

## **Towards a Framework for Examining Policy Success**

David Marsh and Allan McConnell

Refereed paper submitted to the  
Australasian Political Studies Association Conference  
University of Queensland  
6 – 9 July 2008

*Governance and Public Policy Stream*

Professor David Marsh, Research School of Social Sciences, Australian National University, [david.marsh@anu.edu.au](mailto:david.marsh@anu.edu.au)

Associate Professor Allan McConnell, Department of Government and International Relations, School of Social and Political Sciences, University of Sydney, [a.mcconnell@usyd.edu.au](mailto:a.mcconnell@usyd.edu.au)

## **ABSTRACT**

Claims that a particular policy has been a ‘success’ are commonplace in political life. They emanate from government and government agencies, interest groups, political parties, media, think tanks, non-governmental organisations etc. However, a few of these claims are justified in any systematic way. This article seeks to remedy this omission by offering a heuristic which practitioners and academics can utilise to approach the question of whether a policy is, or was, successful. It builds initially on two sets of literature: work on public sector improvement; and the work of Mark Bovens, Paul ‘t Hart and collaborators on success, failure and policy evaluation. We begin with a discussion of the epistemological issues involved raising the question of whether it is possible to produce an objective measure of ‘success’. Subsequently, we present a framework for assessing success, focusing on three dimensions: process success; programmatic success; and political success. Each of these dimensions is discussed, before we move on to raise a series of what we term complexity issues, that is issues which make any judgment in relation to any of the dimensions of success difficult: the question of success for whom; the variations across time, space and culture in assessing success; and the methodological issues involved.

## INTRODUCTION

Claims that policies are successful, or not, are implicit, if not explicit, in most discussions of policy. As just one example, Boyne (2003, 211) argues:

In the last decade there has been a pandemic of public sector reforms... Governments across the globe have reorganised and restructured public organisations in an effort to produce better services. The current Labour Government in the UK is engaged, as are others, in this quest for public service improvement. Indeed, the Blair government has stated that delivery on its pledge to raise service standards is the single most important criterion for judging the success or failure of its second term in office (for similar observations see Jas and Skelcher 2005, 195 and Hodgson, Farrell and Connolly 2007, 356).

There were 438,000 hits when we inserted the phrase “policy success” into a Google search (accessed 19/4/2008) of five main types: media pieces assessing the success/consequences of policy<sup>1</sup>; claims by government and government agencies of policy successes, either in the media or in official documents<sup>2</sup>; interest group or voluntary organisation’s assessments/claims about policy successes<sup>3</sup>; blogs on policy outcomes<sup>4</sup>; and academic articles assessing policy success, usually in the form of evaluation studies. As such, both political actors, whether politician, bureaucrats or interest group leaders, and observers, whether academics, journalists or bloggers, assert, even if they don’t demonstrate, the ‘success’ of policy initiatives.

However, the key problem is that these claims/assessments about policy outcomes do not establish any systematic criteria for assessing success or failure. The non-academic literature skates over the problem and even the academic literature, which we consider in more depth below, most of which is concerned with the evaluation and explanation of ‘public service improvement’, generally fails to outline and discuss criteria against which success/improvement could be judged. As Boyne (2003, 212) puts it: ‘Academic contributions to the definition of improvement are also, as yet, extremely limited’. Our aim here is to help remedy this omission in the literature by outlining a framework, a heuristic, within which practitioners and academics can approach the question of whether a particular policy was/is successful.

In pursuing this aim, the article is divided into three sections. In the first section we look at the existing literature on policy success and related topics in more detail, paying particular attention to the key epistemological issue of whether any ‘objective’ assessment of success is possible. The second section then outlines our framework for approaching the issue of assessing policy success. In the final section we further explore some of the complexities involved in assessing policy success and conclude

---

<sup>1</sup> See, for example, ‘Border Policy’s Success Strains Resources’ – [washingtonpost.com /wp-dyn/content/article/2007/02/01/AR2007020102238.html](http://washingtonpost.com/wp-dyn/content/article/2007/02/01/AR2007020102238.html) .

