ELSEVIER

International Public Management Journal 3 (2000) 23–46

International Public Management Journal

The challenge of evaluating systemic change: the case of public management reform

Jonathan Boston*

School of Business and Public Management, Victoria University of Wellington, PO Box 600, Wellington, New Zealand

Abstract

At what stage of reform in the public sector does it become possible to conduct a thorough appraisal of results and how does one know when this stage has been reached? How should such an assessment be undertaken? By what methods can comprehensive and far-reaching *systemic* reforms be evaluated in the arena of public management during recent decades, particularly in countries like Australia, Britain and New Zealand? Most assessments have focused upon specific changes in management practice including the introduction of performance pay, the move to accrual accounting, the growth of contracting-out, the separation of policy and operations or the devolution of human resource management responsibilities. Alternatively, they have dealt with management changes in particular policy domains –such as health care, education, community services or criminal justice –or within a particular organization (department, agency or private provider). By contrast, there have been relatively few *macro* evaluations –comprehensive assessments of the impact of root-and-branch changes to the system. The problems of evaluation in the arena of public management are inherently complex and the way ahead is by no means clear. This article offers some broad reflections on the limitations to policy evaluation in the field of public management, and more particularly explores the obstacles confronted when assessing the consequences of systemic management reforms. It focuses on recent changes in the New Zealand public sector to illustrate the general themes because these reforms constitute one of best examples of systemic change anywhere in the world. © 2000 Elsevier Science Inc. All rights reserved.

* Tel.: +64-4-463-5456; fax: +64-4-463-5084.

E-mail address: Jonathan.Boston@vuw.ac.nz (J. Boston).

PII: S1096-7494(00)00033-7

1. Introduction

Since the late 1980s a steady stream, if not a flood, of ministers, government officials, academics and consultants from many different parts of the world have made the long trek to the distant islands of New Zealand, and more specifically to its capital, Wellington, to examine first-hand the radical policy changes which have transformed that country's system of public management. Invariably, these earnest investigators besiege their hosts with countless probing questions. Common among them are the ostensibly simple, yet profoundly difficult, questions: Have the reforms been good for the country? Has the experiment been successful? Has it worked? And should the new model of public management embraced by New Zealand be emulated by other jurisdictions?

As one of those who is a regular recipient of such questions, I am often reminded of the response which Mao Tse-Tung is reputed to have given when asked his opinion concerning the consequences of the French Revolution: 'It is too early to tell'. While at one level this answer is rather frivolous, it also makes the crucial point that the effects of major political, economic or technological changes or radical shifts in key areas of public policy are likely to be felt over long periods of time; it is thus imprudent to attempt a full and definitive assessment of their consequences too early. But this begs further questions: at what stage does it become possible to conduct a thorough-going appraisal and how does one know when this stage has been reached? Furthermore, how does one actually go about the task of undertaking such an assessment? Or to be more specific, how should one evaluate comprehensive and far-reaching *systemic* reforms, such as those that have occurred in the arena of public management during recent decades, particularly in countries like Australia, Britain and New Zealand?

Such questions, of course, are not new. Indeed, they have engaged the energies of a good many scholars and policy advisers over the years. But the relevant literature is by no means extensive or beyond scrutiny. To be sure, there is a substantial and rapidly expanding body of material, including case studies and comparative analyses, which evaluates aspects of the new public management. However, as Pollitt (1998a: 65) has observed, much of the scholarly research thus far has been at the *micro* level and 'highly context specific'. That is to say, it has focused upon specific changes in management practice, such as the introduction of performance pay, the move to accrual accounting, the growth of contracting-out, the separation of policy and operations or the devolution of human resource management responsibilities. Alternatively, it has dealt with management changes in particular policy domains –such as health care, education, community services or criminal justice – or within a particular organization (department, agency or private provider). By contrast, there have been relatively few *macro* evaluations–that is, comprehensive assessments of the impact of root-and-branch changes to the whole system of public management (e.g., situations where there have been major policy changes across many different dimensions of management within multiple policy sectors). One notable exception was the large-scale inquiry undertaken in the early 1990s by the Task Force on Management Improvement (1992) which examined the reforms to the federal public service in Australia over the preceding decade. But notwithstanding its considerable length (over 600 pages), depth and breadth, the Task Force's report has been the subject of various criticisms, including those of a methodological nature (see Pollitt, 1995). Moreover, as readily acknowledged by its authors, the assessment was partial and incomplete.

This leads to a further point: despite the large literature which recent public management reforms has generated, relatively little of it grapples with the epistemological, conceptual and methodological problems that beset policy evaluation in this arena. All too frequently important methodological issues are ignored or glossed over. As a result, claims are made about the impact of particular management changes for which the evidence is either thin or of questionable validity. Likewise, while the difficulties of evaluating *specific* changes in management practice (e.g., the introduction of performance-based remuneration) have attracted attention (certainly in the academic literature), much less consideration has been given to the problems of assessing *systemic* changes. Nor is there an extensive literature dealing with the issue of how to evaluate the relative merits of different public management paradigms or models.

It is with some trepidation that I draw attention to such matters since the problems of evaluation in the arena of public management are inherently complex and the way ahead is by no means clear. Nevertheless, in the interests of stimulating some debate, this article offers some broad reflections on the limitations to policy evaluation in the field of public management, and more particularly explores the obstacles confronted when assessing the consequences of systemic management reforms. I have deliberately focused upon New Zealand's recent reforms to illustrate the general themes covered in the article, partly because these reforms constitute one of best examples of systemic change anywhere in the world and partly because of my familiarity with the subject matter.

Two caveats should be noted before commencing the task. First, this article is not designed to provide a comprehensive, rigorous and in-depth evaluation of the New Zealand model of public management, or indeed any other model. Such a task, if it were possible at all, would be a massive undertaking. Rather, my concern is to identify the various kinds of hurdles that must be surmounted if a comprehensive assessment were to be seriously contemplated.

Second, the fundamental problems in the field of policy evaluation are not so much technical in nature (though there is no shortage of such issues), but ethical and political. As Parsons (1995: 550) argues, policy evaluation 'is knee-deep in values, beliefs, party politics and ideology'. Not merely do the normative and ideological commitments of the evaluator influence the nature of the assessment conducted, they also affect the way the evidence is interpreted and hence the conclusions which are drawn. Among other things, this means that all claims about the effects or merits of particular policies need to be treated with caution.

2. The challenges of evaluating systemic change

How, then, should one begin the task of evaluating systemic change, and more particularly systemic change in the field of public management? An obvious answer is to employ the standard tools and techniques of policy evaluation. Yet this approach raises two immediate difficulties. First, all the usual evaluation techniques pose serious conceptual, methodological and, in some cases, ethical issues. Second, the question arises as to whether standard evaluation methods are adequate for the task of assessing changes of a large-scale or systemic nature.

2.1. General evaluation issues

The general problems of policy evaluation are well recognized by researchers and widely discussed in the relevant literature. Some of the issues are very complex and it be would be impossible to do them justice in this article. For the sake of brevity, however, the key problems facing policy evaluators can be summarized under four headings:

1. Choosing the appropriate criteria for evaluation and determining what constitutes ‘success’;
2. Determining and securing the relevant evidence;
3. Interpreting the available evidence, including the problem of establishing appropriate counterfactuals and determining causation; and
4. Arriving at an overall assessment.

