimping”—under-provision of services to the relatively high-cost patients within each DRG payment category (e.g., not giving elderly hip fracture patients the full treatment they need),

⁵⁶ Particularly in the United States context of multiple health payers and private hospitals, there were fears that, in addition to the feared erosion of quality and outcome, casemix funding might lead to cost-shifting to other health costs payers (e.g., private insurers), and cuts in “uncompensated care”—i.e., reductions in pro-bono treatment of uninsured persons.

- “Dumping”—the avoidance of patients whose expected costs of treatment exceed the relevant DRG price (by explicitly or covertly avoiding admitting them for treatment, or by inappropriate transfers to other hospitals),
- “Creaming”—over-servicing of patients who are expected to be relatively low-cost and whose treatment is therefore expected to be profitable (including encouraging admission to hospital of patient who do not actually require inpatient treatment because, for example, of the mildness of their symptoms).

Note that the latter two forms of behavior are often amalgamated together under the label “cream-skimming.”

As a result of these fears, casemix funding systems have always been accompanied by quality-oriented regulatory structures, particularly in the form of monitoring regimes to detect inappropriate conduct on the part of hospitals. In addition, the pure output-payment principle was from the outset qualified at the margins through devices such as additional input-based payments for anomalous “outlier” patients who for special reasons require treatment the cost of which greatly exceeds the relevant DRG price. There have, also, been progressive improvements to the casemix output classification, with a particular emphasis upon the reduction of heterogeneity within each output group (so that, for example, the hip fracture output group was eventually disaggregated into three or more groups which were each more “iso cost” than the original single hip fracture DRG had been). Nevertheless, as important as these safeguards were, they were far from a comprehensive or watertight response to the quality/outcomes problem. Whatever way the issue was approached, managing-for-results in the hospital sector had to contend with an environment of inherently imperfect performance measures.

Fears of what would happen to quality and outcomes under the casemix funding system are representative of the concerns about perverse and unintended consequences which, as discussed in earlier sections of this paper, are so often raised in relation to performance budgeting systems and managing-for-results more generally. The extent to which casemix funding does in fact generate such adverse quality/outcomes effects—and, if so, whether they are serious enough to outweigh efficiency improvements⁵⁷—is therefore a crucial empirical question with much broader relevance to public management.

B. The Empirical Evidence

Although the U. S. PPS is not actually a performance budgeting system, the fact that even a “private” hospital sector has, arguably, many quasi-public characteristics makes it instructive

⁵⁷ It is important to remember that even some erosion of quality and outcome in the treatment of the average patient would be an acceptable price to pay for a sufficiently large efficiency improvement, given that inefficiency imposes a heavy price on society—either through a need to ration health services more than would otherwise be necessary, or through the opportunity cost of the additional tax burden.

to review the U. S. experience prior to looking at other countries. An additional consideration here is that there is once again, considerably more empirical literature on casemix funding in the U. S. than in other countries.

The large empirical literature up to 1991 on the impact of PPS in the United States was thoroughly reviewed and synthesized by Coulam and Gaumer (1991). In summary, this literature demonstrated that the system had been highly successful in improving hospital efficiency and reducing the rate of increase of Medicare health costs. In respect to feared adverse consequences, it showed that to that time “none of the worst fears raised at the outset have been borne out by experience.” More specifically, there was no evidence of any significant deterioration of health outcomes,^{58 59} nor of other feared perverse effects more specifically relevant to the U.S. health system⁶⁰ (Coulam and Gaumer, 1991, pp. 62–70. As for impacts on technology and research, these were more difficult to test, but the empirical literature did at least demonstrate that there had not been a “large and systematic reduction in the rates of adoption of new technology” (Coulam and Gaumer, pp. 63-64). Reinforcing the impression of cost containment driven by efficiency rather than quality reduction, a survey of 10 pairs of hospitals which had been “winners” and “losers” in profitability terms under PPS in the 1980s found that “in the majority of losers, interviewees representing a wide cross-section of staff ...stated that staffing was excessive and that reductions could be made without sacrificing quality of patient care,” while “staff at winner hospitals generally related lean staffing that was perceived to be about right” (Bray et al., 1996).

There appear to be relatively few empirical studies of the impact of PPS relevant to the period since Coulam and Gaumer’s literature review.⁶¹ Most of the pre-1991 studies adopt a

⁵⁸ In the context of a review of the findings on quality, Coulam and Gaumer noted that there was quite a considerable body of evidence suggesting that there had been “an increase in the degree of [health status] instability of patients at discharge after PPS was introduced” (68). They noted that this raised quality concerns, but observed that discharge planning had been a long-term problem prior to PPS. Perhaps more significantly, they noted that “the impression one gets from this research is that, by reducing the hospital role in patient episodes, PPS has placed a premium on what happens after the hospital stay.” It has since become clear that PPS drove a major re-allocation of services, under which hospitals focused upon “treating only the most seriously ill patients,” and there was a huge growth of outpatient, home care, nursing home and other ancillary services to care for post-discharge patients and patients who now are not treated in a hospital setting at all (Coffey, 1999). This means that, to a large degree, earlier discharge arises from the fact that no longer are hospitals treating large numbers of patients who could be more cost-effectively be treated elsewhere. Insofar as this is the case, this re-allocation of services has been a major source of improvements in the efficiency of health care.

⁵⁹ Additional support for this conclusion came in a 1993 paper by Manton, Woodbury, Vertrees, and Stallard, which examined mortality rates in the 12 months before and after the introduction of PPS.

⁶⁰ Namely, cost shifting, and cuts in uncompensated care (Coulam and Gaumer, 1991, pp. 64–70).

⁶¹ This impression receives additional confirmation from the fact that the almost all of the literature reviewed in a 2001 review of the PPS literature (Rosenberg and Browne, 2001) pre-dates 1991.

“pre/post implementation comparisons,” comparing rates and trends before and after the introduction of PPS. Over time, with the accretion of other confounding factors, this methodology became increasingly inappropriate. More recent studies tend therefore to adopt different approaches. Thus, for example, a recent empirical examination by Boccuti and Moon (2003) of Medicare’s long-term cost containment record adopted the approach of comparing the rate of Medicare cost increases with those of private insurers, who still predominantly use a cost reimbursement approach. Their conclusion was that, comparing like with like, “Medicare has proved to be more successful than private insurance in controlling the growth rate of health care spending per enrollee.” It has been possible to identify only one post-1991 paper on quality/outcome effects of PPS—a paper by Meltzer, Chung, and Basu (2002) which find some evidence of “skimping” in a study of California patients.⁶² The predominant contemporary view of PPS seems to be that “overall ...the system has worked pretty well in establishing better incentives without having a systemic adverse effect on quality.”⁶³

Outside the United States, studies of the impact of casemix funding are much more limited. In respect to Portugal—which, as noted above, was one of the earliest nations outside the United States to adopt casemix funding⁶⁴—Dismuke and Guimaraes (2002) studied mortality rates for the most frequent non-obstetric DRG category before and after the introduction of casemix funding. They concluded that “no evidence was found that the case-mix based payment system ...has had a pernicious effect on hospital quality as measured by hospital mortality.” In a separate paper, Dismuke and Sena (1999) found that the casemix funding system had a positive effect on productivity in the use of three of the most common diagnostic technologies.

Two studies are available which are relevant to the introduction of casemix funding in the Victoria, which the first Australian state to adopt the system (in 1993). One of these presents results which are “quite consistent with the expectation that casemix funding is providing hospital with appropriate incentives for cost efficiency” (Young and Harris, 1999, p. 14), although the methodology employed was, the authors acknowledged, not well suited to assessing the specific impact of casemix funding on efficiency.⁶⁵ The other study was

⁶² Meltzer, Chung, and Basu also refer to an earlier Newhouse paper as providing evidence of “patient selection according to profitability.” In fact, Newhouse’s paper (1989, p. 41) points to “mixed” evidence, which “is as yet too weak to conclude that it is a serious problem.”

⁶³ Mr. Glenn Hackbarth, Chairman of the Medicare Payment Commission, which advises the U.S. Congress on Medicare issues, quoted in Tieman (2003).

⁶⁴ See Wiley (1992) for the background to this.

⁶⁵ The paper applies stochastic frontier analysis—which is suitable for identifying *relative* efficiency levels between producers at a point in time—to post-casemix funding (1994/94) data. It finds very low differences in efficiency levels between hospitals. The inference is then drawn that the low differences in efficiency levels suggest a high level of overall absolute efficiency (“it is difficult not to conclude from the evidence presented that Victorian hospitals are remarkably efficient”). The authors explicitly acknowledge that “we cannot be sure

(continued)

conducted by the Victorian Auditor-General. This found, firstly, that the goal of “substantial efficiency gains ...has been effectively met” (Auditor-General of Victoria, 1998, p. 5). It pointed in this context to official cost data showed quite dramatic cost reductions, and also to other evidence of changes in treatment patterns which suggested that these cost reductions were associated with significant increases in efficiency (1998, pp. 187–222). On the other hand, the Auditor-General considered that part of the price for these dramatic savings had been an “overall decline in the quality of care” (1998, pp. 5, 65). This conclusion on quality was methodologically questionable—even though plausible and supported by other anecdotal evidence—as it was based upon opinion surveys of hospital administrators, hospital medical staff and other parties whose interests were, in many cases, adversely affected by the reforms. There appears, unfortunately, to be only one empirical study (Brown and Lumley, 1998) relevant to the impact of casemix funding on quality in the Victorian hospital sector. This study, based on a patient satisfaction, provides a small degree of support for the quality erosion thesis.⁶⁶

Whatever the reality about impacts upon quality may have been in Victoria, the prevalence of continuing major *concerns* about quality erosion some time *after* the introduction of casemix funding stands in contrast to experience elsewhere. What was, however, distinctive in Victoria was that the introduction of casemix funding coincided with substantial immediate cuts to overall levels of hospital funding, driven by the imperative of state-wide fiscal consolidation (Steering Committee on Government Service Provision, 1997, p. 43). Under these circumstances, it is very difficult to distinguish quality impacts arising from the casemix systems from those arising from severe funding cuts. In the United States the position had been quite the opposite—PPS had been introduced with payment rates based broadly on existing treatment cost levels, and it was only after a few years that there was a tightening of rates. The story has been quite similar in most other countries where the system was introduced as the basis for public hospital funding—that is, the changeover was typically designed to be initially cost neutral, with savings then being realized over time as efficiency improved.

of any causal connection [between efficiency and the introduction of casemix funding] since we do not have any comparative data from other funding systems in Victoria or comparative data prior to 1993” (14). The other problem, again recognized by the authors, is that the methodology takes no account of possible quality erosion. To make these points about the methodology of the paper is not, it should be emphasized, either to question the quality of the research, or to question the plausibility of the proposition that casemix funding plus major funding cuts forced efficiency improvements in Victorian hospitals.

⁶⁶ The study reported patient satisfaction surveys, conducted in 1989 and 1993 respectively, of women who had just given birth. There was a doubling of critical comment on the quality of care, leading the authors to conclude that “standards of care are being compromised in the current economic and policy environment.” The 1993 survey was, however, conducted a mere 3 months after the introduction of casemix funding.

