Justification by Works or by Faith? : Evaluating the New Public Management

Christopher Pollitt

_Evaluation_ 1995 1: 133
DOI: 10.1177/135638909500100202

The online version of this article can be found at:
http://evi.sagepub.com/content/1/2/133

Published by:
SAGE
http://www.sagepublications.com

On behalf of:
The Tavistock Institute

Additional services and information for _Evaluation_ can be found at:

Email Alerts: http://evi.sagepub.com/cgi/alerts
Subscriptions: http://evi.sagepub.com/subscriptions
Reprints: http://www.sagepub.com/journalsReprints.nav
Permissions: http://www.sagepub.com/journalsPermissions.nav
Citations: http://evi.sagepub.com/content/1/2/133.refs.html

>> Version of Record - Jan 1, 1995

What is This?
Justification by Works or by Faith?

Evaluating the New Public Management

CHRISTOPHER POLLITT
Brunel University

The article first defines a set of policy measures that constitute the ‘New Public Management’ (NPM). Evidence is cited to support the contention that the NPM has been widely adopted, with local variations, in many Western states. The impacts of such large-scale reform are considerable. The article therefore explores the extent to which the NPM has been subject to serious and systematic evaluation. It is argued that there have been relatively few broad-scope evaluations and that the methodologies adopted for these have tended to leave some fundamental questions about the effectiveness of the NPM unanswered. Despite this, political enthusiasm for reforms of this type apparently continues to run at a high level. The article concludes with some modest suggestions for ways in which the evaluation of NPM-style reforms could be improved.

Introduction

‘There is a common core to public sector modernization in the OECD countries’ (Naschold, 1995: 1). There is also extensive (though not complete) agreement as to the contents of the core. Tolerably similar ingredients are listed by a variety of commentators (e.g. Barzelay, 1992; Dunleavy and Hood, 1994; Gore 1993; Naschold, 1995; OECD, 1993a; Pollitt, 1993; St-Martin, 1994). ’By the mid-1980s, it became clear that there was a remarkable degree of consensus among the political leadership of various countries about what was wrong about the civil service’ (Peters and Savoie, 1994: 419). The elements shown in Table 1 are now proclaimed and practised by a wide range of Western European and North American countries. They have come to be known collectively as the ‘New Public Management’ or NPM.

These eight elements comprise a kind of ‘shopping basket’ for those who wish to modernize the public sectors of Western industrial societies. Not every element is present in every case. Characteristic mixtures vary somewhat from country to country (e.g. more emphasis on performance management in Scandinavian countries, a stronger accent on MTMs, the contractualization of the public service and systematic approaches to improving service quality in New Zealand, the UK and the US; internal decentralization in France). This is only to be expected—after all, different countries
have experienced different historical trajectories and seek to reform themselves within very different constitutional frameworks (Jones, 1994; St-Martin, 1994: 33–5). Some countries—most notably Germany and Japan—seem not to have bought into NPM at all.

Nevertheless, after allowing for these real and important differences, most modernization efforts have involved the simultaneous deployment of a number of NPM elements. There are even some signs of convergence since the early 1990s. At the very least it is obvious that the elements listed in Table 1 interact one with another, often quite strongly. Thus, for example, the separation of purchasing and provision is a prerequisite for the introduction of MTMs, and also a way of disaggregating monolithic bureaucratic organizations. Similarly the setting of performance targets is a useful precursor to moving the terms of employment towards a term contract, PRP basis. In short there appears to be every justification for regarding these eight elements as an interacting set or system (NPM) rather than evaluating each one as though it were a single and distinct project or program. As the Secretary of the Australian Department of Finance said at the end of 1993:

...the various strands of reform are no accident of history. They are mutually supportive and their integrated nature is crucial to the overall success of the reforms. (Sedgewick, 1993: 14)

The working lives of millions of public officials are being substantially altered by these innovations. The management of vast budgets has been significantly reshaped. The basic services received by hundreds of millions of citizens have been and will be affected—for the better, if the proponents of NPM are to be believed. In the US the National Performance Review report (Gore, 1993) promised to ‘downsize’ the federal workforce by 232,000 (subsequently raised by Congress to 272,900). In the UK the Citizen’s Charter program launched in 1991 had, by 1993, been heard of by 71 percent of the population and had published some 40 charters setting out standards for a wide
variety of public services and utilities (ICM Research, 1993; Pollitt, 1994). Also in the UK the process of disaggregating central departments into core functions plus a network of executive agencies had, between 1989 and 1995, produced a situation in which, rather suddenly, 63 percent of all civil servants were working in agencies (Chancellor of the Duchy of Lancaster, 1994). In New Zealand even the most senior civil servants—the departmental chief executives—now relate to their ministers through quasi-contractual ‘performance agreements’. Standards have been prescribed not only for operational matters but also for the quality of policy advice (State Services Commission, 1992). In Australia the percentage of the paid workforce represented by the public service fell from 5.7 percent in 1975 to 2.6 percent in 1993.

Given the scale and significance of these and other impacts it might be supposed that the NPM would have attracted intensive evaluation, particularly so since some of the techniques which make up the NPM’s armamentarium lay great stress on evaluation as a means for ‘closing the feedback loop’. Yet inspection of key documents in a number of countries seems to show that, after more than a decade of NPM-style change, across-the-board evaluations have been far from numerous. Narrower evaluations—of change in one institution or dimension—have been more common, but are not the principal focus of this article (and in any case usually appear to suffer from the same theoretical and methodological limitations as their more ambitious counterparts). There is a paradox here: while NPM doctrine insists that public services must invest much more heavily in the currency of measurable outputs (e.g. HM Treasury, 1988), some fundamental aspects of NPM reforms themselves appear to have remained almost immune from such requirements.

Aims

Given the apparent shortage of broad yet rigorous evaluations of NPM reforms this article addresses three main questions:

1. How far do those evaluations which have been undertaken confirm that the gains claimed by proponents of the NPM have actually taken place?
2. By what methods have their findings been arrived at?
3. What kinds of evaluation could now be undertaken to improve our understanding of the advantages and disadvantages of the NPM?