Under each of these headings, a range of more specific issues can be identified, as illustrated in Table 1. For instance, in many instances there will be different perspectives on what constitutes ‘success’, or even what constitutes ‘an improvement’. This may arise because of conflicts over the goals and objectives of the policy or program in question, which in turn may reflect wider disputes of a moral or ideological nature. Alternatively, there may be agreement on the appropriate criteria for assessment, but disagreement over the weighting of these criteria or over the threshold which must be crossed in order for the policy or program to be deemed successful (Savoie, 1998: 396). A related problem is that views (whether on the part of immediate stakeholders or the wider community) concerning what constitutes success are not static. Consequently, a policy deemed ‘successful’ at one particular point in time may be subsequently judged less positively. Thus, as Pollitt (1998b: 7) argues, ‘what constitutes a “good” public service in the eyes of a diverse, multicultural citizenry will itself be (at any given moment) somewhat diverse and (over time) subject to change. No amount of measurement can “solve” value differences of this kind’.

A second difficulty encountered when evaluating public policies is determining what to measure and how to measure it. Take, for instance, the issue of ‘quality’: in the field of public management the quality of the goods and services funded (and often provided) by the state is of pivotal importance in assessing performance. Yet, as is widely recognized, measuring quality is notoriously difficult. The use of client or customer surveys may provide a broad indication of whether the quality of a service has changed, but such surveys are not without their drawbacks and limitations. After all, the users of a particular service may have only limited information, and in some cases may have impaired judgment.

When assessing the impact of a particular policy initiative it is common to compare, if possible, the results achieved after the relevant change with those before the new policy was introduced. But ‘before and after’ studies of this nature require reliable benchmarks or yardsticks against which comparisons can be made, and such benchmarks are often unavailable. Indeed, in the area of public management it is common for performance measures (or indicators) to be altered frequently (sometimes yearly), thus making intertemporal compar-

Table 1
Standard problems of policy evaluation

Conceptual issue	Examples of the kinds of problems encountered
Choosing the appropriate criteria for assessment and defining what constitutes 'success'	• The objectives set by policy makers are usually multiple and often conflicting • The objectives set by policy makers are often poorly specified, and are usually partial and incomplete • Evaluators often differ over the selection of the appropriate objectives • Evaluators often differ over what would constitute a successful policy, or even over what would constitute an improvement on the previous arrangement
Determining what to measure and securing the relevant evidence	• The relevant data are often unavailable or inadequate; in particular there is typically a lack of readily quantifiable data • Appropriate benchmarks are often lacking, so that it is difficult to compare 'like with like' or to make 'before and after' comparisons • Some of the relevant data are very expensive and time-consuming to obtain • The time period is too short for all the significant impacts to have become apparent
Interpreting the evidence	• The evidence is often conflicting • The appropriate counterfactuals are uncertain • It can be difficult to determine what weight to place on particular pieces of evidence, especially 'hard' versus 'soft' data • It can be difficult to isolate the specific impact of different elements of a reform package and determine whether specific outcomes are the product of the reforms or external factors
Arriving at an overall assessment	• Lack of agreement on the policy goals and the criteria for assessment will generate different verdicts depending on the values and weighting methods adopted by different evaluators • Measurement difficulties and the problem of incommensurability render it impossible to develop a composite index or express the varied data in a common currency • Different goals are typically achieved to different degrees and it is difficult to determine what weight to give to the realization, partial realization or non-realization of particular objectives • There is likely to be significant scope for disagreement about the policy implications of the analysis, as well as the applicability of the findings to other jurisdictions or contexts

isons all the harder. Even where reliable and consistent benchmarks are available, they may not represent the appropriate counterfactuals against which to measure the impact of a particular policy initiative. After all, various changes may have occurred to the relevant benchmark, for example, the demand for an agency's services, or its funding levels, regardless of any changes in the policy framework under investigation. Thus as Pollitt (1995: 144) argues, the appropriate measure of the impact of new policy or 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*'. Yet determining such a counterfactual is by no means straightforward and there will always be room for disagreement over the most realistic or objective measure of what would have happened if a particular reform had not occurred.

A related issue is the partial and incomplete nature of the performance information that is available. For instance, reasonably robust 'hard' data may be collected covering changes in an organization's staffing, funding, financial performance and output levels, but there may be little reliable 'soft' information concerning matters of a more intangible nature, such as trends in staff morale, capability, probity, goodwill or creativity. The fundamental problem, of course, is that all performance measures are subject to imperfection. Also, in most cases it is not possible to develop performance indicators that are sufficiently diverse and sophisticated to assess the full range of impacts that a particular policy may be having. To compound matters, the larger the number of measures employed, the more costly the exercise and the more complex the task of reaching an overall assessment.

Data interpretation represents a further challenge. Common difficulties include: conflicting data; the absence of crucial data; questions over the reliability or accuracy of particular data; and the challenge of balancing the more easily accessible quantitative data with considerations of a qualitative nature. Yet another difficulty arises from the fact that some policy changes may have longer-term impacts which take many years to become apparent. Equally, there is the problem of distinguishing between *transitional* and *continuing* effects, both in terms of costs and benefits.

A related difficulty centers upon the issue of establishing causation. It may appear, for instance, that a particular change in the framework of public management has produced its desired effect—fiscal savings, higher productive efficiency, more equitable access, greater client satisfaction, or whatever. But the observed change in the desired output or outcome measure may have been the product of factors unrelated, or largely unrelated, to the policy change in question. Determining causation, and isolating the impact of particular policy initiatives, is necessarily harder when multiple policy changes occur simultaneously—all the more so if they coincide with significant turbulence in the external environment.

Next, there is the challenge of reaching an overall assessment of the merits or otherwise of the policy or program in question, taking proper account of the wide range of (often conflicting) criteria, the competing views about how these criteria should be weighted, the imperfect and incomplete nature of the available performance information, the problems of data interpretation and the issues of causation. In the field of public management, as in most areas of public policy, there is the ever present problem of incommensurability; that is to say, the results of policy reforms typically occur in 'currencies' of an incommensurable nature—some are narrow financial measures, others are concerned with matters of an ethical nature, while yet others are concerned with social, environmental or political outcomes. Where the relevant performance measures are incommensurable, necessarily they cannot be reduced to a single currency or aggregated so as to construct a single index of performance (see Pollitt, 1998b: 5). In these circumstances, evaluators must be careful to distinguish the level of performance achieved in each of the relevant 'currencies' and avoid the temptation of reducing complex realities to simple numerical ratings of how 'good' or 'bad' things are.

Finally, there is the issue of drawing policy lessons from the assessment that has been undertaken. For instance, is the evidence sufficiently robust for evaluators to recommend changes to existing policy settings and, if so, what form should these changes take? Equally, there is the issue of whether any lessons drawn from the experiences of one particular jurisdiction (or context) are relevant to other countries (or contexts) and, if so, which ones. Whether the New Zealand model of public management is applicable to developing countries is but one illustration of the kind of issues which can arise (see Schick, 1998).

2.2. Specific problems evaluating systemic change

The preceding section has briefly summarized the kinds of problems that arise when evaluating specific changes at the policy or program level, such as the devolution of responsibility for human resource management, the contracting out of a particular service, the establishment of a new administrative unit or the introduction of output-based budgeting. These difficulties should not be exaggerated, but neither should they be underplayed. Yet if the evaluation of a discrete policy initiative poses significant challenges, the evaluation of major systemic changes, such as those affecting the whole public sector, is all the harder. Not merely is the task more demanding in terms of scale, scope and cost, but it is also more complex analytically and conceptually.