The literature thus appears to suggest that, despite creating financial incentives for the sacrifice of quality and outcomes, the casemix funding system did not generally have that result, at least unless it was combined with immediate severe funding cuts. The question is, why not?

Part of the answer undoubtedly lies in the quality-oriented regulatory structure which has accompanied casemix funding systems. This structure has been enhanced over the years, and has been supported by significant improvements in quality-related performance information and management processes (e.g., quality assurance, accreditation, even in a few instances the public release of information about hospital and practitioner outcomes). The role of these mechanisms is an example of the synergy between any good performance budgeting system and the broader performance management framework. However, this probably provides only part of the answer. After all, not only are these regulatory structures for quality maintenance imperfect even today, but they were far more imperfect in the early years of PPS.

The other relevant factor is that financial incentives such as those introduced by casemix funding do not operate in isolation, but instead interact with other behavioral drivers, including socialized, value-based drivers. In the health sector, professional ethics and commitments to maintenance of quality care play a very important role. Indeed, many of the quality-related processes which have spread over recent decades are designed expressly to mobilize and strengthen these forces (e.g., peer review processes). It is simplistic to predict perverse and unintended consequences from the linking of financial incentives to imperfect performance measures without taking into account the potentially important counterbalancing role of these other behavioral drivers. Indeed, as Coulam and Gamuer (1991, p. 65) put it, “PPS in effect viewed hospitals, physicians, and others as buffers between the purely financial incentives of PPS and the patient needs for quality care.” Unfortunately, there does not appear to be any empirical literature which attempts to assess the extent of the contribution of socialized, value-based behavioral drivers in safeguarding quality following the introduction of casemix funding. But the impression remains that they may have played a major part in the benign impact of the casemix funding system.

None of this is to suggest that casemix funding—or any other performance budgeting system focused exclusively upon output quantity, with no funding rewards for outcomes or output quality—should be viewed as performing strongly in terms of quality and outcomes. It merely suggests that casemix funding is no worse for quality and outcomes than more traditional modes of hospital funding, which are generally agreed to have performed badly in these areas. It is therefore not surprising that a major contemporary theme of health sector reform around the world is the need to considerably strengthen quality management. In this context, there are many today who suggest that casemix funding needs to be adjusted to incorporate explicit rewards for quality (e.g., Berwick et al., 2003; Salber and Bradley, 2002). It is also widely argued that appropriate forms of competition, combined with the provision of better information on quality and outcomes to patients and payers, has a greater role to play in future in creating incentives for quality than it has generally played in the past.

With these qualifications, it can be said in summary that there are good empirical grounds for viewing this specific form of performance budgeting in the hospital sector as a clear success, delivering strong efficiency gains without any demonstrable adverse effect on quality and outcomes.

The success of the casemix funding system has been built upon the twin pillars of performance information and clear procedures and mechanisms for the use of that performance information in funding decisions and other managerial applications. Health sector experience elsewhere seems to confirm that to introduce performance management and budgeting without a large investment in improved performance information is to invite disappointment. Thus, for example, the prominent health economist Alain Enthoven (2000) has identified poor information as one of a handful of key factors⁶⁷ explaining the relatively disappointing impact of the 1980s internal market reforms of the British National Health Service. However, the casemix experience also suggests that it is not necessary to wait for performance measures to be near-perfect between putting them to use. As Coffey (1999) observes with particular reference to the early days of PPS when performance measures were considerably more imperfect than they are today, “even with seemingly inadequate databases, Medicare’s casemix reimbursement ...had profound effects.” It certainly cannot be said that performance budgeting has only worked in the hospital sector because performance measurement was easy. Quality and outcome measurement is not necessarily any less difficult than for much of the rest of the public sector. Moreover, hospital outputs, while not as extremely heterogeneous as some other public outputs, are very far from being standardized mass-manufactured “widgets”. What is so impressive about casemix funding is precisely the success achieved by a strong form of performance budgeting even in the face of considerable performance measurement difficulties.

C. Implications for Sectoral Performance Budgeting Systems More Generally

How relevant is the experience of the casemix funding system to the public sector more generally? Clearly, casemix funding is merely one example of a sectoral performance budgeting system, and there is no reason to assume that experience of casemix funding is representative of all other such systems. It is more reasonable to assume that the efficacy of performance budgeting—and also the most appropriate model of performance budgeting—will vary from one sector to another, depending upon a range of considerations. What works

⁶⁷ “Before [the internal market] reforms there were no systematic reliable data on the costs of services. The government did launch efforts to improve this situation, including directives on costing for contracting and the Casemix Office to develop a British counterpart to United States diagnosis-related groups (DRGs). However, progress has been slow. Such cost data were not available during the internal-market period ...The importance of good information on quality and cost is not limited to market models. It is essential to any properly managed system, but especially to a centralized system. Without it, the center would have not way of recognizing an improvement if one were to happen” (Enthoven, 2000).

best for hospitals will not necessarily work best for education, or for police, or perhaps even for other areas of the health sector.

What is least surprising about the casemix funding record—at least viewed with the benefit of hindsight—is its success in cost containment. It is the apparent absence—or at very least the almost complete lack of evidence—of perverse effects arising from the use of imperfect performance measures which is much more remarkable. There is no reason to assume that this is representative of other sectoral performance budgeting (and managing-for-results) systems elsewhere in the public sector. There is, in fact, some empirical evidence of such perverse effects in at least two other parts of the public sector—school education and labor market programs⁶⁸ (evidence succinctly summarized in Propper and Wilson (2003) and Courty and Marschke (2003)). The stark historical record of such effects arising from target-setting in Soviet-style central planned economies (Nove, 1984) is also persuasive on this point. It would therefore be inappropriate to draw from the benign casemix funding experience the general conclusion that concerns about the perverse effects of the use of performance measures in the public sector are without foundation.

A more plausible interpretation of the casemix experience, however, is that it suggests that the extent of such perverse effects may be substantially mediated by the presence of an altruistic commitment on the part of public sector employees to the maintenance of service quality. That is, workers will not take advantage of imperfections in performance measures to the extent that pure self-interest might suggest. If so, then it is reasonable to assume that the strength of altruistic commitment to service quality varies between sectors, and that it is particularly strong in the hospital sector because of the extent of professional socialization and perhaps also because of the particularly serious ramifications for clients of lapses in quality of care. If so, this would explain the casemix funding experience. It would also predict, however, that in many other parts of the public sector, expectations of the magnitude of perverse effects which are based upon the assumption of pure self-interest may be somewhat exaggerated.

Another possibility suggested by the casemix experience is that it is a mistake to think of performance budgeting or other managing-for-results systems only as a means of constraining self-interested behavior which undermines performance. It may be that successful managing-for-results systems also generate substantial benefits by helping to reshape and balance the altruistic commitment to service quality held by public sector employees. This rests upon the proposition that the conception of service quality held by public sector employees may often tend to be myopic and flawed. They may, in particular, be unduly focused upon the perceived needs of individual clients and give too little weight to overall cost-effectiveness. In a hospital context, for example, the patient-focus of treating practitioners may, if unchecked by cost considerations, create a systematic bias towards excessive duration of treatment and other forms of over-treatment notwithstanding that,

⁶⁸ More specifically, programs established under the U. S. Job Training Partnership Act.

viewed in a broader perspective, this may be contrary to the interests of patients in general because it might simply lengthen waiting lists for treatment. Casemix funding changed this because it “created a natural language for communicating to the medical community the financial implications of clinical decisions” (Averill et al., 1996). It is relevant in this context that empirical evidence suggests that hospitals which responded successfully to PPS in the United States were those which were “more likely to share responsibility for financial decision-making and information about hospital financial performance with physicians” (Bray et al., 1994).

There would appear to be little empirical evidence on these points in the public sector more generally. However, research on performance-based funding in a major U. S. labor market program (JTPA) provides at least a morsel of relevant evidence. Work by Heckman and others indicates that caseworkers in JTPA training centers tend to “manifest strong preferences for serving the most disadvantaged (and least employable).” They find, firstly, that this altruistic commitment acts as a significant counterweight to the “cream-skimming” (creaming and dumping) incentives created by the funding formula (Heckman et al., 1996). However, from the perspective of policy-makers, the extent of the case-worker bias towards the most disadvantaged is a problem, because it leads to too large a portion of resources going to clients who are, on average, least likely to succeed in re-entering the labor market even with substantial training assistance (Heckman, Heinrich, and Smith, 1997, p. 392). This is where the performance-based funding system plays an important role. Heckman, Heinrich, and Smith (1997, p. 393) find that in one job training center which they view as representative, “in this training center and, we conjecture, in most others, performance standards operate as a partial check on the preferences of case-workers with a social-worker mentality.” In other words, the existence of performance measures linked to center funding induces caseworkers to moderate the degree of their bias towards the least employable. This analysis suggests that altruistic motivation and managing-for-results systems operate, in this instance, synergistically, each moderating the potential excesses of the other.

All this suggests that the interaction of altruistic (“intrinsic”) and self-interested motivation is a matter of considerable importance for the efficacy of performance budgeting and managing-for-results more generally. It is a matter which is considered further in the next (final) section of this paper.

V. CONCLUSIONS AND REFLECTIONS

A. Overview of Findings

The principal aim of this paper has been to ascertain what light the empirical literature sheds on the efficacy of performance budgeting and whether, more specifically, that literature supports the contention that efforts to link funding to results in government budgeting have failed.

Regrettably, the empirical literature on *government-wide* performance budgeting is disappointingly limited in scope and methodology, and does not provide a basis for strong

conclusions about the efficacy of these systems. Nevertheless, it does appear to provide some support for the proposition that, where the necessary investment in the development of performance measurement and other performance information has been made, it is possible to use that information in budgeting to improve both allocative and productive efficiency. It certainly cannot be said that the empirical literature demonstrates the failure of performance budgeting.

The empirical literature on at least some sectoral performance budgeting systems provides clearer results. There is quite strong evidence of the efficacy of the sectoral system chosen for detailed discussion in this paper—“casemix” funding of hospitals. Not only has this particular output-based funding system delivered strong productive efficiency gains, but it appears to have done so with surprisingly little in the way of perverse and unintended effects.⁶⁹ The casemix funding example illustrates the crucial role of performance measurement, management accounting and the relaxation of input controls as underpinnings of successful performance-based funding.

A disappointing feature of the literature on government-wide performance budgeting systems is the dearth of empirical studies of non-U. S. systems. It is paradoxical that the nation which has historically led the way on performance budgeting—the United States—is one in which the system of government is, arguably, relatively unfavorable to the use of performance information in allocative decisions in the annual budget. There is reason to believe that, at the level of budgetary allocations, performance budgeting is potentially more effective in nations where the concentration of budgetary power in the executive arm of government permits unified expenditure prioritization. Nations such as Australia, New Zealand and the United Kingdom have all over the past couple of decades undertaken major policy-driven budgetary expenditure reallocations in the presence of explicit procedures and mechanisms for linking funding and results. Although there has been little close empirical study of the role of performance information in these expenditure reallocations, there is reason to believe that this role may have been substantial.