In pursuing these questions the article concentrates on the management side of the policy/management interface—on changes in the actual running of public services. In several countries management reforms have also led to a reconsideration of the policy-making/policy-advisory functions themselves (e.g. St-Martin, 1994; Trosa, 1994), but proper discussion of these important ‘knock-on’ effects must perforce await a later article.

Scope and Purpose

It is a commonplace of the relevant literature that evaluation may have a variety of purposes and may be conducted across a wide range of levels, varying from individual
self-evaluation to large-scale international comparisons of entire economies or social systems. As the NPM is an international phenomenon—and has probably been most vigorously promoted by national-level governments—this article is mainly concerned with broad-scope evaluations. Such evaluations include those which focus on entire sectors (e.g., education, health care) or on large, multifaceted national organizations (e.g., the Australian public service) or, beyond even these large national studies, international comparisons of the kind carried out by the OECD or the World Bank. Inevitably, therefore, many finer points of detail are bound to be passed over and some local cases which exhibit interesting variations from broader trends will languish unacknowledged.

As far as the purposes of the evaluations reviewed are concerned, a fairly catholic/inclusive stance is maintained. There are too few large-scale evaluations available for it to be sensible to ignore or exclude particular categories. Our tour d’horizon will therefore include, first, relatively ‘pure’ academic analyses (e.g., Hyndman and Andersen, 1995; Dunleavy and Hood, 1994; Robinson and Le Grand, 1994); second, studies undertaken by academics as consultants for state agencies (e.g., Naschold, 1995; St-Martin, 1994); third, evaluations by intergovernmental bodies (e.g., OECD, 1993b, 1994); fourth, in-house analyses carried out by individual governments (e.g., Chancellor of the Duchy of Lancaster, 1994; Employment Service, 1993; Management Advisory Board, 1994; Task Force on Management Improvement, 1992; Trosa, 1994); fifth, scrutinies by state audit institutions (Auditor General of Canada, 1993; General Accounting Office, 1993) and, sixth, the more connoisseurial/discursive assessments that are sometimes offered by senior public officials and politicians (Sedgewick, 1993; Waldegrave, 1994).

The dominant purposes vary somewhat across this portfolio. Some of the more academic work is principally aimed at theorizing—or simply describing—new managerial arrangements. By contrast many of the government publications appear to have been produced mainly to meet the basic requirements of public accountability, or to ‘sell’ the reforms as the achievements of a particular administration. An honourable—indeed unique—exception to this often narrow genre is the much more ambitious $A1.1 million evaluation of a decade of Australian reform (Task Force on Management Improvement, 1992). Other evaluations have a more formative character, that is they are intended to assist those implementing the reforms to carry on their work more swiftly or effectively. Unsurprisingly, the particular methods employed also vary, and these will be an important focus of the later part of the article.

However, before tackling even the first question it is necessary to identify the kinds of claims which are made by the proponents of NPM. The degree to which intended effects are achieved is almost always a prime interest of evaluators, even if they go on to consider, in addition, unintended effects and other, processual features of a project or program. Intended effects (impacts, outcomes) are of interest in themselves, but also have the useful property of indicating (implicitly or explicitly) the criteria by which policy-makers apparently wish to be judged. Thus, if it is claimed that a particular management technique will enable an agency to issue more licences without any increase in staff or expenditure we may presume that efficiency is an important criterion for those promoting the new arrangements. Put in this simple way, the
relationship between, on the one hand, intended outputs and outcomes and, on the other, evaluative criteria may sound too obvious to be worth stating. However, in practice it is not uncommon to find unstated value premises or criteria lurking undeclared beneath seemingly ‘commonsense’ policy objectives, and there may be gains in clarity to be had from spelling out such linkages.

Evaluating the NPM

The NPM: What its Proponents Claim

The central claim of the NPM is embodied in the very title of the report from the US National Performance Review team: *Creating a Government that Works Better and Costs Less* (Gore, 1993). President Clinton, announcing the Review, elaborated upon this as follows:

Our goal is to make the entire federal government both less expensive and more efficient, and to change the culture of our national bureaucracy away from complacency and entitlement towards initiative and empowerment. (quoted on the frontispiece of Gore, 1993)

A similar set of ambitions (though with less emphasis on the empowerment of staff) were stated by the UK Minister of Public Service and Science in his foreword to the 1994 review of the Next Steps program:

I have two clear priorities for the Civil Service. These are to continue to raise the standards of public service; and, at the same time, to continue to improve the efficiency with which these services are provided to minimise the burden on the taxpayer. (Chancellor of the Duchy of Lancaster, 1994: i)

Interestingly, in a 1993 speech, the same minister had made explicit reference to the popular American text which inspired the US Gore report (Osborne and Gaebler, 1992) and described the UK reform program in the following terms:

It took some time for the reform movement to become comprehensive. But I now believe it has become pretty well comprehensive and total. Achieving something like what Osborne and Gaebler have written about: the ‘reinvention’ of government... it is a reinvention designed to produce responsive management and to raise standards of service within existing resource levels.... It has increased, not diminished, the individual’s practical control over the public services provided to him or her. It has replaced a system where that control, exercisable in theory, was routinely subverted in practice by producer interests. (Waldegrave, 1994: 82; for an Australian example, see Sedgewick, 1993: 2)

In some ways Osborne and Gaebler represent a public sector update of an earlier, even more popular book, *In Search of Excellence* (Peters and Waterman, 1982). Certainly the two books share a penchant for the upbeat, mini case study as their main mode of discourse.

Earlier, in 1989, the Canadian Prime Minister launched a Public Service 2000 initiative claiming that its objective was:

... to foster and encourage a public service that: is... imbued with a mission of service to the public; recognizes its employees as assets to be valued and developed; and places as much authority as possible in the hands of front-line employees and managers. (quoted in Auditor General of Canada, 1993: 161)
Our final quotation is from the UK Prime Minister, John Major. Shortly after succeeding Mrs Thatcher, he identified himself in a quite personal way with the Citizen’s Charter. In the introduction to the first Charter White Paper he wrote that his new program:

... sets out the mechanics for improving choice, quality, value and accountability. Not all apply to every service. But all have a common objective: to raise the standards of public services, up to and beyond the best at present available. (Prime Minister, 1991: 2)

In sum, therefore, we can say that leading politicians in a number of countries believe that the NPM will yield greater economy, greater efficiency, rising standards of public service, keener ‘ownership’ and enhanced autonomy for service managers/providers and, last but not least, greater responsiveness by staff to the users of public services of all kinds. This ambitious list provides us with a preliminary set of criteria against which to judge the success or otherwise of this particular genus of reform.