To start with, any systemic change is bound to have multiple objectives. There may, for instance, be some broad overarching objectives for the reform program as a whole, and then a large number of specific objectives associated with each major subcomponent. Other things being equal, the greater the number of objectives, the greater the likelihood that some of them will conflict, the greater the difficulties of assessing their relative importance, and the greater the potential for disputes over what constitutes success.

This points to a further problem: judging the success of a new 'system' or management paradigm cannot be accomplished merely by assessing the success of each of the component parts and then aggregating the results. Presumably, it is also important to consider the merits of the system as a whole. Does it cohere? Is it internally consistent? Are the parts properly integrated? What tensions are evident? Does the system facilitate good interagency coordination? Does it serve the government's collective interests (as opposed to the interests of individual ministers)? How well does it accommodate and adapt to external shocks? How responsive is it to changing political tides and societal preferences? Yet any attempt to assess a new system on the basis of such criteria will encounter formidable measurement difficulties. How, for instance, does one measure 'coordination', 'coherence', 'adaptability' or 'political responsiveness'? What indicators might be useful as assessment tools?

Of course, the measurement difficulties are not limited to those of assessing how well the system as a whole is operating, or comparing this with the previous system (or another reference point). Systemic change also poses greater difficulties when it comes to estimating the impact of individual policy changes. The problem, put simply, is that most of the benchmarks against which performance is traditionally assessed—whether relating to the performance of specific programs, organizations or individuals—will cease to have any relevance (and, indeed, may cease to exist). For instance, a systemic change in public management is likely to entail more-or-less simultaneous modifications at a number of different levels: there will be changes in organizational boundaries and responsibilities, in how public sector organizations are funded and managed, in information systems and processes, and in interorganizational relationships and policy coordination. And at the microlevel there will be changes in job descriptions, incentive systems and interpersonnel relationships. Obviously, where simultaneous changes occur to an organization's internal structure, management systems, financing, staffing arrangements and personnel, it becomes increasingly difficult to determine what to measure, and what performance benchmarks, if any, remain appropriate. Matters will be further complicated if the organization's 'outputs' and/or operating environment are significantly altered while it is undergoing major internal restructuring.

Yet, typically, whenever major systemic changes are initiated in the field of public management, entirely new systems are introduced for information management, performance management, and accountability. And if there is no continuity in the relevant performance indicators, such as output levels, quality measures, and so forth, then it will be all the harder to assess trends over time or undertake 'before and after' studies. This, of course, will not necessarily prevent assessments being made of the performance of individual agencies *under* the new model, that is, assuming that the new performance management framework and the wider policy environment remains reasonably stable over a number of years. But it is one thing to measure the progress of specific organizations against a relatively stable set of performance indicators in a relatively stable policy environment; it is quite another to compare the performance of two radically different models of public management –all the more so if the implementation of the new model coincided with important changes to the wider constitutional, political, economic, social and technological environment.

3. Problems in assessing systemic change: the New Zealand case

The kinds of problems encountered in evaluating systemic change in the field of public management are graphically illustrated by the experience of New Zealand since the mid-1980s. In the space of little more than five years—between 1985 and 1990—New Zealand undertook one of the most comprehensive and far-reaching programs of public management reform attempted anywhere during the latter part of the twentieth century, certainly within the OECD. In most respects the reform agenda was consistent with of the ideas, principles and doctrines of the new public management (Hood, 1991); indeed, some commentators regard the New Zealand model of public management as the best—and perhaps the most 'extreme'—example of managerialism (Kickert, 1999: 487).

Few would dispute that the reforms undertaken during the mid-to-late 1980s constituted a systemic change. After all, virtually every aspect of the public sector was altered to one degree or another. Major changes include:

- a systematic program of corporatization, privatization and commercialization;
- the devolution of human resource management responsibilities from the State Services Commission to the chief executives of individual departments and Crown agencies;
- the introduction of new financial management systems, including accrual accounting, an output-based system of appropriations, capital charges and a distinction (for the

purposes of resourcing, monitoring and accountability) between the Crown's 'ownership' and 'purchase' interests;
- the institution of major structural changes, such as the transfer of many service-delivery functions to separate, non-departmental agencies;
- The introduction of an extensive network of 'contractualist' devices to govern the relationship between 'principals' and 'agents' within the public sector, as well as between public and private organizations; and
- the development of a comprehensive system of strategic management under which ministers specify their strategic objectives over the medium term and use these to set departmental priorities.

We may note that these changes did not occur in isolation, but were part of a much broader program of economic, social and constitutional reform (see Boston, Dalziel and St. John, 1999; Kelsey, 1995). The nature of these latter changes need not detain us here. However, they undoubtedly compound the difficulties of evaluating the public management reforms.

3.1. Evaluations of New Zealand's public management reforms

New Zealand lacks a tradition of serious policy evaluation (see Turner & Washington, 1999). Surprisingly, for a country that has led the world in its preparedness to experiment with new approaches in many fields of public policy, decision makers have tended to place relatively little value on policy research and evaluation. It has been argued, for instance, that evaluation might stifle innovation, thereby inhibiting desirable policy changes (see Bushnell, 1998: 367). Other concerns have focused on the costs involved and the limited analytical capacity available, both within the public service and the academic community. It must be remembered that New Zealand is a small country in population terms (barely the size of a middle-sized city) and thus lacks the resources, both financial and intellectual, of larger democracies. Given the huge effort that has been expended in designing and implementing the extensive reform agendas of recent governments, the public sector has not had the luxury of devoting a significant proportion of its best talent to the task of policy evaluation. This explains the absence of any assessment of similar scope and scale to that undertaken in Australia in the early 1990s (see Task Force on Management Improvement, 1992).

It would, however, be utterly misleading to suggest that the public management reforms in New Zealand have not been evaluated. On the contrary, there is now a burgeoning number of studies and reviews—some by government agencies (especially the Audit Office, the State Services Commission and the Treasury), some by international agencies (such as the OECD and World Bank) and some by academics. Many of these are careful, patient, perceptive analyses, and furnish important insights into the strengths and weaknesses of the new model of public management.

Having said this, few of the studies thus far have provided a comprehensive or systematic assessment. Instead, most have dealt with relatively discrete or narrow areas of activity (such as contracting out, corporatization, financial management or strategic management); alternatively, they have concentrated upon particular policy domains (such as health care and education) or specific events (such as the Cave Creek tragedy and the deaths at Canterbury

Health Ltd.) (e.g., see Boston & Pallot, 1997; Campos & Pradhan, 1997; Doolin & Lawrence, 1997; Gregory, 1998; Henisz, 1999; Henry, 1995; Norman, 1995; 1997; Pallot, 1998; Shaw, 1999; Stent, 1998; Thompson, 1999). Where more systematic analyses have been attempted, they have generally been constrained in terms of their depth and rigor (e.g., see Halligan, 1997; Scott, Bushnell & Sallee, 1990; Kelsey, 1995; State Services Commission, 1998; Steane, 1999; Wistrich, 1992).

Other deficiencies in the existing literature can also be identified. Few of the studies give much attention to the problems of evaluating systemic change, some even failing to address the more general difficulties and limitations of policy evaluation; most utilize only a limited range of assessment criteria, with much of the focus being on efficiency and accountability; and few employ multiple evaluation techniques. Equally important, there has been a relative dearth of rigorous 'before and after' analyses. The main exceptions have generally been those concerned with the more commercial parts of the public sector and those dealing with financial performance (e.g., see Duncan & Bollard, 1992; Spicer, Emanuel & Powell, 1995).