There are a number of distinctive features of the literature reviewed in this paper. One is that much of it is primarily focused upon the role of performance *measures*, rather than of performance *budgeting*, and is relevant to performance budgeting only insofar as it addresses the budgetary uses and impacts of performance measures. While this is useful, it can potentially be misleading because performance budgeting is concerned with the budgetary use not simply of performance measures, but of performance information generally. This is particularly true of performance budgeting viewed as a tool for improved allocative efficiency. Although performance measures are a fundamental tool of performance budgeting, even a well-developed system of performance measurement has inherent limitations which mean that, although allocative choices can be informed by, but never determined by, performance measures. Regrettably, simplistic notions of the nature of the

⁶⁹ At least when not accompanied by excessively severe immediate cuts in hospital funding levels.

potential link between measures and budgets are not uncommon, and have to some extent affected the empirical literature.

The literature has also been characterized by divergent and occasionally unclear definitions of performance budgeting, and has in some cases been a little too ready to accept at face value the claims of some jurisdictions to have introduced performance budgeting. Future researchers attempting to assess the impact of performance budgeting in any given jurisdictions may be well advised to give more consideration to how serious the jurisdiction's commitment to performance budgeting has actually been.

A third feature of the literature on government-wide systems is that strikingly little use has been made of actual expenditure *data* to assess allocative impacts of performance budgeting, and of output/outcome measures to assess impacts on productive efficiency and program effectiveness. Opinion surveys have, instead, been the primary research tool. Methodological and data availability problems provide part of the explanation for this limited use of expenditure and performance data. However, notwithstanding these difficulties, there would appear to be scope for more future data-based empirical research.

A threshold challenge in studying the allocative impact of performance budgeting is distinguishing between policy-driven and endogenous changes in functional and program expenditure allocations. Recent advances in “constant policy” fiscal forecasting and projection methods are potentially useful in this context. The challenge of distinguishing policy-driven from endogenous changes is greater when examining medium and longer-term changes in expenditure allocations, because many endogenous changes take place only gradually over time. In the short run, the problem is less severe, and pertains mainly to the cyclical components of expenditure. This is another reason why the study of episodes of substantial short-run budgetary expenditure reallocation—such as those experienced by Australia, New Zealand and the United Kingdom—may prove particularly useful.

The task of disentangling the impact of performance budgeting on policy-driven budget reallocations from that of other factors, such as changes in the configuration of political forces, remains a challenging one. It is questionable how far statistical methods can help. Case studies of concrete examples of the uses and impacts of performance information on major expenditure reallocation decisions are perhaps more likely to be fruitful.

In respect to impacts on productive efficiency and program effectiveness, it is probably not practicable to seek to distinguish the impact performance budgeting from that of managing-for-results reforms more generally. However, as the quality of performance measures across the general government sector continues to improve, it will at least become increasingly possible to base analysis of the impacts of managing-for-results upon actual measures of efficiency and effectiveness. With progress in measurement of outcomes and quality—including through the steadily widening use of instruments such as client satisfaction surveys—empirical work to test the extent of perverse and unintended effects arising from target-setting should also become possible.

It was also hoped that the review of the empirical literature would shed light on the issue of behavioral distortions arising from the use of imperfect performance measures and performance information generally in performance budgeting systems. This is, as noted previously, an issue of considerable importance not only to performance budgeting, but to managing-for-results more generally. Regrettably, it is an issue upon which, again, the literature on government-wide performance budgeting sheds little light. However, the empirical literature on sectoral systems is rather more useful. The case study explored in depth in this paper—casemix funding in the hospital sector—is notable for the extent to which what appeared, *ex ante*, to be quite reasonable fears about the dysfunctional and perverse responses to the introduction of significant financial incentives linked to highly imperfect performance measures were *not* realized. This is not taken as suggesting that on that these fears are generally without foundation. There is, as noted in Section IV, some empirical evidence in broader literature on sectoral systems of performance budgeting and managing-for-results of perverse behavioral responses to the choice of performance measures. However, this literature does not attempt to assess the net impact upon performance of these systems—that is, to assess whether any performance improvements outweighed such perverse behavioral responses or not.

The question of the efficacy of performance budgeting—and that of the magnitude of any perverse and dysfunctional responses—arguably need, however, to be put in a broader behavioral context. What follows, by way of conclusion to this paper, is some reflections of certain underlying behavioral issues.

B. Behavioral Foundations of Performance Budgeting and Managing-for-Results

A weak link in the literature on performance budgeting—and, indeed, in much of the literature on managing-for-results generally—is the connection with behavioral theory. As noted previously, given the strong contemporary tendency to view performance budgeting as an instrument for encouraging or pressuring agencies to improve their performance, there is a need for a more explicit articulation of precisely how it is that performance budgeting systems are assumed to motivate individual public officials to perform better. For example, is the motivating force seen as arising from some type of link between performance budgeting and stronger results-based extrinsic incentives for individuals (e.g., performance pay)? If not, what exactly is the mechanism?

This points, however, to a more fundamental point: that effective performance budgeting systems, and managing-for-results strategies more generally, will be those which reflect an accurate appreciation of what motivates people in an organizational context, and particularly in a public sector organizational context.

Public sector management has over recent decades been heavily influenced by the belief that human organizational behavior is fundamentally self-interested, and that this is as true in the public sector as in the private sector. There has, as the social economist Le Grand (1997, 2003) observed, been a “fundamental shift in policy makers” beliefs concerning human

motivation and behavior,” in which the self-interest postulate has replaced the earlier tendency to assume that those who manage and deliver public services were “public spirited altruists.” Economists have been very influential in this, not only because of the growth of the sub-discipline of organizational economics (of which agency theory is the most influential strand)—which has been overwhelmingly predicated on the standard homo economicus behavioral assumption—but also because of the large body of economic literature over recent decades applying the self-interest postulate specifically to the analysis of the public sector.

There is a widespread belief that the corollary of the self-interest postulate is that the best way of improving public sector performance is to create much stronger extrinsic incentives (rewards and sanctions) for measured individual performance relative to clearly-defined objectives. This type of thinking has been reinforced by the long-standing tendency of many to look to business and market models as the basis for the re-design of the public sector. It has been an important influence on the way many think about managing-for-results systems, and has also impacted on approaches to performance budgeting—most notably in the idea that performance budgeting should involve funding “rewards” to agencies for good performance.

Such thinking is, however, somewhat problematic. In the first place, it does not necessarily follow from the self-interest postulate that the best way of motivating agents is to make maximum use of high-powered extrinsic incentives. Organizational economics has analyzed extensively—albeit deductively more than empirically—the perverse and dysfunctional responses to such incentives which can be expected of maximizing agents under circumstances not only of poor performance measurability, but also of uncertainty, and the presence of multiple goals (“multi-tasking”). It is a well established conclusion of this literature that the greater the extent of these types of problems, the more reliance should in general be placed upon lower-powered, “fuzzier”⁷⁰ extrinsic incentives (career systems, use of subjective evaluation rather than formularized results-based payments, bonuses based upon team rather than individual performance etc.),⁷¹ or even upon fixed remuneration combined with intensive monitoring (Gibbons, 1998, p. 120). It is, moreover, widely acknowledged by organizational economists who have given specific consideration to the public sector that these types of problems tend to be particularly severe in the public sector, and that in addition there are further more specifically public sector problems—such as the presence of multiple principles—which are of relevance. From this it follows that the optimal role for high-powered extrinsic incentives is generally less in the public sector than in the private sector (Tirole, 1994; Dewatripont, Jewitt, and Tirole, 1999; Dixit, 2002; and Grout and Stevens, 2003). These points have long been familiar to public administration scholars,

⁷⁰ In Kreps’ (1997) terminology.

⁷¹ See Prendergast (1999) and Gibbons (1998) for excellent overviews of this literature.

who have also recognized other broader considerations such as the presence in some parts of the public sector problems of significant “goal ambiguity.”⁷²

In the second place, the behavioral postulate of pure self-interest is certainly wrong, and particularly wrong in the public sector. Public administration scholars have long correctly emphasized the importance of what is usually referred to as “public service motivation”, often conceptualized as an internalized commitment to a “mission” (Wilson, 2000). Even in a private sector context, the psychology literature⁷³ recognizes that “intrinsic motivation”—in particular, the feeling that work is important and gives a sense of accomplishment—is important. Both the psychology and management literature recognize that “organizational commitment”/“goal alignment” (of which intrinsic motivation is an important component) and, where relevant, professional ethics are important determinants of the productivity and approach to work of employees. Moreover, the empirical literature⁷⁴ provides substantial evidence that intrinsic motivation is significantly more important in the public sector: government employees are on average significantly more driven by a sense of mission than are private sector employees, but also that amongst public sector workers, “service-oriented” employees are “more productive than economic-oriented employees” (Crewson, 1997, p. 515).

A crucial conclusion of much of the literature on intrinsic motivation is that excessive reliance upon extrinsic incentives can actually reduce (“crowd out”) intrinsic motivators, resulting in worse rather than better performance (e.g., Osterloh and Frey, 2000). The more important intrinsic motivators are as a source of organizational performance, the greater the risk of this occurring. In a public sector context, this reasoning has been advanced as an argument not only for against continued reliance upon relatively low-powered incentives—in particular upon career-building motivation—but also as a consideration suggesting that the optimal degree of outsourcing of the production of publicly-funded services is less than might otherwise be imagined (Arnett, 2003; Martin, 1994; Sandmo, 2002).

Economists have, in general, paid little attention to these themes, and have for the most part continued to model organizational behavior—including within the public sector—on the

⁷² Tankersley and Grizzle (1994, p. 8) define goal ambiguity as circumstances where “some public projects are undertaken on a best-efforts basis when a problem is known to exist, but where it may be unknown if anything can be done about”—or, one might add, *how* the problem is best addressed. A good example may be public policy directed at the reduction of rates of unwanted teenage pregnancies. In response to the setting of targets for this objective under the British PSA, one British public health administrator recently wrote that “a recent systematic view concluded that we simply do not know how to do this: none of the current strategies work” (Goodman, 2002, p. 567). Even in the face of this type of ambiguity, governments often persist with efforts to address the problem.

⁷³ See literature summarized in Crewson (1997, p. 507–8) and Osterloh and Frey (2000).

⁷⁴ See, for example, the literature reviewed in Wright (2001), the literature cited in Francois (2000), p. 276, and some of the earlier specifically public sector literature summarized in Crewson (1997, p. 500). Houston (2000) is also relevant.

basis of the self-interest postulate. An encouraging development of very recent years, however, is the emergence of some economic literature addressing the role of intrinsic motivators in general and of “public service motivation” in particular.⁷⁵ One of the earliest contributors has been Bruno Frey (1994, 1997a, 1997b, Frey and Oberholzer-Gee, 1997), who has not only introduced the intrinsic/extrinsic motivational theory to economics, but coined the “crowding out” terminology in this context. Bénabou and Tirole (2003) develop on the foundations of psychological research an economic theory of the interplay of extrinsic incentives and intrinsic motivators, and which seeks to shed light on the circumstances under which “crowding out” might and might not occur. Specifically addressing the public sector context, Francois presents a model formalizing the intuition that “public sector reform based on implementing management and incentive practices from business can diminish employee effort based on public service motivation” (2000, p. 279). Besley and Ghatak (2003) develop a formal theory based on the commitment of public officials to their missions, in which the lesser reliance of those public sector agencies upon extrinsic motivators serves the important function of facilitating the “matching” of individuals with appropriate mission commitment to agencies in the recruitment process. Experimental economists have also begun to make an empirical contribution which reinforces the findings of earlier research in other disciplines on the potential for high-powered monetary incentives under some circumstances to produce worse outcomes.⁷⁶

However, it would be quite wrong to regard this literature as providing a case either against any increase in the role of extrinsic motivators, or against the greater use of performance measurement in the public sector management and budgeting.