<sup>2</sup> See, for example, ‘Managing Prosperity’ [www.treasury.gov.au/documents/1183/RTf/Managing\\_Prosperty.rtf](http://www.treasury.gov.au/documents/1183/RTf/Managing_Prosperty.rtf)

<sup>3</sup> See, for example, [www.bcchamber.org/Policy\\_and\\_Positions/Policy\\_Success\\_2004-2005.asp](http://www.bcchamber.org/Policy_and_Positions/Policy_Success_2004-2005.asp)

<sup>4</sup> See, for example, ‘Bush Claims Climate Policy Success’ [blog.wired.com/wiredscience/2007/04/bush/claims\\_cli.html](http://blog.wired.com/wiredscience/2007/04/bush/claims_cli.html)

by translating them into critical choices to be made by those seeking to ascertain the 'success' of a particular policy

## **THE EXISTING LITERATURE ON POLICY SUCCESS**

There are some exceptions to our general conclusion that the literature on policy success is thin and here we deal briefly with two: the literature on evaluation and public service improvement; and Boven et.al.'s work, first, on policy failure and, more recently, on policy success.

### **Policy Evaluation: Measuring Public Sector Improvement**

Unsurprisingly, there has been a relative proliferation of evaluation studies in more recent years. Neo-liberal accusations of 'big government' which drains the public purse and fails to deliver effective and accountable public services, has put pressures on public sectors worldwide to demonstrate that they deliver 'public value' (Moore 1995). Equally, strong public sector advocates have been keen to demonstrate, through 'evidence' that the public sector still has a legitimate place in delivering public services. However, an apparent abundance of evaluation studies only take us so far if we want to deal with policy success. As Dye (2005, 332) argues:

Does the government generally know what it is doing? Generally speaking, no...(E)ven if programs and policies are well organized, efficiently operated, adequately financed, and generally supported by major interest groups, we may still want to ask, So what? Do they work? Do these programs have any beneficial effects on society? Are the effects immediate or long range?... Unfortunately, governments have done very little to answer these more basic questions'.

It is perhaps unsurprising that assessments of 'success' are narrowly focused. Much of the evaluation literature is produced from within government but it rarely, if ever, moves beyond the assumption that success equates with meeting policy objectives or producing 'better' policy. Certainly, a great deal of US academic literature exists on the subject of policy evaluation (e.g. Weimer and Vining 1989; Nagel 2002; Davidson 2005,) but it fits into a similar pattern. Most is highly quantitative, as well as highly normative, given its assumption that the purpose of evaluation and policy analysis is 'client-oriented advice relevant to public decisions' (Weimer and Vining 1989, 1). In the UK, where policy studies tilt more towards problem solving or critique than advice, there are far fewer academic studies of evaluation than one might expect. In fact, an extensive search by Hodgson, Farrell and Connolly (2007, 362-363) located only 100 studies, of which many more were concerned with how to introduce a particular improvement measure than with any assessment of its success or failure (see also Boyne 2004). The evidence-based policy movement has been a powerful force within government, strengthening the assumption that evaluation is the study of policy improvement. Indeed, when Michael Barber, the former Head of Tony Blair's Delivery Unit, sought to reflect about 'what success looks like', he concluded that it involved the integration of delivery, spending and efficiency (Barber 2007, 249 and 374).

Others have highlighted the narrow focus of evaluation studies. Head (2008), for example, suggests that the evidence-based policy movement gives scant recognition

to the existence of three very different policy analysis lenses, or form of knowledge, which produce different forms of 'evidence' rather than one: scientific research, practitioner perspectives and political knowledge. The point is reinforced by the argument of Taylor and Balloch (2005, 1) that: 'evaluation itself is socially constructed and political articulated'. Thus, in our attempt to grapple with the nature of policy success, we need to move beyond the perception that the purpose of policy analysis is simply about 'client-oriented advice', but that it can also be about puzzle solving (Winship 2006), democratic dialogue (Ingram and Schenider 2006) and critique (Dryzek 2006).

There are a number of contributions to the evaluation literature which have a more sophisticated discussion of policy success, or, as this literature puts it, public sector improvement. Here, we briefly discuss the work of Boyne (2003), which, in our view, represents the most developed discussion to date of the criteria against which we can assess public sector improvement.

Boyne (2003) is concerned to answer the question: what is public service improvement?<sup>5</sup> His focus is on the improvement in services, so, in the terminology we discuss in the next section, his focus is upon programmatic, rather than process or political success. Boyne identifies five models of organisational improvement in the literature. The goal model is the most straightforward; here the organisation's 'success' is assessed in terms of achieving its goals. As such, in the example Boyne (2003, 214) gives, if the goal of the health service is to raise the standards of citizens' health, then this can be measured in terms of indicators of mortality or morbidity. In contrast, the systems-resource model and the internal processes model use proxies to measure improvement/success. So, in the former, an organisation is successful to the extent that it can attract the resources to survive/grow, while, in the latter, success is measured in terms of the quality/standard of the internal processes of the organisation, both technical (eg. information and budgeting systems) and human (HR practices). Both these models seem to us to be problematic for two reasons. Firstly, the problem of criteria is not really resolved, because the immediate question becomes: how do we measure the quality of information or HR systems? Secondly, and more importantly, these models focus on inputs not outputs and, in turn, these are the same inputs that are usually invoked to explain goal or evaluation success; so the discussion is circular.