Of the attempts to assess the new model of public management in a more holistic way, two were commissioned by the government. The first of these was conducted by Basil Logan in 1991, and the second by the distinguished American academic, Professor Allen Schick in 1996. Various academics have also undertaken more systematic assessments of the public sector reforms (see Boston, Martin, Pallot and Walsh, 1996; Easton, 1997). But while such studies have merit, it is fair to say that they are all characterized by important omissions and a failure to grapple adequately with the problems of assessing systemic change. For instance, the reviews by Logan and Schick apply only a limited range of criteria (principally those of an economic and financial nature) to the assessment of the reforms, and ignore significant elements of the new model (e.g., both focus primarily upon the departmental sector and both largely ignore certain important dimensions of public management, such as human resource management and service delivery). These remarks apply with equal validity to studies of New Zealand's reforms undertaken by organizations such as the Office of the Auditor General of Canada (1995) and the OECD (1996).

In recent years, the State Services Commission—the body with the legal responsibility for overseeing the performance of the public sector, especially the 40 or so government departments—has commenced a series of reviews of some of the key reform initiatives. Likewise, the Office of the Controller and Auditor-General has taken an increasingly active role in evaluating particular aspects of the new model, including the management of contracts, the quality of institutional governance, aspects of service delivery, the effectiveness of the new systems of financial management, and the quality of service performance reporting (e.g., see Controller & Auditor-General, 1996, 1999a, 1999b). Over time, studies of this nature will supply an increasingly rich database for evaluating organizational performance *under* the new model. However, neither the reviews by the SSC nor those by the Audit Office will provide a basis for comparing the performance of the new model with that which it replaced.

The weaknesses and limitations of the current literature assessing the public sector reforms in New Zealand are not due primarily to any lack of professionalism or expertise on the part of those who have attempted to evaluate the new model of public management —though there are undoubtedly cases where the analyses are methodologically flawed or seriously biased. Rather they reflect the sheer magnitude of the problems—as outlined in the previous section—that arise when evaluating major policy changes. Not surprisingly, the literature abounds with comments like: ‘It is difficult to assess the effects of these organizational changes’ (OECD, 1996: 123); ‘there is no reliable data’ (Logan, 1991: 26); ‘there is no reliable base point’ (Logan, 1991: 27); and the relevant information ‘is not generally available’ (Campos & Pradhan, 1997: 436). It might be added that such qualifications and limitations have not prevented some evaluators making remarkably bold claims about the virtues, benefits and wider applicability of the new model. The remainder of this article considers the key challenges evaluating New Zealand’s public management reforms as an illustration of the general problems of assessing changes of a systemic nature.

3.2. Selecting the criteria for evaluation

Most of the assessments of the New Zealand model have been based upon a relatively limited number of criteria. Evaluators have also displayed a tendency to choose those indicators that serve their particular interests, predilections or ideological orientation. Critics of the reforms, for instance, have tended to downplay or ignore economic and financial criteria (such as allocative and productive efficiency), while proponents have generally said little about the wider political, social and ethical ramifications of the reforms. Those evaluations undertaken (or commissioned) by government agencies have typically focused upon how the reforms have affected the following outcomes:

- the level of allocative and productive efficiency;
- the level of cost-effectiveness (or cost savings);
- the degree of managerial and political accountability;
- the level of public expenditure (both in absolute terms and as a proportion of GDP); and
- the quality and quantity of public services.

The selection of such criteria is usually justified—if it is justified at all—on two main grounds: they are consistent with the main objectives set by the architects of the reforms; and they are among the standard criteria applied internationally by those involved in evaluating public sector performance.

Yet while there is truth in both these claims, any appraised based solely upon these criteria will necessarily be partial and incomplete, if not seriously misleading. New Zealand’s reformers, at both at the political and bureaucratic levels, had many goals other than those just mentioned: some were articulated explicitly, others were rarely mentioned or even perhaps concealed; some were internally consistent, others were mutually incompatible. Among such objectives were the following: reducing the potential for producer ‘capture’, enhancing equal employment opportunities within the public sector, making public institutions more bicultural (in accordance with the principles of the Treaty of Waitangi), removing essentially commercial activities from direct public ownership, enhancing hierarchical control (including political control of the bureaucracy), and increasing dynamic efficiency (by making the public sector more innovative and entrepreneurial). There were also numerous lower-order objectives, such as recruiting more private sector managers into the public sector and freeing ministers from the ‘tedium of administration’ (Upton, 1999: 14) so that they could devote more of their energies to priority setting and strategic issues. Added to this, the management reforms in discrete policy domains, such as education, health care and scientific research, all had specific objectives, many of them distinct from the wider reform agenda.

Yet it would not be sufficient simply to list the reformers’ objectives and then employ these as the relevant assessment criteria. For one thing, the objectives were of widely varying importance, and this needs to be taken into account in any evaluation. For another, some of the objectives emerged during the latter stages of the reform program, often in response to perceived deficiencies in the original design. The goal of enhancing the government’s strategic management capacity is a case in point (see Boston & Pallot, 1997). Additionally, many of the reformers’ objectives were vague and cannot be operationalized (for evaluation purposes) without further clarification: what is meant, for instance, by the term ‘biculturalism’ when applied to the public sector?

More fundamentally, any proper evaluation of the New Zealand model of public management must examine not merely whether (or to what extent) the hopes and aspirations of the reformers were achieved, but also how the changes affected other important societal goals and values. It must also explore the unintended consequences of the reforms, both positive and negative. When considerations of this nature are brought into the equation, a large number of additional criteria can be identified, such as:

- the transitional costs in moving from one model to another, and how these costs have been distributed among those affected;
- the capacity of the new regime to cope with shocks, risk, uncertainty and cross-sectoral issues, and its effectiveness in responding to emergencies.

Other criteria might include the impact of the reforms on:

- the scope for judicial review (see Chen, 1999; Taggart, 1990);
- the distribution of power between politicians and public servants, and the policy influence and role of various groups, including employees, unions, professional groups, sector groups, and so forth;
- ethical standards within the public sector, including the level of fraud;
- the role of voluntary agencies and the level of social capital;
- citizen engagement in public life, the level of trust in public institutions, and the degree of confidence in the system of government;
- the willingness of citizens to cooperate with the authorities and engage in coproduction;
- the quality of information for decision making; and
- the capacity for policy learning, both by individual agencies and by the governmental system as a whole.

Needless to say, the above list is far from exhaustive. It ignores, among many other issues, the question of implementation gaps (i.e., the extent to which the reforms as actually implemented differed from those that were intended), the management of the reform process, the costs of compliance (both for providers and users of public services), the durability of the reforms and the impact of the new model on the political process, electoral outcomes, constitutional conventions, and the pressure for constitutional reform.

One way of making progress in such circumstances might be to adopt a two-stage process under which certain broad categories of criteria are delineated and then subcriteria (or more specific measures) are identified within each category. Table 2 outlines the kind of approach that could be adopted. Note, however, that the criteria identified in this table are not intended to be comprehensive or complete. Under each broad category many additional measures could be identified; and in respect of each of the so-called 'specific measures' further disaggregation could be undertaken. For instance, under the measure 'improving human resource management' one could readily employ a host of indicators, including: recruitment and retention; staff turnover; succession planning; staff training; the number, length and kind of industrial disputes; staff morale; the employment opportunities for women and ethnic minorities; and so forth. Of course, any attempt to apply such a wide range of criteria would be a daunting task.