In the first place, although public officials are not purely self-interested, they are most definitely not purely altruistic. They are, rather, driven by a mix of motivations. Accordingly, the optimal motivational system will be some mixture of extrinsic incentives and public service motivation. The problem is to achieve the right balance between reliance upon intrinsic motivators and (the right type of) extrinsic incentives. There is certainly no reason whatsoever to conclude that the low degree of reliance upon extrinsic performance incentives in traditional public administration has been optimal. Indeed, not only is the proposition that too little reliance has traditionally been placed upon these extrinsic incentives highly plausible, but there is also at least some empirical evidence to that effect (see Burgess and Ratto, 2003, p. 298).

There needs also to be some caution about the application of the crowding out hypothesis. There is no basis to assume that *any* expansion of the role of extrinsic incentives will be at the expense of intrinsic motivation. In the context of private sector performance pay systems, Deckop, Mangel and Cirka (1999), for example, produce empirical results which support the

⁷⁵ Some of this building upon a limited earlier literature on volunteerism and nonprofit entities.

⁷⁶ See the research summarized in Bénabou and Tirole (2003, p.490).

proposition that in organizations with high value-alignment, the right type of performance pay system can complement, rather than weaken, a crucial source of intrinsic motivation (“organizational citizenship behavior,” defined as “employee behavior that goes above and beyond the call of duty, that is discretionary and not expressly recognized by the employing organization’s formal reward system, and that contributes to organizational effectiveness”). Taylor and Pierce (1999) provide some evidence of a similar effect in a public sector context.

Importantly, it is quite possible that the presence of strong intrinsic motivations in the workforce may actually permit a somewhat greater role to be played by (the right types of) extrinsic performance incentives than would otherwise be the case in a context of substantial measurement problems, uncertainty, multi-tasking and multiple principles—that, as Osterloh and Frey (2000, p. 540) put it “intrinsic motivation also helps to overcome the so-called multi-task problem.” If agents were purely self-interested, they could be expected to exploit to the maximum any imperfections in performance measures. However, the presence of strong “public service motivation” and commitment to the mission may at least partly neutralize this tendency, because it means that agents care not only about their own interests, but also about the results which they achieve. The casemix funding example examined in Section IV provides, perhaps, some support for this proposition—prior predictions of serious perverse and dysfunctional effects from the highly imperfect output measures upon which the system was based appear not to have been realized, and this may well have been because of the commitment of health practitioners to quality of care and professional standards. It may that the casemix funding system may be a particularly positive example of this neutralization effect, given that—as noted in Section IV—there is some empirical evidence of “gaming” responses to the choice of performance measures used in other sectoral performance budgeting and management. This does not, however, invalidate the general point.

Two further points are also of relevance in this context. First, it does not follow from the presence of some undesirable responses to imperfect measures that there has not been a net improvement in performance as a result of the use of performance information in management and budgeting. Secondly, some of the supposedly perverse effects arising from the use of performance measures may not in fact be perverse at all, but might reflect legitimate policy choices such as a desire to weight efficiency considerations somewhat more vis-à-vis equity.

More generally, the behavioral literature reinforces a point which is, again, a very familiar one to management and public administration scholars—that building and molding public service motivation and mission-orientation are at the core of effective public management. In considering the motivational impact of managing-for-results and performance budgeting systems, therefore, it is clearly essential to consider explicitly their impact on public service motivation and mission orientation. Although no attempt can be made here to do justice to this issue, a couple of points may perhaps usefully be made. The first arises is that it is necessary not only to nurture and reinforce the sense of mission-orientation, but also to be able to change and re-mold it. The conception of their mission held by public officials may vary from that held either by their political masters or the broader community. This might be for any of a range of reasons, including ideological orientation (and the failure to adapt that

orientation to changing circumstances), professional values/norm, and myopia (an understandable failure to see the bigger picture, as in the example of the trade-off between equity and efficiency in the context of limited resources). Managing-for-results processes—starting with the process of making the desired outcome as explicit as possible, and linking outputs and activities to those outcomes—would appear to make sense in this context as a means of improving “goal alignment.” Target-setting—preferably as closely linked to the budget formulation as possible—may, have an important signaling role to play in this context. Thus, for example, Courty and Marschke (2003b), present a model in which the use of imperfect performance measures in the public sector plays the useful role, even in the absence of any linkage with extrinsic incentives, in communicating priorities between competing objectives to agents in a multi-tasking environment. Finally, insofar as measures and targets actually do successfully capture the success of the organization in achieving the outcomes which public officials care about, their use in management and budgeting may help to focus more attention on the bigger picture, partially offsetting the myopia which so easily affects people operating as small parts of larger production processes.

The fundamental *managerial freedom* theme of managing-for-results—and its expression in a performance budgeting context as advocacy of the relaxation of input controls—may also make much greater sense if one recognizes the importance of intrinsic as well as extrinsic motivation. A common simple rationale for the managerial freedom doctrine is that strong accountability for results (outcomes and outputs) will fill the vacuum left by the drastic reduction of controls over processes and inputs. While there is much in this—namely, that the major improvement of performance information is important to the success of the what has sometimes rather hyperbolically been called “liberation management”—it nevertheless fails to take fully into account the inevitable serious limitations to our ability to measure results in many parts of the public sector. It is a commonplace proposition of organizational economics that, to the extent that it is hard to measure the results delivered by self-interested agents, more intensive monitoring and control of agents’ effort (i.e., processes) is likely if the degree of shirking is to be limited. This seems to undercut the managerial freedom theme. This is, however, were the sense of mission and public service motivation may play a crucial role—the stronger are these altruistic motivators, the more likely is that de-control of processes and inputs will result in improved performance rather than increased shirking.

The above analysis is, it must be admitted, rather speculative. As Wright (2001) observes, the topic of public sector work motivation has received far too little attention by scholars. We need much better theory—but theory with stronger empirical foundations—to build what Crewson (1997, p. 500) calls “a better understanding of civil service motivation and behavior.” All over the world, experiments are at present underway in changing the nature of the motivators and incentives faced by public officials, in the interests of improved performance. This will certainly provide a wealth of data for future researchers, which will hopefully generate information which will ultimately help us improve the design of managing-for-results systems generally, and performance budgeting in particular.

U. S. STUDIES RELEVANT TO THE EFFICACY OF PERFORMANCE BUDGETING

Do Performance Measures Change Budget Allocations?

| <i>Date (when empirical work carried out)</i> | <i>Author(s)</i> | <i>Level of Government</i> | <i>Nature of Study</i> | <i>Findings</i> |
|---|----------------------------------|--------------------------------|--|---|
| 1999-2000 | Melkers, Willoughby et al. | State and Local | Survey of (i) "state budget offices" (i.e., both executive budget offices and legislative budget officials) (responses from 36 states); (ii) state agency officials with budgeting and/or performance measurement responsibilities (responses from 48 states), (iii) budget officers & department heads of city and county governments (responses from 253). | The percentages of respondents who asserted that performance measures were "very effective" or "effective" in changing budget appropriations were 7.5 percent (state budget offices), 24.2 percent (state agencies) and 16.9 percent (city/county) (Melkers, Willoughby, James, Fountain and Campbell, 2002, p. 21). |
| 1997 | Melkers and Willoughby | State | Survey of executive budget offices and legislative budget officials. 104 responses from 49 states. | In response to the proposition that "some changes in appropriations are directly attributable to outcomes from the implementation of performance-based budgeting in my state," 9.4 percent of respondents strongly agreed and 29.7 percent agreed. The figures were significantly lower for legislative branch respondents than for those from the executive branch (Melkers and Willoughby, 2001, p. 61). When asked to rate the effectiveness of performance-based budgeting in "changing appropriation levels" on a four point scale from 1 (not effective at all) to 4 (very effective), the average rating was 1.54 (Willoughby and Melkers, 2000, p. 114). Note: Willoughby and Melkers define performance-based budgeting as, approximately speaking, strategic planning plus performance measures (see discussion in Section II). |
| 1997 | Jordan & Hackbart | State | Survey of executive budget offices. 46 state responses. | 29 out of 45 respondent states agreed that "achievement of performance standards affects budget recommendations in the Governor's Executive Budget." 33 out of 46 respondent states claim that performance affects funding in the next fiscal year, and 11 of the 46 claim that performance affects funding even in the current fiscal year (Jordan & Hackbart, 1999, p. 75, 77). |

U. S. STUDIES RELEVANT TO THE EFFICACY OF PERFORMANCE BUDGETING
Do Performance Measures Change Budget Allocations?

| <i>Date (when empirical work carried out)</i> | <i>Author(s)</i> | <i>Level of Government</i> | <i>Nature of Study</i> | <i>Findings</i> |
|---|--------------------|--------------------------------|--|--|
| 1997 | Poister and Streib | Local | Survey of city government. 695 respondents, most of whom were city managers or assistant managers. | Note: Unfortunately, this survey did not give respondents an opportunity to indicate the degree of impact on budget allocations. 60 percent of respondents from jurisdictions which had “centralized, citywide performance measurement systems that incorporate most departments and programs” reported moderate or substantial changes in budget allocations as an impact of performance measures (Poister and Streib, 1999, pp. 332–33). Note: Responses may pertain not only to city-wide budgeting, but also to some extent to internal budgeting with organizational sub-units. |

Are Performance Measures Used in Resource Allocation Decisions?

| <i>Date (when empirical work carried out)</i> | <i>Author(s)</i> | <i>Government Budget (GW) or Internal Agency Budgeting (I)</i> | <i>Level of Government</i> | <i>Nature of Study</i> | <i>Findings</i> |
|---|----------------------------|--|--------------------------------|--|--|
| 2000 | GAO | I ¹ | National | Survey of sample of managers and supervisors in 28 agencies of the federal government. 2,510 useable responses. | 43 percent of respondents reported using performance measures to a “great” or “very great” extent in allocating resources. A similar percentage (45 percent) reported making use of performance measures for priority setting (GAO, 2001b, pp. 11, 29–32). |
| 1999–2000 | Melkers, Willoughby et al. | GW | State, Local | Survey of (i) “state budget offices” (i.e., both executive budget offices and legislative Budget officials) (responses | 26.3 percent of respondent state budget offices affirmed that output or outcome measures were either “very important” or “important” in decisions about budget appropriations to agencies. The corresponding |

¹While the survey apparently did not distinguish between internal and across-government resource allocation, the vast majority of respondents were from “line” agencies.