Is it too Soon to Evaluate?

Some features of the NPM are too new for their full effects yet to be observable—especially in those countries which did not begin to adopt NPM approaches on any appreciable scale until the late 1980s or early 1990s, such as Germany, Norway or Finland (Naschold, 1995: 13–54). Elsewhere, however, deeper experience has already accumulated. The UK and New Zealand have each had large-scale reforms underway for more than a decade. In other countries particular elements of the NPM have been in place for still longer periods, e.g. performance-related pay in the US federal service or decentralized executive agencies in Sweden. Even some of the more recent entrants to the NPM stable have now garnered 5 years or more of ‘track record’—Canada launched its Public Service 2000 initiative in 1989 and Australia embarked upon its Financial Management Improvement Plan (FMIP) in 1983 and subsequently restructured its cabinet and created a Management Advisory Board in 1987 (Halligan and Power, 1992; Zifcak, 1994).

In short, it is not too early to begin to assess progress and to look for results, though it probably is too soon to come to any firm conclusions about some of the latest NPM fashions such as the application of ‘business process re-engineering’ or radical decentralization of the authority to set the pay and conditions of staff.

Some Problems in Evaluating the NPM

Before inspecting the findings of existing evaluations it appears sensible to ask what would be the methodological requirements for studies likely to generate reliable and valid information on NPM. Yet the majority of the analyses referred to in this article afford either limited space or no space at all to the analysis of methodological difficulties. The purpose of this section is therefore simply to summarize what these difficulties are. In the following section we shall examine the extent to which existing
evaluations have addressed them. Subsequently, in the final section, we shall explore the possibilities for more adequate evaluations of the NPM in future.

Some of the more obvious, gross difficulties are the following:

1. New Public Management reforms are usually multifaceted (e.g. budgetary reform accompanied by a decentralization of authority, setting of quality standards and installation of new IT systems). When a particular effect or impact is identified how can the evaluator decide which reform facet or component produced it?

2. More often than not other changes in the political/administrative environment are going on at the same time as NPM reforms (e.g. budget cuts at the same time as decentralization of managerial authority). How can the evaluator know whether a particular effect is generated by the NPM reform, or the simultaneous shifts in other environmental variables? Both this problem and the previous one are essentially issues of attribution.

3. Even when particular effects can confidently be attributed to this or that NPM reform, what is the appropriate comparator for the new state of affairs? The baseline performance of the organization before the NPM reform may not be appropriate because in some (many?) instances the organization’s performance under the pre-NPM regime may have been changing anyway (for better or worse). So there may well be a need for the evaluator to construct a ‘counterfactual’ (for an extended discussion of the role of counterfactuals in social science argument, see Elster, 1978).

4. There is a need to theorize the politico-organizational contexts into which particular reform elements are being injected as well as analysing the characteristics of the reforms themselves. What works in one context may not work in another (a particular reform ‘plug’ may only fit one kind of organizational ‘socket’—see Pollitt, 1995).

5. Defining and measuring the costs and benefits of NPM programs pose a number of difficulties. There is a need to identify and measure the transitional (one-off) costs (and benefits) of implementing the reforms. These could include the direct costs of investment in new IT, documentation, training and informing the public of changes. There could also be ‘Hawthorne-effect’ benefits. More ambitiously one might wish to estimate the opportunity costs (e.g. staff time and attention diverted to learning about new practices, meaning that the quality of existing services declines for an interim period). Equally, if not more important, there is the question of continuing (recurrent) costs and benefits. Among these may well be significant changes in transaction costs. Such changes may be positive (i.e. the service is easier to run following the reforms) or they may be negative (many commentators have remarked upon the increase in transactions costs which can accompany shifts from unified hierarchical administration to markets or quasi-markets with their paraphernalia of contracts and supporting IT).

6. Most fundamentally, what criteria should be used to evaluate the reforms? The most common and in many ways easiest approach is to measure the identifiable effects of a particular reform against its stated objectives or goals. There are, however, at least two drawbacks. First, such an evaluation may miss unintended
Christopher Pollitt

effects (because it is not looking for them, e.g. if a reform is ‘sold’ in terms of improving the efficiency of a public service but actually achieves this at the expense of a covert loss of equity). Second, some of the ‘official’ goals may themselves be very difficult to pin down (e.g. to empower service users, to increase the pride of public service staff in their work).

Each of these problems deserves further discussion. The first and second immediately raise fundamental epistemological and methodological issues. The large-scale, multi-faceted, simultaneous nature of many recent changes in public management will often rule out controlled, experimental evaluations. Yet the promise of the standard OXO quasi-experimental design is also distinctly limited (Pawson and Tilley, 1994). Another alternative, well-designed case study could perform useful exploratory and descriptive functions, and some testing of the implicit theories underlying NPM (Yin, 1984), but case studies alone will not provide a synoptic picture of the extent of reform and neither are they a particularly convincing way of measuring the transitional and continuing costs and benefits of reform. In general terms some sort of mixed method approach appears to be necessary, and this idea will be elaborated in the final section of the article.

How do the Existing Evaluations Measure up?

In one sense it is easy to summarize the findings of existing evaluations. The official ones are almost uniformly strongly positive, albeit with some words of caution about the scope for further improvements (Chancellor of the Duchy of Lancaster, 1994; Employment Service, 1994; Task Force on Management Improvement, 1992; State Services Commission, 1990, 1992). For example:

...performance has been impressive. As the individual entries in the Review show, agencies are producing consistent improvements in performance for the benefits of customers and taxpayers. (Chancellor of the Duchy of Lancaster, 1994: i)

A quick reading of the Next Steps Review would seem to provide warrant for this optimism. After all, 102 executive agencies have met 80 percent of their targets—is this not adequate confirmation of the value of this program? An evaluator’s response might be ‘only if the targets are themselves evaluated’.