3.3. Determining what to measure and finding the evidence

As already noted, those evaluating the New Zealand model face formidable hurdles in terms of securing relevant and reliable performance data. The specific problems are readily identifiable:

1. Under the former model of public management, the use of performance measures was relatively limited. Accordingly, in most policy areas it is hard to find appropriate benchmarks or baselines against which performance under the new model can be compared with that under the old.
2. Even where relevant benchmarks exist (e.g., concerning the performance of particular agencies, or the provision of specific outputs), the magnitude of the policy changes since the mid-1980s renders most of them of dubious value. For instance, in many cases there has been a lack of organizational continuity, with numerous public sector agencies having been privatized, corporatized or radically reshaped (in some cases numerous times). And in the few cases where there has been a degree of organizational continuity (e.g., the State Services Commission and the Treasury), significant changes have often occurred in relation to the level and kind of services provided by the agency in question. Additionally, virtually every organization has been affected by changes to its wider operating environment (e.g., the regulatory framework, public expectations, technological innovations, etc.).
3. Since the late 1980s, new performance management systems have been introduced throughout the public sector. Departments, for instance, are required to negotiate annual 'purchase agreements' with the relevant minister and provide Parliament with annual 'statements of service performance'. This has led to a vast improvement in the range and quality of the performance data available to evaluators. However, in most cases, there have been substantial changes during the 1990s in the way departmental outputs have been defined and in the nature of the performance measures employed. Such problems have been compounded by continuing organizational restructuring and

Table 2
Assessing public management reforms: possible criteria

General criteria	Specific criteria
1. Economic and financial criteria	● Improving productive efficiency ● Enhancing allocative efficiency ● Improving labor productivity ● Enhancing cost-effectiveness ● Minimizing transaction costs ● Minimizing compliance costs ● Controlling public expenditure (as a proportion of GDP)
2. Constitutional criteria	● Retaining a meritocratic public service ● Minimizing the arbitrary use of public power ● Avoiding inappropriate ministerial interference in management, regulatory and judicial decision-making ● Ensuring adequate ministerial and parliamentary control over public organizations ● Allocating public power to the appropriate level of government ● Consulting properly with parties affected by policy decisions ● Ensuring that public agencies fulfill their statutory
3. Accountability criteria	● Enhancing political accountability ● Enhancing managerial accountability ● Enhancing financial accountability ● Enhancing legal accountability ● Enhancing openness and transparency
4. Ethical criteria	● Minimizing corruption and fraud ● Minimizing conflicts of interest ● Ensuring probity ● Minimizing extravagance and waste ● Avoiding cronyism and nepotism (e.g. in ministerial appointments)
5. Institutional design criteria	● Choosing organizational mechanisms that are appropriate for the function in question ● Minimizing overlap and duplication ● Minimizing the scope for producer (and other forms of) ‘capture’ ● Minimizing turf protection and inter-agency rivalry ● Enhancing policy coordination ● Enhancing the clarity and quality of organizational governance ● Enhancing democratic control
6. Management criteria	● Improving human resource management ● Improving financial management ● Improving change management processes ● Improving risk management ● Improving strategic management ● Improving information management ● Improving project management ● Improving performance management
7. Ownership interest criteria	● Enhancing the efficient use of public resources; ● Improving the management of cash and fixed assets ● Preserving and enhancing organizational capabilities ● Enhancing organizational leadership

(continued on next page)

Table 2
(continued)

General criteria	Specific criteria
8. Purchase interest criteria	• Enhancing service delivery, in terms of access, equity, responsiveness, etc • Improving the quality of policy advice • Improving policy research and evaluation
9. Client/citizen responsiveness criteria	• Enhancing the quality of services, including public safety • Enhancing customer/client satisfaction • Enhancing cultural sensitivity in service delivery • Enhancing access to, and availability of, public services • Enhancing equity of access • Enhancing consumer choice (where appropriate) • Enhancing the quality of information about services • Enhancing consumer advocacy services • Enhancing grievance procedures • Enhancing opportunities for redress in the event of inadequate services
10. Information-related criteria	• Improving the quality of information available to policy makers • Improving the quality of information available to managers • Improving information flows within individual organizations and across the public sector • Improving the integration of budgetary, strategic and performance management data
11. System-wide criteria	• Enhancing the government's capacity to manage its collective interests • Enhancing policy coordination and inter-agency cooperation • Enhancing the capacity of the public sector to respond to changing political imperatives and external shocks • Enhancing public confidence in the institutions of government

regular changes to the volume and kind of outputs purchased by the government. Accordingly, there remains a lack of good time series data.

4. In terms of purchasing and reporting arrangements, the New Zealand model has focused primarily on 'outputs' (i.e., the goods and services provided by agencies) rather than 'outcomes' (i.e., the impacts on, or consequences for, the community of the outputs or activities provided by the government). This means that ministers purchase outputs, not outcomes, and that agencies report their performance almost exclusively on an output basis. Moreover, with only a few exceptions, ministers are not obliged to provide regular reports to Parliament on whether their desired outcomes are being achieved. Thus, although the annual budget and departmental annual reports contain references to hundreds of outcomes, relatively few outcome measures are reported on a consistent or regular basis. For such reasons, the performance management regime offers little help when it comes to assessing the wider social and economic impacts of the public sector reforms, or indeed the effectiveness of specific policy interventions.
5. Yet even with respect to organizational outputs, the measurement of performance is far from complete. Not all outputs have well-developed performance measures; some

performance indicators are of doubtful relevance; and, as a general rule, the quantity measures are much better than those concerned with quality. This, of course, simply reflects the inherent difficulties of measuring the quality of many outputs—policy advice being but one example (Boston et al., 1996; Di Francesco, 1999). But even for those outputs where performance measurement is somewhat easier, there remains 'little hard evidence concerning technical efficiency, unit costs, productivity and the like' (Campos & Pradhan, 1997: 436; see also Norman, 1997: 27). Similarly, there have been few robust studies conducted examining the costs and benefits of external contracting—which has increased markedly since the late 1980s.

6. If the measurement of technical efficiency and productivity poses difficulties, it will always be much harder to assess the impact of systemic change on the more intangible aspects of organizational performance and the quality of public life. Included here are such factors as: organizational culture; staff morale, dedication, loyalty and commitment; the capacity and willingness to innovate; the level of trust, cooperation, reciprocity and information sharing between agencies; the degree of 'rent-seeking' behavior; management 'responsibility' as opposed to 'accountability'; and the virtues of probity, decency and integrity. While a number of researchers have offered thoughtful reflections on such issues, it is fair to say that a great deal remains unknown (and perhaps unknowable) (see Barnes, 1999; Gregory, 1998, 1999; Martin, 1994).

In sum, there are serious difficulties securing the necessary evidence to assess many aspects of the New Zealand model. As noted, most performance indicators since the late 1980s have focused upon levels of activity or process issues; much less attention has been given to quality, effectiveness or substantive results (see Ryan & Brown, 1998; Ryan, 1999). But this is only part of the problem. The dearth of reliable evidence is readily acknowledged even by those at the center of the machinery of government—who presumably have access to the widest possible range of quantitative and qualitative data. This was underscored in late 1999 by the State Services Commission in its briefing to the incoming government. In a remarkable observation, the Commission (1999: 25) stated: 'We know a lot more than a few years ago about the quality of management processes within departments, but we do not know quite enough to form authoritative judgments about: overall organizational performance; whether the capability of departments is growing or eroding over time; and their ability to respond to government priorities and their external environment over the next two or three years'. Bear in mind that this statement was made more than a decade after the enactment of the major public sector reforms, and notwithstanding a huge increase in the volume of performance data being provided to the central agencies by individual departments.