Are Performance Measures Used in Resource Allocation Decisions?

| Date (when empirical work carried out) | Author(s) | Government Budget (GW) or Internal Agency Budgeting (I) | Level of Government | Nature of Study | Findings |
|---|-------------------|--|------------------------|--|---|
| | | | | <p>from 36 states); (ii) state agency officials with budgeting and/or performance measurement responsibilities (responses from 48 states), (iii) budget officers and department heads of city and county governments (responses from 253).</p> | <p>figure for city/county was 38.7 percent (Melkers, Willoughby, James, Fountain and Campbell, 2002, p. 16).</p> <p>40 percent of state budget offices, 25.4 percent of state agency officials, and 33.9 percent of city/county respondents “performance measurements” not carrying enough weight in budgetary decisions’ was a “significant problem” for the effective use of performance measurement. (Melkers, Willoughby, James, Fountain and Campbell, 2002, p. 29).</p> <p>35.1 percent of respondent state budget offices believed that output or outcome measures were important or very important in agency “budget development.” For city/county, the corresponding figure was 46.0 percent.</p> <p>28.1 percent of respondent state budget offices believed that these measures were important or very important in agency “budget execution.” For city/county, the corresponding figure was 35.9 percent.</p> |
| 1996-97 | GAO | I | National | <p>Survey of sample of managers and supervisors in 24 agencies of the federal government. Useable responses from about 72 percent of the 1,300 persons surveyed.</p> | <p>21 percent of respondents reported that performance measures were used to a “great” or “very great” extent in developing agency budgets. The use of performance measures as the basis for funding decisions had a similar response rate (20 percent). (GAO, 1997b, p. 49).</p> |
| 1997 | Jordan & Hackbart | GW | State | <p>Survey of executive budget offices.</p> | <p>23 of 41 respondent states “agree” with the proposition that “performance indicators are an important tool for making budget allocation decisions in my state.” 3 of the 41 “strongly agree” (Jordan & Hackbart, 1999, p. 78).</p> |

| Are Performance Measures Used in Resource Allocation Decisions? | | | | | | |
|---|---|---|---------------------|---|--|--|
| Date (when empirical work carried out) | Author(s) | Government Budget (GW) or Internal Agency Budgeting (I) | Level of Government | Nature of Study | Findings | |
| 1997 | Poister and Streib | GW, I | Local | Survey of city government. 695 respondents, most of whom were city managers or assistant managers. | Almost two-thirds of respondents from jurisdictions which had “centralized, citywide performance measurement systems that incorporate most departments and programs” believed performance measures were important or very important for budgeting purposes (Poister and Streib, 1999, p. 330). | |
| 1996 | GASB [directed by Melkers and Willoughby] | I ² | State, Local, Other | Survey of state, local, educational (school, university) and some other public sector entities. 900 responses. | Of state government departments responding, 44.7 percent and 41.7 percent of those which claimed to have developed performance measures claimed to use output and outcome indicators respectively for resource allocation purposes. For municipalities, the figures were 20.1 percent and 17.7 percent, and for counties, 33.6 percent and 28.1 percent. The corresponding figures for the entire sample of 900 public sector entities were 27.7 percent and 25.2 percent. (GASB, 1997, pp. 1, 6–8). However, it should be noted that only a portion of the entities claiming to have developed performance measures claimed to have developed outcome or output measures. Recalculating the percentages claiming to have used outcome/output measures for resource allocation purposes as a percentage of those claiming to have developed outcome/output measures yields the following: state government departments—68.6/64.0 percent; municipalities—66.1/58.1 percent; counties—84.5/70.7 percent. | |
| 1996 | Wang | I, GW? | Local | Survey of Florida municipal officials (elected officials, city managers, police chiefs, and finance directors). | When asked to rate the role of performance measures in “making budget allocation decisions in setting up funding priority and funding levels” on a scale from 1 (“not favorable”) to 5 (“very favorable”), the average rating | |

²Only a very small percentage of the respondents would have been central budget offices.

| Are Performance Measures Used in Resource Allocation Decisions? | | | | | | |
|---|-----------|---|---------------------|---|--|--|
| Date (when empirical work carried out) | Author(s) | Government Budget (GW) or Internal Agency Budgeting (I) | Level of Government | Nature of Study | Findings | |
| 1970–1995, at 5-yearly intervals. | Lee | GW | State | 178 responses. Follow up qualitative interviews with selected respondents. Survey of the executive budget offices in all U.S. States. Responses from great majority. | accorded by respondents was 3.7 (3.7 from finance directors) (Wang, 1999, pp. 543–542). [Note: (i) it is not absolutely clear that the survey was being asking about the extent of use of performance measures for this purpose as opposed, say, to their view as to whether indicators should be used in such a way; (ii) the survey focused on police services, and it is not clear whether municipal-wide officials were expressing views about the use of performance measures in budgeting generally, or in determining the police budget specifically. The percentage of states in which the state budget offices claimed that “substantial use” was made of outcome/output information in the formulation of the executive budget rose from 15/15 percent in 1970 to 26/45 percent in 1990, and then fell to 18/30 percent in 1995 (Lee, 1997a, pp. 136–37). The percentage of states in which the state budget offices believed that “substantial use” was made of outcome/output information in legislative budget decision-making broadly fell from 25/19 percent in 1970 to 11/10 in 1995 (Lee, 1997a, pp. 136–37). The percentage of states in which the state budget offices who claimed that they made “substantial use” or “some degree” of program analysis in the formulation of the executive budget broadly rose until 1990, and then declined between 1990 and 1995 (a drop from 90 percent to 72 percent in the use of effectiveness analysis, and from 100 percent to 84 percent in productivity analysis (Lee, 1997b, pp. 21–22). [Note: it is not possible to disaggregate the “substantial use” respondents from the “some degree” of use respondents, although it is clear that the numbers of the former are significantly smaller than the latter.] Connelly and Tompkins, 1989—need to look at when received. | |

U. S. GASB CASE STUDIES
Government Accounting Standard Board Case Studies on the Use and Effects of Using Performance Measures for Budgeting, Management, and Reporting

| State/City/County [States] | Background | Use of PMs in Resource Allocation | Effects of PMs on Resource Allocation | Use of PMs for Performance Improvement | Effects of PMs on Performance Improvement |
|---|--|---|---|---|--|
| Maine (Tucker and Campbell, 2002) | Performance budgeting legislated 1997, for implementation by 2001. Primary intent of legislation is use of performance information in resource allocation. | “had not occurred at the time of this study.” | “the first performance budget had not been developed at the time of this study. Because of this, little effect on the resource-allocation process can be demonstrated.” | Unclear at this stage, because measurement requirements are new. | Unclear at this stage, because measurement requirements are new. |
| Wisconsin (Melkers and Mhatre, 2002) | Performance measurement used for “many years.” Efforts to use them in the budget process quite recent, starting with a pilot project in 1999. | “uncertain” | “interviewees ...saw little evidence of an impact on resource allocation.” | Increasingly being used. | Case study does not make clear. |
| Arizona (Tucker, 2000a) | Legislation requiring performance measurement etc (along GPRA lines) was passed in 1993. Attempt to integrate with budgeting in the 1999-2001 biennium budget cycle. | See next column. | “not one individual provided an example of a direct relationship between the use of performance measures and funding levels.” | Are being used. | Significant effects. |
| Illinois (Tucker, 2000b) | Performance measurement initiative relatively new. No indication in case study of effort to link with budget. | Little use. | “There has not been a measurable result in terms of dramatic allocation changes or trends realized” | “measures being developed, collected, and reported were not necessarily being used to further performance improvement, yet” | Little. |
| Iowa (Epstein and | “Budgeting for Results” system introduced in 1995, as | “Agencies have used budgeting for results | “In some cases, it has resulted in a reallocation | Many examples of use. | “Departments reported significant benefits and |

U. S. GASB CASE STUDIES
Government Accounting Standard Board Case Studies on the Use and Effects of Using Performance Measures for Budgeting, Management, and Reporting

| State/City/County | Background | Use of PMs in Resource Allocation | Effects of PMs on Resource Allocation | Use of PMs for Performance Improvement | Effects of PMs on Performance Improvement |
|---|---|--|---|---|--|
| Campbell, 2000a) | part of a broader managing for results system. | and performance measurement to develop and justify budgets and to target allocation of resources after they are budgeted. The governor uses the information in making budget decisions. Some legislative committees use the budgeting for results structure and performance data in their deliberations, and others do not." | of resources" | | results from using measurement for strategic planning, performance monitoring, and performance improvement." |
| Louisiana (Epstein and Campbell, 2000b) | In 1997, legislation requiring a "managing for results cycle," including performance budgeting, was passed. | "Legislative staff, executive planning staff, and department staff all remarked that the inclusion of performance information in budget proposals...had affected the budget process." | "performance information has mostly influenced decisions on proposed new programs, or funding increases and decreases to specific programs" | Examples of use. | Unclear. Single example of performance improvement effect. |

U. S. GASB CASE STUDIES
Government Accounting Standard Board Case Studies on the Use and Effects of Using Performance Measures for Budgeting, Management, and Reporting

| State/City/County | Background | Use of PMs in Resource Allocation | Effects of PMs on Resource Allocation | Use of PMs for Performance Improvement | Effects of PMs on Performance Improvement |
|--------------------------|---|--|--|--|--|
| Oregon (Fountain, 2000) | Many years work on performance indicator development, particularly benchmarks. Efforts from early 1990s to link to budget process. | “the use of performance measures in the budget process seems to have evolved into one of providing elected officials with information that will assist them in decision making and in understanding the role of various state programs in achieving the results deemed important by the legislature and the governor, not in directly helping establish funding levels.” | Significant, but not in mechanical way originally conceived. | “performance measurement has been used in a wide variety of ways that have contributed to ...performance improvement within the state” | Substantial. |
| Texas (Tucker, 2000) | Texas is “an excellent example of a state government that has effectively institutionalized the use of performance measures.” Legislation on performance measurement, budgeting etc in 1991. Considerable effort to link to budget process. | Used extensively. | Case study does not make clear. | Substantial. | Substantial. |

U. S. GASB CASE STUDIES
Government Accounting Standard Board Case Studies on the Use and Effects of Using Performance Measures for Budgeting, Management, and Reporting

| State/City/County [Cities and Counties] | Background | Use of PMs in Resource Allocation | Effects of PMs on Resource Allocation | Use of PMs for Performance Improvement | Effects of PMs on Performance Improvement |
|--|--|--|---|--|--|
| City of Austin (Epstein and Campbell, 2000c) | New “managing for results” system, including performance budgeting, was being implemented at time of study. However, significant previous measurement efforts, and some performance budgeting efforts. | “Performance information is used on several levels in Austin for resource allocation.” | “The effects of using performance measurement for resource allocation were more likely to come about because of departmental uses of measures than Council uses, either because departments shifted resources in their proposed budgets, or they shifted how their budgeted funds were used.” | Significant. | “The effects of using performance measurement cited by Austin manager included ...explicit cost savings and service performance improvements.” |
| Multnomah County, Oregon (Bernstein, 2000a) | Substantial efforts in performance measurement for many years. Performance budgeting since 1993–94. | Not clear. | “It is not clear, however, whether decisions ...are really driven by the use of performance measures.” | Substantial. (“There was a general sense that using performance measures for performance improvement and achieving targeted outcomes is becoming more of an expectation, and perhaps even ingrained behavior”) | “overall ... positive effect on the performance of County programs and the achievement of goals and outcomes.” |