The final two models which Boyne identifies, the competing values model (Boyne 2003, 219-220) and the multiple constituency model (Boyne 2003, 220-221), are more interesting, and indeed raise some of the key issues we discuss below. Boyne (2003 p. 219) presents the competing values model as: 'an attempt to synthesize element of the (previous) three models.' Its main contribution is in recognising that ideas about effectiveness are contested and that such contestations, in part, reflect power relations. The multiple constituency model is more radical because it suggests that there can be no agreed criteria for establishing effectiveness/improvement/success. Rather, each

---

<sup>5</sup> He begins by making a distinction between effectiveness and improvement. He argues that effectiveness is a narrower concept, usually applied to the performance of single organisations, while improvement refers to the performance of a series of linked organisations. Similarly, he contends that the effectiveness literature deals with the relative achievements of different organisations in a single time period, whereas improvement is a dynamic concept. Finally, he argues that the effectiveness literature focuses on the effectiveness of organisations, while the improvement literature concentrates on the delivery of services.

group of stakeholders uses different criteria and attaches different weights to them. In addition, the importance of these criteria ebb and flow within the changing power relations between groups. As such, this model has a distinctly post-structuralist tone (see Boyne 2003, 222), which makes it difficult to produce generalisations about policy effectiveness/success.

Boyne claims to combine the goals model and the multiple constituency model, but there are also elements of the competing values model in his formulation. His main contention (Boyne 2003, 222), and we have considerable sympathy with this view, is that the goals model, unlike the systems-resource model and the internal processes model, focuses on outputs, on improvements, rather than the putative causes of improvement. In our view, that is the first question: has there been improvement/is the policy successful? However, we have to acknowledge, as both the competing values model and the multiple constituency model do, that different groups may assess success differently and the matter of which assessment is seen as 'accurate' may reflect power relations in society. Of course, as Boyne hints, ontological and epistemological questions are crucial here, because the difference between the goals model and the competing values model is that the first two are foundationalist/realist in ontological terms and positivist and critical realist respectively in epistemological terms, while the last is anti-foundationalist/constructivist in ontological terms and post-structuralist/interpretivist in epistemological terms (on these debates see Marsh and Furlong, 2002). Boyne advocates, but doesn't discuss, a realist position, because he wants to reject what he calls the: 'postmodernist morass of extreme relativism' (Boyne 2003, 222). As such, he suggests that it is possible to agree a working definition of improvement, albeit one which acknowledges that 'operational definitions of service improvement will be contingent rather than universal' (Boyne 2003, 223), contingent on the ecology of the organisation and the characteristic of its environment.

Overall then, we would suggest that Boyne's work is interesting because it recognises the ontological and epistemological issues which underpin discussions of success and acknowledges that criteria for establishing success are contested and reflect power relations. However, his discussion also has major weaknesses. It focuses fairly narrowly on discussing improvements (or otherwise) in service delivery; so, in Bovens et.al's terminology on programmatic success/improvement. In addition, in so far as Boyne's work broadens the discussion, it does so too cursorily. He raises important ontological and epistemological issues, but does so with a passing reference to realism and doesn't make it clear whether he sees realism as an ontological or an epistemological position. We revisit these issues more systematically below.

### **The Work of Bovens, 't Hart and Collaborators**

Although the work of Boyne (2003), and indeed Taylor and Balloch (2005) and Hodgson, Farrell and Connolly (2007), is interesting, in our view the best work dealing with the issues with which we are concerned is by Bovens, 't Hart and their collaborators. Our main concern here is to look in more detail at their edited book on policy success (Bovens, 't Hart and Peters 2001), but we shall briefly consider their book on policy fiascos/failures (Bovens and 't Hart 1996) and their article on policy evaluation (Bovens, 't Hart and Kuipers, 2006).

Bovens and 't Hart's initial contribution (1996) dealt with the issue of policy fiascos/failure, a fact they and Peters acknowledge (Bovens, 't Hart and Peters, 2001, p. 9): However, they emphasise (Bovens and 't Hart 1996, 157) that policy success is, in many senses, a 'mirror image' of policy failure and turn specifically to that issue in their edited work with Peters (Bovens, 't Hart and Peters 2001).