One other point is worth noting at this juncture: as efforts are made to improve the quality of performance management systems and the data available to those charged with steering the ship of state, the impact that such systems have on the actual behavior of those being assessed must not be ignored. For instance, in New Zealand's case the regime governing the appropriation of public funds and the assessment of service performance (introduced via the Public Finance Act 1989) has had a number of negative and unintended consequences. One of these stems from the undue emphasis placed upon outputs and the commensurate undervaluing of processes and the government's 'ownership interest' (i.e., its interest in the longer-term capability of public agencies). To quote Thompson (1999: 518–519): 'From this limitation flows a short-termist view of the enterprise, as funders seek to obtain returns on their contributions within the accounting period and ignore investments and divestments affecting other periods. . . . Service performance reporting . . . disallows the recognition of inputs required to maintain or enhance the machinery'.

3.4. Interpreting the available evidence

The limited nature of the available data is but one problem; another is making sense of such information as is available. The challenge here is not merely determining what has changed, and by how much, but how the various observed changes can be accounted for. Take, for instance, the question of fiscal savings. There is reliable evidence that significant savings have been achieved since the late 1980s in relation to the cost of administering governmental programs, especially within the departmental sector (Treasury, 1996). Yet which of the various management reforms (e.g., the decentralization of personnel management, the removal of most input controls or the new accountability framework for departmental heads) contributed most to this outcome is far from clear. Moreover, some, if not most, of the reported savings may well have occurred if the previous bureaucratic regime had been retained but subjected to a similarly heavy and protracted dose of fiscal stringency.

All this, of course, points to the enduring problem of counterfactuals. Suppose, for the sake of argument, that none of the major management reforms had been introduced. Suppose there had been no systematic program of corporatization, no expansion in the use of contractualist devices, no significant devolution of human resource management, and no reform of the financial management regime. What outcomes might then have been expected? What kind of performance might the old order have delivered? Needless to say, determining the appropriate counterfactuals is well nigh impossible. Plainly, the public sector would not have stood still. Various innovations would still have occurred. Many organizational boundaries would doubtless have been altered, and new agencies created. There would still have been changes of government, not to mention changes in the level of demand for various public services. Furthermore, the revolution in information technology, with all its ramifications for employment opportunities, internal management systems and the pattern of service delivery, would have occurred regardless. To complicate matters, not much help can be gleaned by exploring the performance of other systems, such as Australia, Denmark, Ireland or Sweden, where public management reforms followed other paths; there are simply too many contextual differences to enable robust international comparisons to be made.

Conflicting evidence represents a further challenge. Almost regardless of which aspect of the New Zealand model is examined—be it the purchaser-provider split, the separation of policy and operations, the output pricing system, the strategic management regime, staff training and succession planning, the management of information technology, the governance and management of Crown entities, the extent to which public agencies are fulfilling their obligations under the Treaty of Waitangi, the level of corruption, or the mechanisms for political accountability—there are indicators pointing in different directions and contrary opinions on the merits of the new order. Let me mention three brief examples.

1. The program of corporatization, introduced in the mid-1980s, has been widely acclaimed for the benefits it generated, both in terms of improved service delivery and better financial results (see Duncan & Bollard, 1992; Spicer et al., 1995). Yet much less attention has been paid to the 'costs' of corporatization, including the loss of certain forms of accountability (Taggart, 1990) and the social and expenditure implications of the mass redundancies necessitated by the establishment of the new state-owned enterprises (SOEs). Equally, the financial performance of the SOEs remains uneven, and the application of the SOE model to the health sector in the early 1990s is generally agreed to have been less successful (on virtually any measure) than the initial corporatization program. In a recent evaluation of the SOE model, Henisz (1999: 366) offered the following assessment: 'The extended agency-theory-based experiment with corporatization of state-owned enterprises, despite initial efficiency gains, failed to yield private sector levels of efficiency or produce the sustained dynamic efficiency gains hoped for. Political intervention became increasingly common and disclosure standards frequently evaded'.
2. One of the defining features of the New Zealand model is its reliance upon contractualist modes of control and performance management, both at the individual and organizational level. Yet assessments of this approach (e.g., see Boston, 1995, 1998; PUMA, 1999; Schick, 1996; Scott, 1997) have produced very mixed findings—some claiming that contractualism has delivered substantial improvements in efficiency, accountability and control; others maintaining that any benefits have been offset by the negative effects, such as higher transaction costs, protracted negotiations, reduced operational flexibility, less adaptability in dealing with unforeseen events, reduced trust and goodwill, goal displacing behavior, 'contract fixation' and what Schick (1996: 84) calls 'accountability by specification'. If external commentators are divided in their views, so too are those responsible for negotiating the contracts. In recent years, senior ministers have both commended and criticized aspects of the contractualist framework, in some cases challenging its theoretical foundations as well as its practical application. As a minister in the former National-led government put it: 'The theory . . . relies heavily upon Ministers playing their role as principals in a contractual regime comparable to a marketplace. We are expected to be energetic and well-informed purchasers, monitoring output delivery and bringing particular sanctions and pressures to bear as required. The reality is far from a market model. It is characterized more by monopoly supply, compliant demand, arbitrary prices and asymmetry of information' (Upton, 1999: 12).
3. Moving from the general to the particular, controversy has raged in recent times over the performance of the department responsible for providing welfare assistance to beneficiaries (Work and Income New Zealand) and its chief executive, Christine Rankin. On the one hand, there is good evidence that the management reforms since the late 1980s have contributed to a significant improvement in service delivery in this area (as measured, for instance, by the time taken to process requests for welfare assistance) (Petrie, 1998). On the other hand, the new department has been the subject of intense public scrutiny since early 1999 as a result of claimed extravagance on staff training, the mismanagement of applications for student allowances, and long delays in

the processing of applications for student loans. For instance, in February 2000 some 23,000 students were sent loan contracts containing seriously inaccurate information. Performance failures can, of course, occur even in the best of organizations. But the number and scale of the recent problems at Work and Income New Zealand certainly raises questions about the quality of the organization's management. More generally, they pose an ever-present problem for evaluators: in assessing organizational performance, how much weight should be given to specific cases of major success or failure, and how much account should be taken of the broader context in which the organization is operating?

In each of the above cases there are important issues concerning causation. For instance, the corporatization program might have been a greater success, at least in strictly financial terms, if it had been introduced in more buoyant economic conditions. Similarly, some of the problems with the contractualist framework might be lessened if resources were not so tight or if ministers (as the main 'principals') had different expectations or capabilities. And the performance issues which have arisen at Work and Income New Zealand might have been avoided if the department had not recently undergone a major restructuring, or if it had not been required to undertake major new responsibilities, or if different appointments had been made at senior management levels. In short, many of the problems that have arisen under the new model might not have occurred if the external environment had been different, or if the reforms had occurred in a different way or at a different pace.

The analysis thus far has been concerned primarily with the issue of conflicting evidence over the merits and effects of *particular* aspects of New Zealand's reformed public sector. Yet there is another level at which the evaluation can be conducted. How satisfactorily is the new model operating as a whole? How well does it cohere? And how well does it cope with the enduring dilemmas of public management, such as the balancing of conflicting political imperatives, and the balancing of short-term and long-term priorities? Again difficulties arise in addressing such questions. To start with, as already noted, there is the problem of how system-wide performance should be assessed. Next, for each of the possible indicators of performance at this level—the views of stakeholders, the durability of the new framework, instances of serious internal inconsistencies, duplication or dysfunction, the extent to which the collective interests of the government are being protected—there is evidence pointing in contrary directions.