U. S. GASB CASE STUDIES
Government Accounting Standard Board Case Studies on the Use and Effects of Using Performance Measures for Budgeting, Management, and Reporting

| State/City/County | Background | Use of PMs in Resource Allocation | Effects of PMs on Resource Allocation | Use of PMs for Performance Improvement | Effects of PMs on Performance Improvement |
|---|---|--|--|---|--|
| Portland, Oregon (Bernstein, 2000b) | Substantial performance measurement efforts over many years. Extent of performance budgeting effort not made clear. | Apparently only "marginally." | Apparently negligible. | Substantial. | Case study does not make clear. |
| Tucson (Bernstein, 2000c) | Substantial performance measurement efforts throughout 1990s, and performance budgeting since "early to mid 1990s." | Elected councilors don't use the measures in budget decisions. However, City department do use it in their internal budgeting. | Apparently little. | Case study does not make clear. | Case study does not make clear. |
| Winston-Salem (Bernstein, 2000d) | Performance measurement used for 25 years. | "The Budget Office looked at workload in almost every department to ... understand budget requests. Department heads interviewed indicated that they too used performance measures to identify where programs needed additional resources ..." | "performance measurement ... helped managers better allocate resources" | Substantial. | "performance measurement use had helped fine tune services offered, saved some money, allowed government to do more with less" |
| Prince William County, Virginia (Bernstein, 2002) | "Throughout the 1990s ... incrementally implemented processes to convert traditional budget and reporting to a managing-for-results program." | "according to the majority of those interviewed, use of performance measurement in the budget process is | "The decline in the overall cost of government, the shifting of resources to strategic goal areas ... are cited by the County as examples that the outcome | Substantial. | Substantial. |

U. S. GASB CASE STUDIES
Government Accounting Standard Board Case Studies on the Use and Effects of Using Performance Measures for Budgeting, Management, and Reporting

| State/City/County | Background | Use of PMs in Resource Allocation | Effects of PMs on Resource Allocation | Use of PMs for Performance Improvement | Effects of PMs on Performance Improvement |
|--|---|--|--|---|---|
| | | <p>pretty much as advertised. Performance measures are considered, both in terms of prior-year results and as targets for the current and future years.”</p> <p>Not entirely clear, but appears limited.</p> | <p>budget process is working...Between 1992 and 1998, percentage change in per capita cost of programs demonstrated a shift or resources to those areas that are strategic initiatives.”</p> <p>Apparently very limited.</p> | | |
| <p>Dekalb County, Georgia (Tucker and Willoughby, 2002)</p> <p>Sunnyvale, California (Epstein, Campbell and Tucker, 2002a)</p> | <p>New system of “performance management” being implemented.</p> <p>“consistent, systematic use of performance measures for over twenty years, for resource management and continual performance improvement”</p> | <p>Substantial.</p> | <p>“Sunnyvale’s uses of performance measurement appear to have met officials’ stated intentions and expectations for resource allocation and policy decisions.” A number of concrete examples cited.</p> | <p>Being increasingly used.</p> <p>Substantial.</p> | <p>Case study does not make clear.</p> <p>Claimed to have boosted productivity and service quality substantially.</p> |
| <p>San Jose, California (Epstein, Campbell and Tucker, 2002b)</p> | <p>New “results-driven performance management system” being introduced over recent years.</p> | <p>“Performance measurement has been used to allocate resources in the two performance-budgeting pilot departments.”</p> | <p>Limited at this early stage.</p> | <p>Some use.</p> | <p>Successful examples cited. Perception of overall impact not made clear.</p> |

References

- Allen, R., and P. Gertler, 1991, "Regulation and the Provision of Quality to Heterogeneous consumers: The Case of Prospective Pricing of Medical Services," *Journal of Regulatory Economics*, Vol. 3 (December), pp. 361–75.
- ANAO (Australian National Audit Office), 1997, *Program Evaluation in the Australian Public Service*, Audit report No. 3, 1997–98) (Canberra: Australian Government Publishing Service).
- Adnett, Nick, 2003, "Reforming Teachers' Pay: Incentive Payments, Collegiate Ethos and UK Policy," *Cambridge Journal of Economics*, Vol. 27, No.1, pp. 145–57.
- Andrews, Matthew and Herb Hill, 2003, "The Impact of Traditional Budgeting Systems on the Effectiveness of Performance-Based Budgeting: A Different Viewpoint on Recent Findings," *International Journal of Public Administration*, Vol. 26, No. 2, pp. 135–55.
- Auditor-General of Victoria, 1998, *Acute Health Services under Casemix: a Case of Mixed Priorities* (Melbourne: Victorian Government Printer).
- Averill, Richard F., Michael J. Kalison, James C. Vertrees, and Norbert I. Goldfield, 1996, "Achieving Short-Term Medicare Savings Through the Expansion of the Prospective Payment System," *Health Care Management Review*, Vol. 21, No. 4, pp.18–25.
- Ball, Ian, 1994, "Reinventing government: Lessons learned from the New Zealand Treasury," *The Government Accountants Journal*, Vol. 43, No. 3, pp. 19 passim.
- Barnow, Burt S., 2000, "Exploring the Relationship Between Performance Management and Program Impact: A Case Study of the Job Training Partnership Act," *Journal of Policy Analysis and Management*, Vol.19, No.1, pp.118–41.
- Behn, Robert D., 1994, "The Wrong Way to Motivate," *Governing*, Vol. 8 (December), p. 70.
- Bellamy, Sheila, and Ron Kluvers, 1995, "Program Budgeting in Australian Local Government: A Study of Implementation and Outcomes," *Financial Accountability And Management*, Vol. 11, No.1, pp. 39–56.
- Bénabou, Roland and Jean Tirole, 2003, "Intrinsic and Extrinsic Motivation," *The Review of Economic Studies*, Vol. 70, pp. 489–520.
- Bernstein, David J., 2000a, *GASB SEA Research Case Study: Multnomah County, Oregon: A Strategic Focus on Outcomes* (city/: Government Accounting Standards Board.

- , 200b, *GASB SEA Research Case Study: Portland, Oregon: Pioneering External Accountability*, Government Accounting Standards Board.
- , 2000c, *GASB SEA Research Case Study: Tucson, Arizona: An Evolving Performance Measurement Culture*, Government Accounting Standards Board.
- , 2000d, *GASB SEA Research Case Study: City of Winston-Salem, North Carolina: Focusing on Government Efficiency and Public Confidence*, Government Accounting Standards Board.
- , 2002, *GASB SEA Research Case Study: Prince William County, Virginia—Developing a Comprehensive Managing-for-Results Approach*, Government Accounting Standards Board.
- Berry, William D., 1990, “The Confusing Case of Budgetary Incrementalism: Too Many Meanings for a Single Concept,” *Journal of Politics*, Vol. 52, No.1, pp. 167–96.
- Berwick, Donald M., et. al., 2003, “Paying for Performance: Medicare should Lead,” *Health Affairs*, Vol. 22, No. 6, pp. 8–9.
- Besley, Timothy and Maitreesh Ghatak, 2003, “Incentives, Choice, and Accountability in the Provision of Public Services,” *Oxford Review of Economic Policy*, Vol. 19, No. 2, pp. 235–49.
- Blondal, Jon, 2003, “Budgeting in the United States,” *OECD Journal on Budgeting*, Vol. 3, No. 2, pp. 7–54.
- Boccuti, Christina and Marilyn Moon, 2003, “Comparing Medicare and Private Insurers: Growth Rates in Spending Over Three Decades,” *Health Affairs*, Vol. 22, No. 2, pp 230–37.
- Botner, Stanley, 1985, “The Use of Budgeting/Management Tools by State Governments,” *Public Administration Review*, September/October, pp. 616–20.
- Boyne, George, Rachel Ashworth and Martin Powell, 2000, “Testing the Limits of Incrementalism: an Empirical Analysis of Expenditure Decisions by English Local Authorities, 1981–1996,” *Public Administration*, Vol.78, No.1, pp. 51–73.
- Bray, Nancy, Carol Carter, Allen Dobson, Michael J Watt and Stephen Shortell, 1996, “An Examination of Winners and Losers under Medicare’s Prospective Payment System,” *Health Care Management Review*, Volume 19, No. 1, p. 44 passim.
- Breul, Jonathon, 2003, “The Government Performance and Results Act—10 Years After,” *Journal of Government Financial Management*, Spring, pp. 58–64.

- Brown, Stephanie and Judith Lumley, 1998, "Are Cuts to Health Expenditure in Victoria Compromising Quality?," *Australian and New Zealand Journal of Public Health*, 22(2), pp. 279–81.
- Brumby, Jim, Peter Edmonds and Kim Honeywell, 1996, "Effects of Public Sector Financial Reform (FMR) in New Zealand, unpublished paper presented to Australasian Evaluation Society Conference, August 1996.
- Burgess, Simon and Marisa Ratto, 2003, "The Role of Incentives in the Public Sector: Issues and Incentives," *Oxford Review of Economic Policy*, Vol. 19, No. 2, pp. 285–300.
- Campbell, Colin, 2004, "Juggling Inputs, Outputs, and Outcomes in the Search for Policy Competence: Recent Experience in Australia," *Governance*, Vol. 4. No. 2, pp.253–82.
- Campos, Edgardo J. and Sanjay Pradhan, 1999, "Budgetary Institutions and the Levels of Expenditure Outcomes in Australia and New Zealand," in James M. Poterba and Jürgen von Hagen (eds), *Fiscal Institutions and Fiscal Performance*, Chicago: The University of Chicago Press.
- Carlin, Tyrone M., 2003, "Accrual Output-Based Budgeting Systems in Australia—a Great Leap Backwards?," *Australian Accounting Review*, 13(2), pp. 41–47.
- Coffey, Rosanna M., 1999, "Casemix Information in the United States: Fifteen Years of Management and Clinical Experience," *Casemix Quarterly*, Vol. 1, No. 1.
- Congressional Budget Office, 1993, *Using Performance Measures in the Federal Budget Process*, Congress of the United States.
- Coulam, Robert F and Gary L Gaumer, 1991, "Medicare's Prospective Payment System: A Critical Appraisal," *Health Care Financing Review*, 1991 Annual Supplement, pp. 45–77.
- Courty, Pascal and Gerald Marschke, 2003a, "Performance Funding in Federal Agencies: a Case Study of a Federal Job Training Program," *Public Budgeting and Finance*, Vol. 23, No. 3, pp. 22–48.
- , 2003b, "Dynamics of Performance Measurement Systems," *Oxford Review of Economic Policy*, Vol.19, No. 2, pp. 268–84.
- Crewson, Philip E., 1997, "Public-Service Motivation: Building Empirical Evidence of Incidence and Effect," *Journal of Public Administration Research and Theory*, Vol. 7, No. 4, pp. 499–518.