They argue that there are at least two justifications for an increased focus on success. Firstly, failure is not as ubiquitous as many, particularly the neo-liberals who attacked 'big government', claim. Indeed, one of the conclusions Bovens, 't Hart and Peters (2001, 643) draw from the case studies in their monograph is:

When we turn to examine the outcomes of these cases across the policy sectors, the first thing that is apparent, and rather heartening for students of the public sectors, is that there are perhaps more successes than might have been expected, given the conventional wisdom about the prevalence of failures and fiascos in the public sector.

Second, a focus on failure may have important political effects. Indeed, Bovens, 't Hart and Kuipers (2006, 8) suggest that it may have acted as a crucial stimulus for the growth of neo-liberalism:

Often unintentionally, these studies (of failure) have provided part of the impetus for the ideological swings in thinking about governments and public governance that took place throughout the West during the 1980s and the 1990s.

Similarly, some recent political development literature (Bebbington and McCourt 2007) emphasises the negative consequences for developing countries of focusing on failure, because it gives the impression that the politicians and bureaucrats in all developing countries are incompetent/corrupt and that there are no models of best practice or success.

Bovens, 't Hart and Peters' (2001) aim is to assess policy success/failure in four areas where there have been challenges to governance in six European countries<sup>6</sup>: i) management of decline - the steel industry; ii) management of institutional reform - health services; iii) management of innovation - the financial sector; and iv) management of crisis - HIV Aids. Of course, it is the criteria they establish against which success can be judged that is of relevance here. The key distinction they make is between programmatic and political success. Programmatic success is established in relation to three criteria - effectiveness, efficiency and resilience. Political success is characterised as follows: 'the political dimension of assessment refers to the way in which policies and policy makers become represented and evaluated in the political arena' (Bovens, 't Hart and Peters 2001, 20). More specifically, they assert:

Indicators of political failure or success are political upheaval (press coverage, parliamentary investigations, political fatalities, litigation) or lack of it, and changes in generic patterns of political legitimacy (public

---

<sup>6</sup> UK, Netherlands, France, Germany, Sweden and Spain.

satisfaction with policy or confidence in authorities and public institutions) (Bovens, 't Hart and Peters 2001, 21).

However, they also acknowledge that the detailed measures of success will be policy dependent, and so their collaborators in this edited volume were asked to develop specific criteria for their four cases (the particular programmatic and political assessment criteria are laid out in Bovens, 't Hart and Peters 2001, Table 2.2.,21). They also acknowledge that these criteria might produce contradictory results. So, policies may be unsuccessful in programmatic terms, but 'spun' to appear a political success, or successful in programmatic terms, but represented as a failure and, thus, politically negative for the government (Bovens, 't Hart and Peters 2001, 20).

This seems to us to be an important attempt to establish criteria to judge policy success and, in our view, it is a useful starting point from which to address this crucial issue of policy success. However, there is one significant weakness with the discussion of success by Bovens, 't Hart and Peters (2001) because they have no discussion of the process dimension of policy. Yet, in many cases, a government may regard a policy as successful if it survives the legislative processes virtually unamended from the original proposal. So, a government faced with an urgent political dilemma/issue may propose a solution, possibly transferred from abroad, which moves the issue off the political agenda. In common parlance this may be a 'quick fix', which, in a longer timeframe, results in programmatic failure, although in the short term it may be a political success because the government is viewed as decisive, perhaps in the run-up to an election. As such, below we distinguish between the process, programmatic and political aspects of success.

However, even if we could agree that these three dimensions of success are adequate ones (or even more adequate than previous ones) and then agree measures for how these criteria could be assessed, we are still left with the prior question of the extent to which an objective assessment is possible. So, epistemological issues can't be avoided and this issue is raised much more directly by Bovens, 't Hart and Kuipers (2006) than by Boyne (2003). Bovens, 't Hart and Kuipers (2006, 319) argue that 'policy evaluation is an inherently normative act' (2006, 319) and 'It is only a slight exaggeration to say, paraphrasing Clausewitz, that policy evaluation is nothing but the continuation of politics by other means' (2006, 321). They distinguish between two evaluation approaches: a rationalistic, positivist approach; and an argumentative, interpretive or constructivist approach. In their view (2006, 325), rationalist policy evaluation:

assumes the existence of an exogenously produced, i.e. given, set of clear and consistent policy goals and/or other evaluation standards. It also assumes intersubjective agreement on which indicators can be identified to measure the achievement of these goals.