Many other commentators are, to differing degrees, more cautious (Dunleavy and Hood, 1994; Halligan and Power, 1992; Moe, 1994; Naschold, 1995; Pollitt, 1993; St-Martin, 1994; Stewart, 1994; Zifcak, 1994). What is more, this is not simply a case of optimistic governments vs sceptical academics. A close reading of even the official assessments reveals some significant caveats, often resting quietly in the less salient parts of the respective documents. Where the commenting agency has a high degree of independence the caveats become more obvious:

...some of the executives we met wondered whether PS2000 was ‘dead’, and we detected an atmosphere of scepticism and cynicism surrounding the renewal initiative. (Auditor General of Canada, 1993: 173)

or:

140
...our management reviews of 23 large federal agencies and departments done over the last
decade consistently have shown that many agencies lack the fundamental underpinnings that
will be needed to implement the NPR report’s recommendations. (General Accounting
Office, 1993: 3)

What we are concerned with here, therefore, is not so much the conclusions of these
assessments, interesting though they are, but rather the strength of their foundations.
How adequate are the assessments upon which governments and other experts are
proclaiming the success—or forthcoming success—of the NPM?

This question can be answered by a detailed examination of how existing studies
tackle the six problems enumerated in the previous section. Insofar as a particular study
addresses all six problems explicitly and appropriately, we may have confidence in its
conclusions. Insofar as some or all of these problems remain unaddressed, or are
tackled in methodologically inappropriate ways, our confidence in the findings of that
particular report will be reduced. Our review is sequenced in the same order in which
the six problems were introduced in the previous section.

Attribution Problems: The Multifaceted Nature of the NPM and
the Presence of ‘Other Factors’

While many of the reports acknowledge these first two difficulties few seem to do more
than that. For example, in what is generally one of the more supportive academic
analyses, Kettl states that:

Assessing the NPR is difficult—there is no such thing as the NPR. In practice, the NPR has
been a messy and sometimes disorganized multi-front war against the government’s
performance problems. (Kettl, 1994: 5)

Yet elsewhere the author makes strong claims that ‘The NPR produced more than
almost anyone, including perhaps the reinventors themselves, believed possible’ (Kettl,

The problem of the multifaceted (and multivalued) nature of most countries’ NPM
reforms is of more than academic interest, not least because it seems quite probable that
reform packages frequently contain contradictory ingredients, so that the effectiveness
of some particular changes is reduced or cancelled out by other changes. Thus, for
example, the bulk of evidence seems to be that PRP often has a neutral or negative
effect on the majority of staff and yet, despite critical analyses and a poor track record,
it continues to be featured as a prominent component of many public sector reforms in
many countries (for doubts about PRP see: Alban-Metcalfe, 1994; Management
Advisory Board, 1994: 10; OECD 1993b; Perry et al., 1989; Pollitt, 1995; Talbot,

More fundamentally, the tensions between, on the one hand, cost cutting and down
sizing and, on the other, quality improvement, are widely acknowledged in the more
independent commentaries, yet ignored or brushed aside in most ‘official’ accounts
(e.g. Chancellor of the Duchy of Lancaster, 1994; Gore, 1993; Prime Minister, 1991).
Kettl puts it vividly in relation to the US NPR:
...the ghost of deficit reduction lurked behind every promise of empowering workers or improving performance. ... There was a major disconnect between the ‘works better’ and the ‘costs less’. (Kettl, 1994: 6)

The Canadian Public Service 2000 initiative appears to have suffered similar problems:

PS2000 is clearly perceived as having been undermined by parallel initiatives, such as expenditure reductions. (Auditor General of Canada, 1993: 177)

Or again, in Trosa’s assessment of the UK’s Next Steps reforms:

...even if agencies have different kind [sic] of targets, financial (productivity), efficiency and quality, the enquiry shows that financial targets are given a much higher priority. (Trosa, 1994: 9)

More generally, a survey of 3800 UK public sector managers found the need to reduce costs to be the single most important factor driving change (Talbot, 1994: 28–31).

In Australia, too, the financial elements of broad-scope reforms appear to have been the most forcefully implemented (Task Force on Management Improvement, 1992: 580, Figure C.1; see also Management Advisory Board, 1994: 8). In a survey of agencies many responses supported this interpretation, for example:

On the other hand, the overall cash limited budget has been an outstanding success... (Attorney General’s Department, in Task Force on Management Improvement, 1992: para. 4.6)

Recent research into the impact of more decentralized management in British schools, hospitals and social rented housing has begun to unearth confirmation of the tensions between different parts of the NPM package. Where decentralization has taken place within very tight budgetary constraints, local managers and professionals tend to take a fairly cool view of its overall impact. In other cases, however, where decentralization has been accompanied by special funding, a much rosier appreciation prevails (Birchall et al., 1995).

The strength of the cost-cutting theme in most NPM reform programs poses a further interpretive difficulty: namely, where significant productivity increases have been achieved, how far may these be attributed to new-style management (decentralized, flexible, etc.) and how far are they simply the result of relatively traditional bureaucratic responses to budget cuts? In other words, would much of the productivity gain have been achieved simply by insisting on budget cuts and leaving officials to get on with adjusting to the consequences? Something of this may be seen in some UK universities, where huge productivity gains have apparently been accomplished in the period since 1990 (many more students, few extra staff, a rapidly falling unit of resource yet no fall in the apparent quality of degrees). But better NPM-style management has by no means necessarily been the principal cause; rather it has been a case of academics reluctantly cutting per student class contact times, teaching much larger classes, reducing the number or length of written assignments, sacrificing time for research and scholarship, and so on. All of which has been driven by relatively arbitrary budget reductions by central government.
This rather sensitive possibility—that much of the measurable productivity improvement may be attributable to the grudging reactions of service providers faced with crude cuts rather than to the benefits of new management structures and processes—remains largely unexplored in all the official evaluations referenced in this article. A review of the New Zealand reforms does at least mention the possibility:

Our terms of reference required us to report on whether the reforms had improved the efficiency and effectiveness of the Public Service. We have no direct way of answering this question, for two reasons: firstly [there is no before and after data]; second, the three years since 1988 has been a period of ... considerable structural change and financial pressure on departments independent of the reforms, which no doubt also contributed to whatever changes in costs might have occurred. (Steering Group, 1991: 26–7)

An independent commentary on the early implementation of the US Government Performance and Results Act (GPRA) puts the point more generally:

There does not seem to be sufficient recognition that most outcome information that will be reported under GPRA will provide only a 'score' indicating how well the characteristic being measured is doing. Outcome indicators will, in general, not tell the extent to which the program has caused the observed outcomes. (National Academy of Public Administration, 1994: 4)

**Benchmarks and Counterfactuals**

The absence of clear benchmarks against which to measure the increments of improvement attributable to the implementation of an NPM reform package is a problem that tends to be understated in most evaluations. The huge evaluation by the Australian Task Force on Management Improvement at least acknowledged the difficulty:

There are few reliable benchmarks with which to compare current performance and, since the reforms took place at a time of rapid social and economic change, there is no definitive way of separating the impact on cost, agency performance and clients (among other things) of these broader changes and of the governmental decisions which accompanied them. (Task Force on Management Improvement, 1992: 8; for a very similar comment on the New Zealand experience see Steering Group, 1991: 27)

A further example may be found in the UK National Health Service, where an independent academic evaluation of the introduction of TQM schemes noted that 'few of the NHS sites carried out any diagnostics or benchmarking at the outset of the TQM initiatives' (Centre for the Evaluation of Public Policy and Practice, 1994: 19). In another case—that of the UK Citizen's Charter—the government did make some attempt to establish benchmarks, in that a commercial firm was hired to undertake a survey to ‘form a baseline against which progress in achieving improvements could be measured’ (ICM Research, 1993: prefatory summary). Unfortunately, however, the survey was carried out nearly 2 years after the Charter program had been launched. What is more, the published version of the report adopted the odd convention of grouping service users’ ratings of recent service performance change into just two categories, ‘Worse job’ and ‘Better job/same’. This produced quite high percentages in
the latter box, but left the reader in the dark as to whether this was because most respondents had said that services were (despite the Charter) just the same or because they had seen an actual improvement (ICM Research, 1993: 14).

Even where clear benchmarks or baselines can be identified, the evaluator still faces a challenging interpretive difficulty, which may be summed up in the question 'what would the outcome have been if the reform program had not been implemented?' To take a baseline, measure the change since that baseline and assert that the difference represents the impact of the program is methodologically insufficient (which does not, of course, prevent politicians and others from doing it all the time). It is insufficient because it assumes that the baseline situation was static, and that all subsequent change was caused by the reform program and the reform program alone. In practice, however, the situation would seldom have been static. More likely organizational performance was already improving, or declining, because of pre-existing forces for change. So the real measure of the impact of the new program should be where the organization is now compared with where it would have been if it had continued developing as it had been immediately prior to reform. The latter yardstick is necessarily hypothetical, and is referred to in much social science literature as a 'counterfactual' (Elster, 1978).

None of the many evaluations we have examined actually develops a detailed counterfactual. A few (only) acknowledge that there is therefore a problem. The Australian Task Force on Management Improvement recognizes the difficulty in respect of expenditure restraint:

... the capacity of government to manage expenditure has been aided by the new expenditure control framework... the difficulty is... in particular, to devise an objective measure of what might have happened in the absence of the forwards estimates system. In the absence of such a measure, the best evidence on the matter is the professional judgement of those most closely involved. (Task Force on Management Improvement, 1992: 20)

Elsewhere in the same report particular agencies allude briefly to their own 'counterfactuals':

Many of the reforms would have occurred anyway because of the greater complexity of Government and society. It is hard to speculate on what might have been achieved. (Attorney General's Department, in Task Force on Management Improvement, 1992: para. 3.3)

or:

It is difficult to isolate one source of cultural change from another but our view is that we would have changed anyway, reflecting a general shift in values in the broader community and the public sector. (Australian Bureau of Statistics, in Task Force on Management Improvement, 1992: para. 3.2)

In an academic analysis of the impact of quasi-market reforms on the UK National Health Service the same issue arises. Ministers have made considerable play of the fact that some of the early NHS trust hospitals have registered notable efficiency gains since the internal market commenced on 1 April 1991. Researchers point out, however, that these hospitals were probably a self-selecting elite. They already displayed above average efficiency before the implementation of the internal market, and may well have opted for trust status partly because they knew themselves to be on a rising trend
Whether a counterfactual would show that improvement since the quasi-market commenced was merely a continuation of earlier trends or represented a new, ‘additional’ acceleration, stimulated by the market, therefore remains open to debate.

As a final example we may quote an in-house study of the benefits of granting agency status (greater freedom to manage within a framework of objectives and targets—a typical NPM reform) to the UK Employment Service. With commendable candidness the evaluation team began by reporting that:

No benchmarks relating to output targets, or to many of the areas covered in this evaluation, were established pre-Agency, and it is therefore difficult to quantify by direct comparison with earlier performance the effects of Agency status on the performance of the ES. (Employment Service, 1994: 5)

They then go on to observe that:

...it is unlikely to be possible to disentangle the effect of Agency status from other elements of ES performance; and ... it is unlikely to be possible to create an exact picture of what would have happened if ES and Agency status had not occurred. (Employment Service, 1994: 10)

Twenty pages later, however, the conclusions are reached and the tone has changed markedly:

The establishment of the Employment Service (ES) as an Executive Agency has created a firmly performance-driven organization, clearly focussed on the achievement of its key objectives. ... The establishment of ES as an Agency has added considerable value to the implementation of labour market policies and has improved the workings of the labour market. (Employment Service, 1994: 29)

We cite this example not with a view to criticizing the team that undertook the evaluation (as a matter of fact it was rather a thoughtful one, and crisply presented) but so as to illustrate a more general feature of official evaluations of NPM reforms—in the UK, the US and Australasia. So often there is a brief early mention of some potentially quite fundamental methodological limitation to the study, then nothing more is heard of it and a set of conclusions is presented that confidently claims that major benefits flow from the reform in question.