- Deckop, John, Robert Mangel and Carol Cirka, 1999, "Getting More Than You Pay For: Organizational Citizenship Behavior and Pay-For-Performance Plans," *Academy of Management Review*, Vol. 42, No. 4, pp. 420–28.
- Deeble, John, 2000, *Resource Allocation in Public Health: an Economic Approach*, National Public Health Partnership: Melbourne, 2nd Edition.
- Dewatripont, Mathias, Ian Jewitt and Jean Tirole, 1999, "The Economics of Career Concerns, Part II: Application to Missions and Accountability of Government Agencies," *The Review of Economic Studies*, Vol. 66, pp.199–217.
- Dezhbakhsh, Hasem, Soumaya M. Tohamy and Peter H. Aranson, 2003, "A New Approach for Testing Budgetary Incrementalism," *The Journal of Politics*, Vol. 65, No. 2, pp. 532–558.
- Diamond, Jack, 2003a, *Performance Budgeting: Managing the Reform Process*, IMF Working Paper WP/03/33.
- , 2003b, *From Program Budgeting to Performance Budgeting: The Challenge for Emerging Market Economies*, IMF Working Paper WP/03/169.
- DIMA/DOFA (Department of Immigration and Multicultural Affairs and Department of Finance and Administration), 2001, *Purchasing Agreement, 1 July 2001- 30 June 2004*, mimeo.
- Dismuke, C. E. and P. Guimaraes, 2002, "Has the Caveat of Case-Mix Influenced the Quality of Inpatient Hospital Care in Portugal?," *Applied Economics*, Vol. 24, pp. 1301–1307.
- Dismuke, Clara Elizabeth and Vania Sena, 1999, "Has DRG Payment Influenced the Productive efficiency and Productivity of Diagnostic Technologies in Portuguese public hospitals? An Empirical Analysis Using Parametric and Non-Parametric Methods," *Health Case Management Science*, Vol. 2, pp. 107–116.
- Dixit, Avinash, 2002, "Incentives and Organizations in the Public Sector: an Interpretive Review," *Journal of Human Resources*, Vol. 37, No. 4, pp. 696–727.
- Ellis, Randall P., 1998, "Creaming, Skimping and Dumping: Provider Competition on the Intensive and Extensive Margins," *Journal of Health Economics*, Vol. 17, No. 5, pp. 537–56.
- Ellis, Randall P. and T G McGuire, 1986, "Provider Behavior Under Prospective Reimbursement: Cost Sharing and Supply," *Journal of Health Economics*, Vol. 5, .pp. 129–51.

- Enthoven, Alain C., 2000, "In Pursuit of an Improving National Health Service," *Health Affairs*, Vol. 19, No. 3, pp. 102–20.
- Epstein, Paul D. and Wilson Campbell, 2000a, *GASB SEA Research Case Study: Iowa*, Government Accounting Standards Board.
- , 2000b, *GASB SEA Research Case Study: State of Louisiana*, Government Accounting Standards Board.
- , 2000c, *GASB SEA Research Case Study: City of Austin*, Government Accounting Standards Board.
- Epstein, Paul D., Wilson Campbell, and Laura Tucker, 2002a, *GASB SEA Research Case Study: City of Sunnyvale, California*, Government Accounting Standards Board.
- , 2002b, *GASB SEA Research Case Study: City of San Jose*, Government Accounting Standards Board.
- Executive Office of the President/Office of Management and Budget, 2002, *The President's Management Agenda, Fiscal Year 2002*, Washington: OMB.
- Fountain, Jay, 2000, *GASB SEA Research Case Study: State of Oregon: a Performance System Based on Benchmarks*, Government Accounting Standards.
- Francois, Patrick, 2000, "Public Service Motivation" as an Argument for Government Provision," *Journal of Public Economics*, Vol. 78, pp. 275–299.
- Frey, Bruno S., 1994, "How Intrinsic Motivation is Crowded In and Out," *Rationality and Society*, Vol. 6, No. 3, pp. 334–52.
- , 1997a, *Not Just for the Money: an Economic Theory of Personal Motivation*, Cheltenham: Edward Elgar.
- , 1997b, "On the Relationship between Intrinsic and Extrinsic Work Motivation," *International Journal of Industrial Organization*, Vol.15, No. 4, pp. 427 passim.
- Frey, Bruno S. and Felix Oberholzer-Gee, 1997, "The Cost of Price incentives: An Empirical Analysis of Motivation Crowding-Out," *The American Economic Review*; Vol. 87, No. 4, pp. 746–55.
- GASB (Government Accounting Standards Board), 2000, *Performance Measurement for Government: State and Local Government Case Studies on Use and the Effects of Using Performance Measures for Budgeting, Management, and Reporting*, Available at: http://www.seagov.org/sea_gasb_project/case_studies.shtml

- General Accounting Office, 1993a, *Performance Budgeting: State Experiences and Implications for the Federal Government*, GAO/AFMD-93-41.
- , 1994a, *Managing For Results: State Experiences Provide Insights for Federal Management Reforms*, GAO/GGD-95-22.
- , 1995a, *Managing for Results: Experiences Abroad Suggest Insights for Federal Management Reforms*, GAO/GGD-95-120.
- , 1997a, *Performance Budgeting: Past Initiative Offer Insights for GPRA Implementation*, GAO/AOMD-97-46.
- , 1997b, *The Government Performance and Results Act: 1997 Governmentwide Implementation will be Uneven*, GAO/GGD-97-109.
- , 1999a, *Performance Budgeting: Initial Experiences Under the Results Act in Linking Plans with Budgets*, GAO/AIMD/GGD-99-67.
- , 2001a, *Results-Oriented Budget Practices in Federal Agencies*, GAO-1084SP.
- , 2001b, *Managing for Results: Federal Managers' Views on Key Management Issues Vary Widely Across Agencies*, GAO-01-592.
- , 2004, *Observations on the Use of OMB's Program Assessment Rating Tool for the Fiscal Year 2004 Budget*, GAO-04-174.
- Gibbons, Robert, 1998, "Incentives in Organizations," *The Journal of Economic Perspectives*, 12(4), pp. 115-32.
- Goodman, Neville, 2002, "Please Give us Objectives We Can Aim At," *Journal of the Royal Society of Medicine*, Vol. 95, November, p. 567.
- Government Performance Project, 2003, *Paths to Performance in State & Local Government: a Final Assessment from the Maxwell School of Citizenship and Public Affairs*. Accessed 13 January 2004 at <http://www.maxwell.syr.edu/gpp/about/facts.asp>.
- Governmental Accounting Standards Board & National Academy Of Public Administration, 1997, *Report On Survey Of State And Local Government Use And Reporting Of Performance Measures: First Questionnaire Results*.
- Grizzle, Gloria A, 1986, "Does Budget Format Really Govern the Actions of Budgetmakers?," *Public Budgeting and Finance*, Vol. 6, No. 1, pp. 60-70.

- Grout, Paul A. and Margaret Stevens, 2003, "The Assessment: Financing and Managing Public Services," *Oxford Review of Economic Policy*, Vol.19, No. 2, pp. 215–34.
- Hatry, Harry P., 1999, *Performance Measurement: Getting Results*, Washington DC: The Urban Institute Press.
- Heckman, James, Carolyn Heinrich, and Jeffrey Smith, 1997, "Assessing the Performance of Performance Standards in Public Bureaucracies," *American Economic Review*, Vol. 87, No. 2, pp. 389–95.
- Heinrich, Carolyn J, 1999, "Do government bureaucrats make effective use of performance management information?," *Journal of Public Administration Research and Theory*, Vol. 9, No. 3, pp. 363–93.
- Hendon, Claude, 1999, "Performance Budgeting in Florida—Half Way There," *Journal of Public Budgeting, Accounting and Financial Management*, Vol. 11, No. 4, pp. 670–79.
- Houston, David, 2000, "Public-Service Motivation: a Multivariate Test," *Journal of Public Administration Research and Theory*, Vol.10, No. 4, pp. 713–27.
- James, Oliver (forthcoming), "The UK Core Executive's Use of Public Service Agreements as a Tool of Governance," *Public Administration* (forthcoming).
- Jones, David Seth, 1998, "Recent Budgetary Reforms in Singapore," *Journal of Public Budgeting, Accounting & Financial Management*, Vol. 10, No. 2, pp. 279–310.
- Jones, L. R. and Donald F. Kettl, 2003, "Assessing Public Management Reform in an International Context," *International Public Management Review*, Vol.4, No.1, pp. 1–19.
- Jordan, Meagan M & Merl M. Hackbart, 1999, "Performance Budgeting and Performance Funding in the States," *Public Budgeting and Finance*, Vol.19, No.1, pp. 68–88.
- Joyce, Philip, 1999, "Performance-Based Budgeting" in Roy T Meyers (ed.), *Handbook of Government Budgeting*, Jossey-Bass Publishers, San Francisco.
- , 2003, *Linking Performance and Budgeting: Opportunities in the Federal Budget Process*, Arlington, Virginia: IBM Center for The Business of Government.
- Kluvers, Ron, 2001, "An Analysis of Introducing Program Budgeting in Local Government," *Public Budgeting & Finance*, Summer, pp. 29–45.

- Kravchuk, Robert S and Ronald W Schack, 1996, "Designing effective performance-measurement systems under the Government Performance and Results Act of 1993," *Public Administration Review*, Vol. 56, No. 4, pp 348–358.
- Kreps, David M., 1997, "Intrinsic Motivation and Extrinsic Incentives," *American Economic Review*, Vol. 87, No. 2; pp. 359–64.
- Lancer Julnes de, Patria and Marc Holzer, 2001, "Promoting the utilization of performance measures in public organizations: An Empirical Study of Factors Affecting Adoption and Implementation," *Public Administration Review*; 61(6), pp 693–701
- Lauth, Thomas P., 1985, "Performance Evaluation in the Georgia Budget Process," *Public Budgeting and Finance*, 5(1), pp. 67–82.
- Layzell, Daniel T., 1998, "Linking Performance to Funding Outcomes for Public Institutions of Higher Education: the U.S. experience," *European Journal of Education*, 33(1), pp. 103–111.
- Le Grand, Julian, 1997, "Knights, Knaves or Pawns? Human Behavior and Social Policy," *Journal of Social Policy*, 26(2), pp. 149–169.
- Lee, Robert D Jr., 1997a, "A quarter century of state budgeting practices," *Public Administration Review*, 57(2), pp. 133–40.
- Lee, Robert D Jr., 1997b, "The Use of Program Analysis in State Budgeting: Changes Between 1990 and 1995," *Public Budgeting and Finance*, 17(2), pp. 18–36.
- Lee, Simon and Richard Woodward, 2002, "Implementing the Third Way: The Delivery of Public Services Under the Blair Government," *Public Money and Management*, October-December, pp. 49–56.
- Lu, Haoran, 1998, "Performance budgeting resuscitated: Why is it still inviable?," *Journal of Public Budgeting, Accounting and Financial Management*, 10(2), pp 151–72.
- Manton, Kenneth G, Max A. Woodbury, James C. Vertrees and Eric Stallard, 1993, "Use of Medicare Services Before and After Introduction of the Prospective Payment System," *Health Services Research*, 28(3), pp. 269–93.
- Martin, Graeme, 1994, "Performance-related pay in nursing: Theory, practice and prospect," *Health Manpower Management*, 20(5), pp. 10 passim.
- McNab, Robert M. and Francois Melese, 2003, "Implementing the GPRA: Examining the Prospects for Performance Budgeting in the Federal Government," *Public Budgeting and Finance*, 23(2), pp. 73–95.