In contrast, argumentative policy evaluation rejects any separation between fact and value and emphasises that there will always be disputes in evaluation research, for example about the construction of categories.

In their earlier work, Bovens and 't Hart (1996) reject the positivist position, arguing that there are no fixed criteria for assessing success and failure. As such, they adopt a constructivist approach, arguing:

Whenever a policy fiasco is “discovered,” many different kinds of people engage in the meaning making that produces it (Bovens and 't Hart 1996, 10).

A useful starting point (...) is to map out the different layers of meaning that are associated with fiasco as a social construct'. (...) whenever a stake-holder, professional evaluator, or other observer labels a particular government decision, programme, or set of actions as a policy fiasco, she makes a (..) judgement. (...) Each of these acts of judgement is (...) essentially contestable; and each is subject to, or plays a part in, a political process of competing interpretations of ambiguous events. Given the political nature of the process, the substantive issues and claims at stake (...) may at the same time be objects and instruments of contestation. (Bovens and 't Hart 1996, 10-11).

Later, they assert: ‘we are inclined to see merit in alternating epistemologies, with a bias towards interpretivism ‘(Bovens and 't Hart 1996, 151).

However, they qualify their position somewhat arguing:

Does this apparent relativity of failure mean that everything is in the eye of the beholder when it comes to the understanding of policy fiascos? Is there no such thing as an empirical reality of policy failure? We are reluctant to adopt a fully relativist position’ (Bovens and 't Hart 1996, 147).

Their response to this dilemma is to quote Majone (1989, 182) with approval (Bovens and 't Hart 1996, 149; Bovens, 't Hart and Kuipers 2006, 329):

It is not the task of analysts to resolve fundamental disagreements about evaluative criteria and standards of accountability; only the political process can do that. However, analysts can contribute to societal learning by refining the standards of appraisal and by encouraging a more sophisticated understanding of public policies than is possible from a single perspective.

Obviously then, there are difficult questions here. However, we would argue both that it is important to judge the success or failure of policies and that it is possible to establish some criteria to do so. We need to acknowledge the issues raised by Bovens and 't Hart (1996) and Bovens, 't Hart and Kuipers (2006), but not become obsessed by them. In essence, like Bovens and 't Hart and their collaborators, we rest on Majone’s (1989, 183): claim that: ‘The need today is less to develop ‘objective’ measures of outcomes – the traditional aim of evaluation research – than to facilitate a wide-ranging dialogue among advocates of different criteria ’ (cited in Bovens, 't Hart and Kuipers 2006, 329).

However, for such a dialogue to take place, we need a framework which allows us to identify different indicators of success, while acknowledging that the internal and

relational aspects of our criteria may reflect power relations and be perceived differently by different individuals/groups, while some would claim policy success to be nothing more than a social construct reflecting power relations.

## **TOWARDS A FRAMEWORK FOR ASSESSING SUCCESS**

We build on the distinction which Bovens, 't Hart and Peters (2001) make between programmatic and political success, refining both those categories and, as indicated, adding a process category. Table 1 outlines the three dimensions of policy success. We identify the indicators which can be used to measure success in relation to each of the dimensions and, then, the evidence which would be appropriate in relation to each of these indicators. It should be immediately clear that, as we have already said, this is a heuristic, not a model, let alone a theory. Our aim here is to offer a more thorough way to discuss and assess policy success. As such, we briefly discuss the indicators related to the three dimensions identified in Table 1.

**[Table 1 about here]**

### **The Dimensions of Policy Success**

The *process* of policy formation is an important, but often unacknowledged, element in any consideration of whether a policy is successful or not. Broadly defined, 'process' refers to the stages of policy making in which issues emerge and are framed, options are explored, interests are consulted and decisions made. Of course, not all policy making follows such a 'logical' policy cycle, but, rather, may reflect path dependencies, 'garbage can' solutions and discussions within closed policy communities. Nevertheless, processes are important, in both practical and symbolic terms. Certainly, policy which is produced through constitutional and quasi-constitutional procedures will confer a large degree of legitimacy on policy outcomes, even when those policies are contested. Furthermore, as Edelman (1977) emphasised in his classic book *Political Language: Words that Succeed and Policies that Fail*, reassuring words from those in positions of political and bureaucratic authority can facilitate political stability (one measure of policy success), even in the face of iniquitous policies and policy failure (programmatic failure).