Let us take, for instance, the views of one important group of stakeholders, namely citizens. In 1998 Perry and Webster conducted a major survey of New Zealanders' attitudes and values concerning politics and government. This revealed that citizens have a relatively low level of confidence in the public service, and that the level of confidence had fallen significantly since the beginning of the reforms in the mid-1990s. For instance, in 1998 over 70% of citizens had either 'not very much' confidence or 'none at all' in the public sector; this compared with 52% when a similar survey was conducted in 1985 (Perry & Webster, 1999). Likewise, whereas only 26.1% of respondents said that they had 'quite a lot' of confidence in 1998, in 1985 the figure was 43.6%. Such evidence might be interpreted as meaning that citizens have a rather dim view of the new model of public management. Against this, client satisfaction surveys (which are regularly conducted by many public agencies) generally show that the public is well satisfied with the quality of services being delivered by individual agencies. Moreover, the low level of confidence in the public service revealed in the 1998 survey may have been strongly influenced by the high degree of disenchantment with the political system at this particular time. If a survey were conducted in early 2000, the results could well be rather different. Alternatively, the results may reflect not so much the virtue or otherwise of the reformed public sector but rather the impact of a decade of fiscal restraint on the funding of public services. Be that as it may, the 1998 survey results should serve as a caution to those who believe that New Zealand citizens have firmly embraced the new model of public management (see also SSC, 1999: 15).

With respect to the durability of the new model, there is also competing evidence. On the one hand, virtually all the major building blocks of the reformed public sector (including the key legislative changes) have remained largely unaltered for more than a decade. Presumably, this indicates that parliamentarians find the new model generally acceptable, if not desirable. On the other hand, the reforms in particular sectors, such as health care, have already been significantly amended. Further, the center-left government elected in late 1999 has expressed disquiet about many aspects of public sector performance and has announced its intention to make various changes. These include the field of health care where it seems likely that the purchaser-provider split will be largely abandoned and where the major public providers of secondary care services (currently Crown-owned companies) will be returned to their former status of statutory corporations governed by semirepresentative boards. Meanwhile, there is growing pressure for a reversal of the institutional reforms that occurred in the area of scientific research, and continuing disquiet over the large number of new organizations that the reforms have spawned. At another level, the Controller and Auditor-General (1999a) has called for major changes to the financial management system introduced in the late 1980s, including a greater emphasis on outcomes and the longer-term capability of Crown-owned organizations. Such recommendations reflect serious misgivings about important elements of the new model.

In sum, regardless of where one turns, and regardless of which performance indicators are employed, there is conflicting evidence on the merits of New Zealand's reformed public sector. This applies whether one is concerned with specific aspects of the model or the model as a whole. Such conclusions should not be interpreted to mean that New Zealand's public management reforms have achieved very little or that no informed judgments can be made about the relative merits of different parts of the model. But they do serve as a caution against rash and extravagant claims, whether positive or negative. They also highlight why verdicts on the New Zealand model, although generally affirmative, have been far from uniform or consistent.

3.5. Arriving at an overall assessment

I have no intention here of furnishing my own appraisal of New Zealand's bureaucratic reforms—various colleagues and I have attempted this task on other occasions and, I fear, without sufficient sensitivity to the complexities involved (see, for instance, Boston et al., 1996). I would, however, like to offer a few brief comments.

First, any overall assessment of the New Zealand model must not ignore the difficult political, economic and social context within which it was designed and implemented. It was a period of significant electoral volatility and dissatisfaction with the nation's political institutions; indeed the disenchantment was sufficient to generate a major change in the electoral system (a relatively unusual development within the life of democratic countries). It was equally a period of radical economic restructuring, fiscal retrenchment and relatively low growth. Almost certainly, the implementation of the new model would have been easier had the economic conditions been more propitious. In all likelihood, too, various elements of the model would have worked more successfully in a more relaxed fiscal environment. For instance, there would have been more funds available for capital injections and for expenditure on staff training and development, the level of demand for certain public services (e.g., welfare assistance and community services) would have been lower, and many public servants would have faced a less stressful workplace environment. Having said this, I readily acknowledge the difficulties of trying to isolate the contribution which specific contextual factors made to the overall performance of the new bureaucratic order.

Conversely, the impact of the public management reforms on the wider environment cannot be ignored. There can be little doubt, for instance, that the way the reforms were implemented added to public frustrations over the integrity of the political system and contributed to the demands for constitutional change. More controversially, it is possible that the reforms have weakened the policy capability of the public sector, thus increasingly the likelihood of imprudent or poorly informed decisions. It would, of course, be impossible to estimate the economic (or other) costs of any such impact, but potentially they could be very large.

Second, despite the fact that the reforms were initiated at least 15 years ago, I suspect that many of the effects have yet to be fully felt. Only in the past few years, for instance, have a significant number of senior managers been appointed who were not recruited—and thus socialized—under the previous bureaucratic order. To put it differently, the character and operation of the new model—including the norms, values, expectations and ethical standards of those who serve within it—is still being strongly influenced by various legacies, not least the ethos and culture of an earlier cadre of officials. It will be more than a decade before those mentored under the old regime no longer inhabit the senior ranks of the public service. Only when this generational transition occurs will it be possible to assess more accurately the impact of the reforms on the culture of the public service.

4. Conclusion

This article has deliberately emphasized the difficulties of evaluating changes in the field of public management, especially changes of a systemic nature. It might appear from the foregoing discussion that the task is simply too complex and too demanding and that we should abandon it. Such a pessimistic conclusion, however, is unwarranted. To be sure, the methodological challenges are formidable. But I do not believe that they are utterly insurmountable. At the same time, they certainly highlight the need for researchers to be humble, cautious and realistic.

In concluding, I would like to make two brief pleas. First, notwithstanding all the difficulties just mentioned, I believe greater effort should be made to conduct research, within the field of public management, which explores performance at the macro or system-wide level. This is not to devalue individual case studies and other kinds of research that focus on specific policy arenas, individual organizations or particular management issues. Such studies are vital if we are to gain a richer understanding of the different parts of the whole. But research at the more micro level typically gives insufficient attention to the interactions between the different parts of the broader system and thus misses potentially important dimensions, for example, interdependencies of various kinds, interorganizational dynamics, coordination issues.

Second, and related to this, more discussion and debate is needed on the question of what criteria should be employed when seeking to evaluate the performance of a whole system (or competing models) of public management, and how these criteria might be applied in practice. For instance, what performance indicators should be used and what evaluation techniques and strategies are likely to be the most fruitful? Unless I am mistaken, the literature dealing with such matters is relatively modest, thereby providing significant scope for some lateral thinking and innovative research agendas.

Acknowledgments

Jonathan Boston is Professor of Public Policy, Victoria University of Wellington, New Zealand. This article is a revised version of a paper presented at the Sydney 2000 International Public Management Conference, March 4, 2000. I would like to thank my colleagues Bob Gregory, John Martin and Richard Norman for their constructive and helpful comments on an earlier version of this article.

References

Barnes, C. (1999). *Public sector corruption under centralized and decentralized models: a theoretical framework. MPP Research Project*. Victoria: University of Wellington.

Boston, J. (Ed.). (1995). *The state under contract.* Wellington: Bridget Williams Books.

Boston, J. (1998). Public sector management, electoral reform and the future of the contract state in New Zealand. *Australian Journal of Public Administration, 57* (4), 32–43.