- Melkers, Julia and Pratrik Mhatre, 2002, *Case Study: Wisconsin. Use and Effects of Using Performance Measures for Budgeting, Management and Reporting*, Government Accounting Standards Board.
- Melkers, Julia E and Katherine G Willoughby, 1998, "The State of the States: Performance-Based Budgeting requirements in 47 out of 50," *Public Administration Review*, 58(1), pp. 66–73.
- Melkers, Julia E and Katherine G Willoughby, 2001, "Budgeters" views of state performance-budgeting systems: Distinctions across Branches," *Public Administration Review*, 61(1), pp. 54–64.
- Melkers, Julia E., Katherine G. Willoughby, Brian James, Jay Fountain, Wilson Campbell, 2002, *Performance Measurement at the State and Local Levels: A Summary of Survey Results*, GASB,
- Mikesell, John L, 1995, *Fiscal Administration: Analysis and Applications for the Public Sector*, 4th edition, Wadsworth; Belmont, CA.
- Moynihan, Don, 2003, "Managing for Results," in Government Performance Project, *Paths to Performance in State & Local Government: a Final Assessment from the Maxwell School of Citizenship and Public Affairs*.
- National Audit Office, 2001, *Measuring the Performance of Government Departments*, London: The Stationery Office.
- Nove, Alec, 1984, *The Soviet Economic System*, London: George Allen and Unwin.
- OECD, 2003a, *Public Sector Modernization: Changing Organizations*, GOV/PUMA(2003)19.
- , 2003b, *Public Sector Modernization: Governing for Performance*, GOV/PUMA(2003)20.
- Office of Management and Budget, 2003, *Assessing Program Performance for the FY 2004 Budget*. Available at http://www.whitehouse.gov/omb/budintegration/part_assessing2004.html. Accessed May 14, 2003.
- Office of Program Policy Analysis and Government Accountability, 1997, *Performance-Based Program Budgeting in Context: History and Comparison*, Report 96–77A. Florida Legislature.
- Osterloh, Margit and Bruno S Frey, 2000, "Motivation, Knowledge Transfer, and Organizational Forms," *Organization Science*, 11(5), pp. 538–550.

- Palmer, George and Beth Reid, 2001, "Evaluation of the Performance of Diagnosis-Related Groups and Similar Casemix Systems: Methodological Issues," *Health Services Management Research*, 14, pp. 71–81.
- Pettijohn, Carole D and Gloria A Grizzle, 1997, "Structural budget reform: Does it affect budget deliberations?," *Journal of Public Budgeting, Accounting & Financial Management*, 9(1), pp. 26–45.
- Pitsvada, Bernard and Felix LoStracco, 2002, "Performance budgeting—the next budgetary answer. But what is the question?," *Journal of Public Budgeting, Accounting & Financial Management*, Vol. 14, No. 1, pp. 53–73.
- Poister, Theodore H and Gregory Streib, 1999, "Performance measurement in municipal government: Assessing the state of the Practice," *Public Administration Review*, Vol. 59, No. 4, pp. 325–335.
- Poocharoen, Ora-Orn and Patricia Ingraham, 2003, "Integration of Management Systems", in Government Performance Project," *Paths to Performance in State & Local Government: a Final Assessment from the Maxwell School of Citizenship and Public Affairs*.
- Premchand, A, 1999, "Budgetary Management in the United States and in Australia, New Zealand and the United Kingdom," in Roy T Meyers (ed.), *Handbook of Government Budgeting*, Jossey-Bass Publishers, San Francisco.
- Prendergast, Canice, 1999, "The provision of incentives in firms," *Journal of Economic Literature*, Vol. 37, pp.7–63.
- Propper, Carol and Deborah Wilson, 2003, "The Use and Usefulness of Performance Measures in the Public Sector," *Oxford Review of Economic Policy*, Vol.19, No. 2, pp. 250–67.
- Pyhrr, Peter A., 1973, *Zero-Base Budgeting: a Practical Management Tool for Evaluating Expenses*, New York: John Wiley & Sons.
- Radin, Beryl, 1998, "The Government Performance and Results Act (GPRA): Hydra-headed Monster or Flexible Management Tool," *Public Administration Review*, Vol. 58, No. 4, pp. 307–16.

- Reddick, Christopher G., 2002a, "Testing Rival Decision-Making Theories on Budget Outputs: Theories and Comparative Evidence," *Public Budgeting and Finance*, Vol. 22, No.3, pp. 1–25.
- , 2002b, "Canadian Provincial Budget Outcomes: a Long-Run and Short-Run Perspective," *Financial Accountability and Management*, 18(4), pp. 355–82.
- , 2003a, "Testing Rival Theories of Budgetary Decision-Making in the U. S. States," *Financial Accountability & Management*, Vol. 19, No. 4, pp. 315–339.
- , 2003b, "Budgetary Decision Making in the Twentieth Century: Theories and Evidence," *Journal of Public Budgeting, Accounting & Financial Management*, Vol. 15, No. 2, pp. 251–74.
- , 2003c, "Long-run and short-run Budgeting: Theories and Empirical Evidence for the Canadian provinces," *International Journal of Public Administration*, Vol. 26, No. 4, pp. 427–53.
- Robinson, Marc, 2000, "Contract Budgeting," *Public Administration*, Vol. 78, No. 1, pp. 75–91.
- , 2002a, "Financial Control in Australian Government Budgeting," *Public Budgeting and Finance*, Vol. 22., No.1, pp. 80–94.
- , 2002b, "Output-Purchase Funding and Budgeting Systems in the Public Sector," *Public Budgeting and Finance*, Vol. 22, No. 4, pp. 17–33.
- , 2005, "Performance Budgeting: Lessons from a Comparative Overview," in A. Bourmistrov and F. Mellemvik, *International Trends and Experiences in Government Accounting*, Oslo: Cappelen Akademisk Forlag.
- Rosenberg, Marjorie A. and Mark J. Browne, 2001, "The Impact of the Inpatient Prospective Payment System and Diagnosis-Related Groups: A Survey of the Literature," *North American Actuarial Journal*, Vol. 5, No. 4, pp. 84–94.
- Rubin, Irene S, 1997, *The Politics of Public Budgeting: Getting and Spending, Borrowing and Balancing*, 3rd edition, Chatham House Publishers, Chatham, NJ.
- Salber, Patricia R. and Bruce E. Bradley, 2002, "Adding Quality to the Health Care Purchasing Equation," *Health Affairs*.
- Sandmo, Agnar, 2002, "Public Provision and Private Incentives: Public versus Private Sectors in Public Finance," Opening lecture delivered to the International Institute of Public Finance Congress, August.

- Schick, Allen, 2003, "The Performing State: Reflection on an Idea Whose Time Has Come but Whose Implementation Has Not," *OECD Journal on Budgeting*, Vol. 3, No. 2, pp. 71–103.
- Shleifer, Andrei, 1998, "State versus Private Ownership," *The Journal of Economic Perspectives*, Vol. 12, No. 4, pp. 133–50.
- Smith, Peter C., 1995, "On the Unintended Consequences of Publishing Performance Data in the Public Sector," *International Journal of Public Administration*, Vol. 18, Nos. 2/3, pp. 277–310.
- Smith, James Fielding, 1999, "The Benefits and Threats of PBB: An Assessment of Modern Reform," *Public Budgeting and Finance*, Vol. 19, No. 3, pp. 3–15.
- Steering Committee on Government Service Provision, 1997, *Reforms in Government Services 1997*, Productivity Commission, Melbourne.
- Tankersley, W and Gloria Grizzle, 1994, "Control Options for the Public Manager: An Analytic Model for Designing Appropriate Control Strategies," *Public Productivity and Management Review*, Vol. 18, No. 1, pp. 1–17.
- Tanzi, Vito and Ludger Schuknecht, 2000, *Public Spending in the 20th Century*, Cambridge: Cambridge University Press.
- Taylor, Paul J. and Jon L. Pierce, 1999, "Effects of Introducing a Performance Management System On Employee's Subsequent Attitudes and Effort," *Public Personnel Management*, Vol. 28, No. 3, pp. 423–52.
- Thorn, Kristian and Mads Lyndrup, 2002, *The Quality of Public Expenditure: Challenges and Solutions of Result-Focused Management in Denmark*, OECD: Paris.
- Tieman, Jeff, 2003, "It was 20 Years Ago Today," *Modern Healthcare*, Vol. 33, No. 39, pp. 6 passim.
- Tirole, Jean, 1994, "The Internal Organization of Government," *Oxford Economic Papers*, Vol. 46, pp. 1–29.
- Tucker, Harvey, 1982, "Incremental Budgeting: Myth or Model?," *The Western Political Quarterly*, Vol. 35, No. 30, pp. 327–338.
- Tucker, Laura, 2000a, *GASB SEA Research Case Study: State Of Arizona—Focus On Performance*, Government Accounting Standards Board.
- , 2000b, *GASB SEA Research Case Study: State Of Illinois—Emphasis on Accountability and Managing for Results*, Government Accounting Standards Board.

- , 2000c, *GASB SEA Research Case Study: State Of Texas*, Government Accounting Standards Board.
- Tucker, Laura and Katherine Willoughby, 2002, *GASB SEA Research Case Study: DeKalb County, Georgia*, Government Accounting Standards Board.
- Tucker, Laura and Wilson Campbell, 2002, *Case Study: Maine. Use and Effects of Using Performance Measures for Budgeting, Management and Reporting*, Government Accounting Standards Board.
- Wildavsky, Aaron B., 1984, *The Politics of the Budgetary Process*, Boston: Little Brown.
- Wiley, Miriam M., 1992, "Hospital Financing Reform and Case-Mix Measurement: an International Review," *Health Care Financing Review*, Vol. 13, No. 4, pp. 119 passim.
- Williamson, Oliver, 1985, *The Economic Institutions of Capitalism: Firms, Markets, Relational Contracting*, New York: The Free Press.
- Willoughby, Katherine G and Julia E Melkers, 2000, "Implementing PBB: Conflicting Views of Success," *Public Budgeting and Finance*, Vol. 20, No. 1, pp. 105–121.
- Wilson, James Q., 2000, *Bureaucracy: What Government Agencies Do and Why They Do It*, New Edition, New York: Basic Books.
- Wright, Bradley, 2001, "Public Sector Work Motivation: a Review of the Current Literature and a Revised Conceptual Model," *Journal of Public Administration Research and Theory*, Vol. 11, No. 4, pp. 559–86.
- Worzala, Chantal, Julian Pettengill and Jack Ashby, 2003, "Challenges and Opportunities for Medicare's Original Prospective Payment System," *Health Affairs*, Vol. 22, No. 6, pp. 175–182.