It is widely assumed that programmatic 'success' is more likely if the policy process involves, and reflects the interests of, a sufficiently powerful coalition of interests. Of course, determining what is a 'sufficiently powerful coalition' is not an exact science and academics have discussed this issue in terms of 'minimum winning coalitions' (Riker 1962), 'veto players' (Tsebelis 2002) 'advocacy coalitions' (Sabatier and Jenkins-Smith 1993) and 'policy monopolies' (Baumgartner and Jones 1993). However, despite such debates, the simple point remains that, from the perspective of policy makers, putting together a strong alliance supportive of a particular policy can be construed as successful policy. For example, George W. Bush launched Operation Iraqi Freedom with the support of what was then called the 49-member 'coalition of the willing' (subsequently reduced to 25). In essence, a particular configuration of multi-lateral support has continually been framed by the Bush Administration as an indicator of the strength and success of operations in Iraq.

From the perspective of policy makers, a legislative process which subjects a bill to scrutiny, but still leaves the main direction and detailed instruments of the policy intact, is also likely to be construed as a success. Of course, there are instances when the victory of opposition-initiated legislative amendments can help 'improve' the workings of a policy from the government's perspective, or, perhaps more likely, reflect a necessary concession in order to gain opposition support. Nevertheless, such refinements do not take away from the fact that political executives want to see their core proposals become law and the achievement of this end is an indicator of policy 'success'. For example, the Australian Coalition Government headed by Prime Minister John Howard introduced a Bill in the Commonwealth Parliament in November 2005, proposing dramatic and controversial changes to industrial legislation which curtailed collective bargaining and replaced it with a new national system based on individual work place agreements between employer and employee. The Bill was rushed through both chambers in less than six weeks. From the perspective of the Government, this was a 'process success', because, despite minor amendments during the passage of legislation, the mainstay of its economic rationalist/neo-liberal remained intact. Subsequently however, this policy may have lost them the 2007 Election and so been a political failure.

Innovation and influence can also be measures of process success, irrespective of the particular policy outputs and outcomes. A policy process may produce new and innovative ways of tackling a problem. For example, in 2003-4, the Canadian Province of British Columbia chose to reform its electoral system based on recommendations from a Commission of 160 randomly selected citizens. Given the anti-establishment mood in the Province in the wake of corruption scandals, one of the successes of this groundbreaking process was that the ruling Liberal Party could take credit for an innovative process which eschewed vested political interests (Lundberg 2007).

Innovation can also be based on policy transfer from another political jurisdiction (Dolowitz, Greenwold and Marsh 1999; Dolowitz and Marsh 2000). If policy instruments or even broad frameworks are imported from elsewhere, the assumption is that there is some value in doing so. In effect, a policy transferred from elsewhere can bring with it not only particular policy instruments, but also the idea, sometimes erroneous, that it was successful in the original jurisdiction. Although some political actors may dispute this 'stamp of approval', the salient point here is that the policy makers can claim that it was a success elsewhere and will be when transferred. For example, Japan's overhaul of its higher education system was inspired and legitimated by the agencification of British government and reforms in the management of UK universities (Goldfinch 2006).

The types of evidence needed to demonstrate process success detailed in Table 1 are largely straightforward, ranging from quantitative data on legislative voting records, through to qualitative information, such as political speeches and practitioner reports. Of course, as we emphasise below, such 'evidence' is subject to interpretation and contestation, so we are not claiming that process success can be determined irrefutably. Rather, we would argue that, using a heuristic of this type, we can examine these issues more systematically.

*Programmatic* success is often seen as synonymous with policy success and the contemporary focus among most western democracies on evidence-based policy making, or, as Tony Blair's put it, 'what matters is what works', is further confirmation that the assessment of success is outcomes-based and judged by 'the evidence' (Parsons 2002; Sanderson 2002). As such, politicians adduce 'evidence' to demonstrate policy success; a typical example being Tony Blair's assertion in a speech in 2004 on public services: '...services are improving. The best-ever school results, waiting lists falling, crime levels down, council performance improving. The facts ought to speak for themselves' (*Guardian* 29 January 2004).