Boston, J., Dalziel, P., & St. John, S. (Eds.). (1999). *Redesigning the welfare state in New Zealand.* Auckland: Oxford University Press.

Boston, J., Martin, J., Pallot, J., & Walsh, P. (1996). *Public management: the New Zealand model.* Auckland: Oxford University Press.

Boston, J., & Pallot, J. (1997). Linking strategy and performance: developments in the New Zealand public sector. *Journal of Policy Analysis and Management, 16* (3), 382–404.

Bushnell, P. (1998). Does evaluation of policies matter? *Evaluation, 4,* 363–371.

Campos, J., & Pradhan, S. (1997). Evaluating public expenditure management systems: an experimental methodology with an application to the Australia and New Zealand reforms. *Journal of Policy Analysis and Management, 16,* 423–445.

Chen, M. (1999). The diminishing scope of judicial review. Paper presented to the Fourth Annual Administrative Law Conference, 24–25 February.

Controller & Auditor-General. (1996). *Report on governance issues in crown entities.* Wellington, November.

Controller & Auditor-General. (1999a). *Third report for 1999.* Wellington.

Controller & Auditor-General. (1999b). *Report on inquiry into certain events concerning the New Zealand tourism board.* Wellington, April.

Di Francesco, M. (1999). Measuring performance in policy advice output: Australian developments. *The International Journal of Public Sector Management, 12* (5), 420–431.

Doolin, B., & Lawrence, S. (1997). Managerialism, information technology and health reform in New Zealand. *The International Journal of Public Sector Management, 10* (1/2), 108–122.

Duncan, I., & Bollard, A. (1992). *Corporatization and privatization: lessons from New Zealand.* Auckland: Oxford University Press.

Easton, B. (1997). *The commercialisation of New Zealand.* Auckland: Auckland University Press.

Gregory, R. (1998). Political responsibility for bureaucratic incompetence: tragedy at Cave Creek. *Public Administration, 76,* 519–538.

Gregory, R. (1999). Social capital theory and administrative reform: maintaining ethical probity in public service. *Public Administration Review, 59,* 63–75.

Halligan, J. (1997). New public sector models: reform in Australia and New Zealand. In J. Lane (Ed.), *Public sector reform. Rationale, trends and problems*. London: Sage.

Henisz, W. (1999). The institutions and governance of economic reform: theoretical extensions and applications. *Public Management: An International Journal of Research and Theory, 1* (3), 349–371.

Henry, D. (1995). Reform of a major service delivery agency–not platitudes but action. *Administration, 43* (1), 16–35.

Hood, C. (1991). A public management for all seasons. *Public Administration, 69* (1), 3–19.

Kelsey, J. (1995). *The New Zealand experiment.* Auckland: Auckland University Press.

Kickert, . (1999). Discussion of the paper of Sandra Dawson and Charlotte Dargie. *Public Management, 1* (4), 483–487.

Logan, B. (1991). *Review of state sector reforms.* Wellington: State Services Commission.

Martin, J. (1994). Ethics in public service: the New Zealand experience. In N. Preston (Ed.), *Ethics for the public sector: education and training.* Annandale: Federation Press.

Norman, R. (1995). New Zealand's re-invented government: experiences of public sector managers. *Pubic Sector, 18,* 22–25.

Norman, R. (1997). *Cases in public sector innovation: accounting for government.* Wellington: Victoria Link.

OECD. (1996). *Economic surveys: New Zealand.* Paris: OECD.

Office of the Auditor General of Canada. (1995). *Towards better governance: public service reform in New Zealand (1984–94) and its relevance to Canada.* Ottawa: Office of the Auditor General of Canada.

Petrie, M. (1998). *Organisational transformation: the income support experience*. Wellington: Department of Social Welfare.

Pallot, J. (1998). New public management reform in New Zealand: the collective strategy phase. *International Public Management Journal, 1* (1), 1–18.

Parsons, W. (1995). *Public policy: an introduction to the theory and practice of policy analysis.* Cheltenham: Edward Elgar.

Perry, P., & Webster, A. (1999). *New Zealand politics at the turn of the millennium: attitudes and values about politics and government.* Auckland: Alpha Publications.

Pollitt, C. (1995). Justification by works or by faith? *Evaluating the New Public Management, Evaluation, 1* (2), 133–154.

Pollitt, C. (1998a). Managerialism revisited. In B. G. Peters & D. J. Savoie (Eds.), *Taking stock: assessing public sector reforms.* Montreal and Kingston: McGill-Queens University Press.

Pollitt, C. (1998b). How do we know how good public services are? Paper for the CCMD Authors' Roundtable, Ottawa, 12–14 November.

PUMA. OECD. (1999). *Performance contracting.* Paris: OECD.

Ryan, N. (1999). A comparison of contracting arrangements in Australia, Canada and New Zealand. *The International Journal of Public Sector Management, 12* (2), 91–102.

Ryan, N., & Brown, K. (1998). Measuring the performance of community organisations in Queensland and New Zealand. *Australian Journal of Public Administration, 57* (3), 41–51.

Savoie, D. (1998). Fifteen years of reform: what have we learned? In B. G. Peters & D. J. Savoie (Eds.), *Taking stock: assessing public sector reforms.* Montreal and Kingston: McGill-Queens University Press.

Schick, A. (1996). *The spirit of reform.* Wellington: State Services Commission.

Schick, A. (1998). Why most developing countries should not try New Zealand's reforms. *World Bank Research Observer, 13,* 1.

Scott, G., Bushnell, P., & Sallee, N. (1990). Reform of the core public sector: New Zealand experience. *Governance, 3,* 138–167.

Scott, G. (1997). Continuity and change in public management: second generation issues in roles, responsibilities and relationships. In *Future issues in public management.* Wellington: State Services Commission.

Shaw, R. (1999). Rehabilitating the public service–alternatives to the Wellington model. In Chris Eichbaum, et al. (Eds.), *The new politics: a third way for New Zealand.* Palmerston North: Dunmore Press.

Spicer, B., Emanuel, D., & Powell, M. (1995). *Transforming government enterprises: managing radical organizational change in deregulated environments*. Auckland University.

State Services Commission. (1998). *Assessment of the state of the New Zealand public service*. Wellington: State Services Commission.

State Services Commission. (1999). *Briefing for the minister of state services.* Wellington: State Services Commission.

Steane, P. (1999). Public management developments in Australia and New Zealand. *Public Management: An International Journal of Research and Theory, 1* (1), 133–142.

Stent, R. (1998). *Canterbury Health Ltd: a report by the health and disability commissioner.* Auckland: Health and Disability Commissioner.

Taggart, M. (1990). Corporatisation, privatisation and public law. Legal Research Foundation, Publication No. 31.

Task Force on Management Improvement. (1992). *The Australian public service reformed: an evaluation of a decade of management reform.* Canberra: Australian Government Publishing Service.

Thompson, G. (1999). What is wrong with New Zealand's service performance reporting model: the case of public museums. *Public Management: An International Journal of Research and Theory, 1* (4), 511–530.

Treasury New Zealand. (1996). *Briefing to the incoming government.* Wellington: The Treasury.

Turner, D., & Washington, S. (1999). Evaluation in the New Zealand public management system. Unpublished paper.

Upton, S. (1999). The role of the state. *IPS Policy Newsletter, 56,* 8–15.

Wistrich, E. (1992). Restructuring government New Zealand style. *Public Administration, 70,* 119–135.