Theorizing Contexts

This fourth evaluatory issue can be briefly dealt with. Basically none of the broad-scale official evaluations reviewed within this article makes any kind of systematic allowance for different organizational contexts (where ‘context’ might be understood as embracing factors such as the organization’s purposes, culture, technology and market situation). Usually the NPM ‘recipe’ (delegated budgets, market testing, contractualization, etc.) is discussed as though it should be applied pretty much everywhere. Just occasionally the importance of different contexts is alluded to—as when the New Zealand Steering Group on State Sector Reforms’ recommendation No. 37 observes that:
Treasury should consider, in consultation with departments, the further differentiation of reporting requirements on departments based on their size and the risks involved. (Steering Group, 1991: 24)

The possibility that fundamentally different management styles, techniques and structures might be appropriate for different functions or levels is excluded from consideration. Yet this has long been a theme of public administrationists (e.g. Pollitt, 1995; Stewart, 1992) and of the contingency literature in organization theory (e.g. Donaldson, 1985: Ch. 14; Mintzberg, 1979). It is in the light of this body of theory and evidence concerning the importance of difference contexts that Dunleavy and Hood (1994: 15), writing of the global spread of the NPM, warn that:

If the history of public management reform tells us anything, it is that much of this cloning and imprinting will be deeply inappropriate to particular cases.

It must therefore be counted as a source of concern that the issue as yet lacks any systematic treatment in governmental assessments of their NPM programs.

Estimating Transitional and Continuing Financial Effects

This constitutes the fifth major problem for evaluators. Administrative change has transitional (one-off) costs and benefits of its own. These comprise direct costs (at a trivial level, the costs of new stationery; more substantially, the costs of training staff to follow new procedures) and opportunity costs (staff time—at all levels—being taken up with adjusting to the new arrangements rather than with their immediate operational responsibilities). In the case of the introduction of the National Curriculum to British schools it is said that more than £600 million was spent in detailed, sequential changes because of the impracticality of the curriculum as originally introduced.

In addition, reforms obviously create continuing or recurrent benefits and costs. For example, if a program can now be delivered with 20 percent fewer staff this saving will repeat each year until the next set of changes comes along. Interestingly, the clear identification of such costs and benefits occurs only occasionally in the set of evaluations we have scrutinized. For example, the Australian Attorney General’s office is fairly typical when it remarks that:

It is impossible to quantify the cost/savings, but a far greater range and volume of demands by ministers is able to be met with ongoing resource levels. (Attorney General’s Department, Task Force on Management Improvement, 1992: para. 4.3)

In the US NPR, behind the headline-catching promises of aggregate savings, the situation appears similar:

In almost all cases, putting hard numbers on savings actually produced, beyond the downsizing of the federal workforce, turned out to require quite extraordinary feats of budgetary analysis. (Kettl, 1994: 9)

Transactional costs are one (but only one) important component in this equation of continuing effects. They can be rather difficult to track down because they may occur in different organizational locations from those in which the more obvious kinds of costs and benefits arise. Take the case of the internal market reforms in the UK NHS. On the
one hand faster throughput of patients seems to be being achieved in many acute hospitals (though whether to attribute that all to the reforms is, as indicated above, problematic). This would appear to be a continuing benefit, and one that can be measured relatively easily. At the same time, however, the business of getting these patients in and out of hospital may well have become rather more complicated. Patient flows now have to be captured in contracts/service agreements between purchasers and providers. Each side of this quasi-market relationship has to maintain staff with expertise in the technical aspects of contracting in a way that was not necessary under the previous administrative hierarchy, where trust and hierarchical authority substituted for detailed accountancy. Much more elaborate information systems now have to be installed so that providers can attribute costs with much greater precision than was formerly required. In other words transaction costs have risen, at least initially (Robinson and Le Grand, 1994: 257-8).

The NHS example can be generalized to other NPM contexts. Considering ‘hierarchist’ critiques of the NPM Dunleavy and Hood remark that:

If contracting becomes the norm, hierarchists insist on the need to develop patterns of high trust ‘obligational contract relationships’ … rather than low-trust ‘arms’-length contract relationships’ … within the public service, in order to retain the capacity to negotiate across organizational frontiers without massive transaction costs in a crisis. (Dunleavy and Hood, 1994: 12)

All of which makes the task of the evaluator extremely complicated. A snapshot-in-time approach to costs and benefits may be quite misleading insofar as it confounds once-only and continuing costs and benefits, and may ignore longer term dynamics (up or down) in continuing costs. A concentration on foreground throughput measures may also mislead, unless it also takes into account possible background sea changes in the level of transaction costs or the vulnerability of the system to a loss of trust between disaggregated NPM agencies. The timing of measurement can thus be crucial, according to the relative magnitude of short-term (transitional) costs and benefits as against longer-term (continuing) costs and benefits. There are also bound to be learning effects—perhaps contracting is a cumbersome, staff-intensive process in the first couple of years but can thereafter be streamlined? Reporting demands can also be very time consuming in a newly decentralized environment (for example, see remarks of the Australian Attorney General, in Task Force on Management Improvement, 1992: para. 4.7), though they may be streamlined when supervisory authorities become more comfortable with the new regime. Ideally some kind of explicit model of the temporal dynamics of change might be set up, so that costs and benefits could be hung on its various branches and various points in time, working out from the ‘ground zero’ of reform.