Operational success occurs if a policy is implemented according to objectives laid down when it was approved. Of course, this is the classic continental European bureaucratic and legalistic measure of success, but it is also evident in much of US policy analysis, as well as in the practice of public bodies throughout the Western world. Policy implementation in the new millennium is generally a much more complex affair than it was, especially given the growth of multi-level governance, public sector fragmentation through arms-length agencies, quangos, non-departmental public bodies, privatisation and outsourcing (Exworthy and Powell, 2004). The post-new public management turn to 'joined up government' and 'whole of government' approaches (Ling 2002) is a modern variation on old assumptions that policy and programme success require effective implementation. As the UK's Public Accounts Committee stated in a report into joined-up working: 'Joint working should help to deliver better public services and address intractable issues which under traditional ways of working have proved difficult to resolve.' (HC 471, 2001-2, para. 23). In other words, successful programme implementation can be considered policy success, even (and perhaps especially) in complex systems of governance.

The nature of programmatic success can also encapsulate the subsequent impact on society. Governments often build outcomes into both policy design and the political rhetoric accompanying policy proposals. The US 'No Child Left Behind' educational reforms enacted by legislation in 2001 are a case in point, as was the statement made by Australian Prime Minister Bob Hawke in 1987 that no Australian child would be living in poverty by 1990. Such aspirations may be difficult, if not impossible, to achieve, but they provide indicators against which programmatic success can be assessed.

The additional aspect of resource success is one that has come to prominence worldwide in developed countries since the mid-1970s world recession. Specifically, efficient use of resources can also be considered a success, and the growth of public audit bodies, value for money studies, national competition councils and productivity commissions are indications of the salience of resource efficiency as a measure of contemporary public policies. Claims can be made of policy success on the grounds that the policy led to 'waste' reduction, cost cutting or more efficient use of available resources. Targets for success can also be set. For example, the UK Home Office (2004, 4) released a report on efficiency and productivity for the police service, stating explicitly in response to 'what success looks like', that it includes 'Authorities/forces achiev(ing)...cashable efficiency & productivity gains worth at least 9.3% over three years<sup>3</sup>, net of costs – sustainably deployed to where they are most needed. This is the basis on which taxpayers have provided the Service with funding.'

A policy may also be successful in a programmatic sense, if it benefits a particular actor, target group or interest, based on issues such as territory, race, religion and gender. Recent examples include: the UK's 2004 civil partnerships legislation entitling gay and lesbian couples to legal recognition of their relationship; the 1995 decision to hold a referendum in Quebec allowing Quebecers to vote on secession; and the UK's Bellwin Formula which compensates local authorities hit by flooding. Framing success in such ways does not mean that success is irrefutable, far from it. In Australia, for example, land rights legislation in 1976 resulted in over 40% of land being handed back to traditional owners and a 'permit' system for entry. Such moves were widely heralded as a success in advancing the interests of Aboriginal peoples who were severely disadvantaged on all key social indicators such as life expectancy and education. However, in 2007, the Commonwealth Government controversially removed substantial elements of the permit system as part of a wider package of measures aimed at tackling indigenous child abuse. Critics accused the government of 'land grab' while the government argued that its measures were necessary to successfully advance the interests of indigenous communities. The fact that the introduction and subsequent removal of land rights can *both* be heralded as a success, indicates that framing and counter framing 'success' lies at the heart of public policy (Stone 2002).

More generally, the evidence needed to judge programmatic success can be found in sources as diverse as party political speeches, legislative committee reports, briefings from stakeholders, and media commentary. This is not to suggest that we can resolve the perennial question of 'real interests' (Hay 2002; Lukes 2005; Marsh, 2008) and say definitively that a particular policy promotes the interests of a particular group or section of society. It is simply to map out a range of measures by which policy success may be measured.

*Political* success is the final benchmark for policy success. In particular, from the perspective of government and the governing party, a policy may be successful if it assists their electoral prospects, reputation or overall governance project. A policy may help rescue the party or government from low popularity (such as the Thatcher Government's intervention in the Falkland Isles), or help consolidate a lead in opinion polls. It may also apply to a policy which helps pave the way for a national election campaign (such as Australian Prime Minister John Howard's 'Pacific Solution' for asylum seekers in the run up to the 2001 Commonwealth election). One does not need to subscribe wholesale to the theory of 'political business cycles' to be aware of instances and accusations of 'election giveaways', 'electoral bribes' 'election budgets', 'vote buying', 'populist policymaking', 'poll driven government' and more.

Proving such connections between electoral success and policy initiatives is the subject of considerable debate within political science circles. Our point, however, is a much simpler one. Policy can be framed as successful because of an assumption that it has positive political impact. Therefore, opinion polls, election results and media commentary, can all be marshalled as evidence that a particular policy initiative was successful in boosting the fortunes of government, governing parties and their programmes.