In practice, however, the studies reviewed in this article do little to disentangle the foregoing complexities. Brief references to the problems are made but they are not pursued in any depth. For example, the New Zealand State Service Commission found that large efficiency gains had accrued from the establishment of state-owned enterprises (SOEs) and then mentioned, without quantification, that ‘some qualifications have to be attached to this success … [because] some of the gains are one-off’
Christopher Pollitt

(State Services Commission, 1990). The UK government’s annual reports on the Next Steps program certainly include many figures for individual year-on-year ‘cost improvements’ but avoid any broader discussion of differences between one-off and continuing effects, or of any of the other interpretive problems referred to above (Chancellor of the Duchy of Lancaster, 1994). Naschold (1995: 45–7) reports a rare Swedish analysis which links changing cost levels to alternative local government management regimes. Interestingly, this study seems to reinforce the importance of the temporal dimension in that:

... the data available tend to suggest that the relative cost advantage for privatized enterprises tends to decline over time. Many privatized activities initially have low costs, but then successively raise their remuneration demands vis a vis local government. (Naschold, 1995: 46)

A number of important studies do not address the issue of specific, quantified costs and benefits at all. The Australian Task Force on Management Improvement acknowledges that there are problems in assigning savings to the reform program but then quite quickly moves on to other considerations (Task Force on Management Improvement, 1992). In New Zealand a review was set up to examine, inter alia, ‘what has happened so far as a result of the reform programme; what benefits are being realized and what ongoing costs are being incurred?’ The resulting 143-page report contained no figures for costs and benefits at all, but was nevertheless able to conclude that:

... the framework of of the reforms is sound and substantial benefits are being realized. (Steering Group, 1991: 11)

Alternatively, where broad-scope and quantitatively sophisticated analyses of public sector productivity have been attempted it has proven virtually impossible to connect macroeconomic changes to specific NPM reforms. An elaborate Swedish study of public sector productivity 1980–92 was obliged to leave the issue of attributing gains and losses to particular changes largely unresolved (Expertgruppen for studier i offentlig ekonomi, 1994). An ambitious analysis conducted for the Finnish government by a German academic tried to find some connection between the adoption of NPM-type reforms (‘market-steering’ and ‘governance by results’) and various countries’ macroeconomic performance. Using OECD figures this study came to the conclusion that the thesis that NPM reforms led directly to better economic performance:

... has to be strikingly refuted: all the countries with bureaucratic governance by rule [the traditional, unreformed regimes such as Germany] exhibit with respect to almost all dimensions a markedly better macroeconomic performance than the other countries. (Naschold, 1995: 39)

Choice of Criteria for Evaluation

Few of the reviews—either official or independent—make a clear statement of the criteria by which they are judging the progress of reform. In most cases general criteria emerge by implication, and are essentially echoes of the declared objectives for the
relevant programs, such as increased efficiency, greater decentralization of management authority, higher quality of service, and so on. As we have seen, these criteria are applied only loosely. None of the documents referenced here attempts an explicit summary of the extent to which the reforms have saved expenditure or increased user satisfaction, though some do parade quantities of agency-specific performance indicator data which appear to show widespread improvements (the problems of interpreting these have been dealt with above).

This approach may be regarded as sound as far as it goes, but as not going very far. It usually omits or backgrounds consideration of other criteria or values which are normally deemed highly relevant for public sector reforms (Ranson and Stewart, 1994). These include equity and equality, honesty and public accountability. The last is the only one to receive much coverage in the evaluations under review, and even that is patchy (e.g. Steering Group, 1991: 28, 32). In several countries the expanding use of market and quasi-market mechanisms has certainly provoked popular and academic fears of declining equity and/or equality (‘two-tier services’), but this concern has not been a significant focus for any of the official evaluations reviewed here. Within public services the possible implications of decentralized human resource management policies for equal opportunities for staff are occasionally referred to, but in total there is little discussion and fewer data (e.g. Task Force on Management Improvement, 1992: question 4.4 to Australian Business Statistics). Yet early indications are that competitive tendering, for example, may have distinctly and disproportionately negative effects on female staff (Escott and Whitfield, 1995). Official concern with the creation of an increasing number of opportunities for self-interested behaviour by public officials (under decentralized budgetary, contractual and personnel regimes) has been similarly muted, despite growing academic and parliamentary anxiety (Public Money and Management, 1995; see also Management Advisory Board, 1994: 12).

Summary and Conclusions: Our Current State of Knowledge Concerning the NPM

To many it seems that the NPM has become ‘the only show in town’ (Dunleavy and Hood, 1994: 10). Certainly there has been no shortage of politicians and senior managers—in each of the countries covered here—who have been willing to testify to its success. It is the argument of this article, however, that the underpinnings of the NPM’s reputation are actually quite fragile. Such evaluations as have been undertaken are either silent about their methodological limitations or acknowledge only sotto voce major problems of attribution, the absence of benchmarking, the lack of relevant cost data and other problems. This is not to suggest that nothing has changed. Indeed, there is much convincing testimony that working life has been transformed for many public servants. There is also some evidence of perceived changes—both positive and negative—for the users of public services. The problem is that we often do not know whether these are generated by NPM measures or by other factors, or which particular NPM elements are working well and which are not. Nor do we know much about what other consequences such reforms may be producing in terms of inequalities, inequities or increased exposure of public services to self-interested or downright dishonest

149
behaviour. The impact of NPM on values and attitudes is underresearched. Support for
the NPM package as a whole is based more on faith and doctrine than on demonstrable
track record.

If this conclusion is accepted, then at least two practical consequences appear to
follow. The first is that an effort is needed to inject some alternatives into current
reform debates. The NPM formula (see Table 1 above) has become too dominant, too
automatically resorted to, too universal for its own good health. ‘Another important
characteristic that has limited the utility of many of the reforms is their assumption that
all organizations are the same’ (Peters and Savoie, 1994: 422). Perhaps more
differentiated solutions could be canvassed for different public sector contexts? Dunleavy
and Hood (1994: 10), for example, point to the dangers of applying the NPM
where organized crime and corruption threaten the integrity of public services and
argue that, in such circumstances, the virtues of the traditional model of a rule-
governed public bureaucracy deserve reconsideration. There is not space here to take
this argument forward into detail, but it would certainly be refreshing if some of the
yardage of text advocating or attacking the NPM could be replaced by comparisons
with alternative managerial ‘horses for courses’.

The second consequence is that we need better evaluations of the NPM. Of course,
there is no ‘quick fix’ for all the evaluatory hurdles identified above. But equally it is
not difficult to see how our current state of knowledge might be significantly improved.
We therefore conclude this article by sketching what some of the components of such
evaluations might look like.

New Public Management reform programs normally comprise a diverse set of
activities which affect large, complex organizations dynamically, at different levels,
over extended periods of time. Most official evaluations have tackled them by taking
relatively brief snapshots of what appeared to be happening at a particular point in time.
The most popular method has been to talk to a selection of key players, especially
political leaders and top public service managers and to report what is in essence the
‘balance of opinion’. This has sometimes made it difficult to distinguish between
changes in language use and deeper changes of value and behaviour. Quantitative data
have frequently been collected, but in many cases the resulting statistics are either
unreliable or hard to interpret.

Of the official evaluations only the Australian one (Task Force on Management
Improvement, 1992) seems to have gone much beyond this. They undertook large-scale
surveys of both Australian public service staff and citizens. Such surveys would be one
way of significantly building on our present platform of knowledge in most ‘NPM
countries’. They hold the promise of yielding information vital to a fuller appreciation
of NPM outcomes. In particular they throw light on how the middle and lower levels of
the public service experience NPM and—most important—on the extent to which the
citizens in whose name so many reforms are launched have noticed any changes (and if
they have, how they value them). It was noticeable in the Australian case, for example,
that important aspects of the reforms were much more critically regarded at middle and
lower levels than among senior managers (see also Auditor General of Canada, 1993:
166) and that a majority of the public had not noticed any improvement in the quality of
service (Task Force on Management Improvement, 1992: 38–40). Survey questions

---

*Christopher Pollitt*
need to be carefully framed to capture factual, experiential knowledge and not just to evoke like/dislike opinions. On the issue of the final impact of reforms, the knowledge of grassroots staff and/or citizens may well be superior to that of the senior managers, whose opinions more frequently figure in official (and many academic) accounts.

Surveys are useful, but there is a clear need to go beyond the single snapshot (not least to answer the questions such surveys may raise). Staff and public views may well fluctuate through different phases of a reform program (for example, staff may transit through, first, apprehension, then high expectations, then disappointment, then boredom). A longitudinal element would seem to be a vital element in better evaluation. Longitudinal, in-depth studies of selected, but typical reform elements (delegated budgets and personnel authority for example) would be particularly valuable. Carefully chosen, such studies could tackle several of the characteristic weaknesses we have noted in the studies to date. They could identify a clear baseline at the commencement of the reform effort. They could track the dynamics of implementation down and up through different levels of the organization(s), from policy-makers to frontline staff and service users. They could make explicit allowance for parallel or prior sources of change, thus beginning to address some of the problems of attribution. They could even test rival causal models of the reform process.

It would be particularly interesting to attempt such studies on a comparative basis, perhaps examining selected functions in one more strongly pro-NPM country with their counterparts in Germany or some other ‘non-NPM’ state.

Finally, it is not simply a matter of asking the same questions but of a wider audience and over a longer period of time. There is also a clear need to broaden the range of questions which are posed. In particular, it is noticeable that few evaluations have devoted much effort to exploring the relationships between NPM, public accountability and wider constitutional issues. In the UK the government’s stance has been that the NPM has improved public accountability without causing any significant constitutional change (see Waldegrave, 1994; and, for a contrasting view, Stewart, 1994). Not all official commentators have been so sanguine, however. The New Zealand Steering Group was an honourable exception:

We sometimes got the impression that some chief executives assumed the reforms were intended to give them complete autonomy over their operations or at least to shear off any accountability other than to their Ministers. Neither assumption is true. (Steering Group, 1991: 28)

In Australia the Task Force on Management Improvement felt it necessary to warn that there was ‘a need in government to avoid any implication that a public servant is directly accountable to the client they are dealing with …’ (Task Force on Management Improvement, 1992: 37).

Indeed, one of the major questions that peeks out from under the corner of some of the official evaluations is that of the role of politicians in the NPM world. Will they be prepared to give up their habit of ad hoc micro management of those cases, projects or programs that catch their attention? Are ministers equipped instead for the strategic role of supervising a complex, highly decentralized system of public service delivery?

A theme which emerged in the New Zealand review was:
Christopher Pollitt

... that Ministers were under-resourced to fulfil their objectives in the areas of strategically sound decision-making and holding their departments accountable.

Also, among public servants:

... some considered that Ministers are disinclined to think strategically, at least in so far as that requires them to commit themselves to a set of outcomes within the context of a corporate plan. (Steering Group, 1991: 64, 47)

Once one moves to commentaries by more independent sources the expressions of concern become more acute. In his survey of 3800 public managers Talbot found that the biggest gap between rhetoric and practice occurred on the issue of political 'interference'. Managers recorded a major decline in their ability to stand clear of political interference in operational decisions (Talbot, 1994: 38-41). Kettl (1994: viii) concludes that in the US 'the NPR seeks to shift power from Congress to the bureaucracy and, within the bureaucracy, from top to bottom levels' and finds little evidence that Congress is willing to go along with this. Dunleavy and Hood (1994: 16), looking across the international experience, reflect that:

... what is at stake in these reforms are not just bread-and-butter issues of operations, costs and short-term response. Ultimately, the issues are constitutional, in that they affect the foundations of political life and capacity.

It is almost certainly naive to hope that internal reviews could ever address these sensitive political issues in a thoroughly open way. The need is for a mechanism that will give evaluators both strong rights of access and firmly rooted independence. Such an approach would be expensive, time consuming and probably somewhat embarrassing to executive politicians and their senior officials. Yet after 15 years of upheaval right across so many public sectors an evaluative response of this kind does not seem disproportionate to the stakes involved.

References


Christopher Pollitt


CHRISTOPHER POLLITT is Dean of Social Sciences and Co-Director of the Centre for the Evaluation of Public Policy and Practice (CEPPP) at Brunel University, Middlesex UB8 3PH, London. In addition to his academic research and teaching he has undertaken work for the OECD, HM Treasury and a variety of other public organizations. (email: Christopher.Pollitt@brunel.ac.uk)