## **COMPLEXITY ISSUES**

As indicated, the broad criteria we have outlined constitute a framework which academics and practitioners can utilise in order to contend with the matter of whether a policy is/was successful, rather than strictly positivist criteria for success. Our reluctance to embrace objective criteria can be explored further through a number of complexity factors.<sup>7</sup>

### **Whose Success?**

We have already dealt with the epistemological issues, but there is a related and very important issue here concerning ‘success for whom’. The nature of politics, especially in liberal democracies, means that ‘success’ will always be contested to some degree. So, we should not expect government, politicians, civil servants, interest groups, citizens and others, to all agree on whether or not any aspect of a particular policy is successful. For example, under Bill Clinton, the United States began a rendition programme which involved sending suspected terrorists to be interrogated in other countries. Groups such as Liberty have described this practice as the immoral ‘contracting out of torture’. However, supporters have no such qualms, even when they have procedural and legal misgivings. Indeed, one of the architects of this programme, senior CIA official Michael Scheuer, stated in response to a Congressional investigation amid concerns raised regarding legality and ethics:

(t)his is a matter of no concern as the Rendition Program’s goal was to protect America, and the rendered fighters delivered to Middle Eastern governments are now either dead or in places from which they cannot harm America. Mission accomplished, as the saying goes (Committee on Foreign Affairs 2007, 14).

Of course, this raises crucial issues about power relations: do outcomes serve particular interests? Most political scientists would answer in the affirmative, hence acknowledging that success can be relative, depending on which interest we focus on. As one example, Guantanamo Bay prison was viewed as a success (programmatic and political) by the Bush Administration because it incarcerated some of the world’s most dangerous terror suspects and thus, in its view, legitimised its war on terror. However, critics condemned the prison on the grounds that it violated international law (a process failure), abused prisoners, many of whom are innocent (a programmatic failure) and damaged human rights in a country which prides itself on democratic processes (a political failure).

A related point is made by Bovens, ‘t Hart and Kuipers (2006, 322) who argue:

(Even in relatively) uncontroversial instances, policy evaluations are entwined with processes of accountability and lesson drawing that may have winners and losers. However, technocratic and seemingly innocuous, every policy programme has multiple stakeholders who have an interest in the outcome of an evaluation: decision makers, executive agencies, clients, pressure groups.

They also emphasise that interested parties can attempt to influence evaluations (Bovens, ‘t Hart and Kuipers 2006, 324). A case in point is the study by Squires and

---

<sup>7</sup> Some of these factors are adapted from Hogwood and Gunn (1984).

Measor (2005) who demonstrate that in the UK, an evaluation of the Youth Justice Strategy by local evaluators, was severely limited by a national template which only allowed evaluators to assess performance against targets, as opposed to asking whether the interventions actually worked or not. In essence, measures of success were nationally driven by the Home Office and the Youth Justice Board, seeking evidence that a flagship programme was delivering a reduction in youth crime.

There is also the issue of whether we can rely on individuals'/groups' reported view of their interests, which highlights the ontological question of whether individuals have 'real' interests independent of their knowledge of them. Platzer (2005), in her role as an evaluator of mental health programmes, confronted precisely this issue. She found herself stuck between a traditional model of evaluation which assumed that services users were sufficiently empowered to articulate their needs, and her own experiences of vulnerable and disempowered psychiatric patients, who needed (expensive) personalised, support and befriending before they could have any hope of articulating their needs.

Power and its relationship to interests is perhaps the thorniest question in political science, let alone policy studies, and we have no time to address it here. However, while our framework allows us to identify different indicators of success, we must keep at the forefront of any analysis a recognition that policy outcomes reflect power relations and are likely to be perceived differently by different individuals/groups who see themselves as benefiting from/being adversely affected by those policies. At the same time, and returning to the ontological and epistemological issues, we must also recognise that some would claim that policy success is nothing more than a social construct reflecting existing power relations.

### **Time, Space and Culture**

Bovens 't Hart and Peters (2001, 20) argue that: 'Assessments of success or failure in government are therefore dependent on temporal, spatial, cultural and political factors.' Unfortunately, they don't develop this argument, but it is an important one.

There is clearly a temporal dimension involved in any assessment of success. So, any assessment of either the programmatic or the political dimension of success can involve a short or a longer timeframe. As such, a policy that appears successful at one time, with a short timeframe, may seem less successful after a longer period. In the 1960s, the growth of public housing tower blocks (highrise apartments in the US) were widely considered to be a successful response to problems of urban overcrowding, but they have since been widely perceived as epitomising urban squalor and its attendant problems of poverty, crime, unemployment and drug abuse. The Sydney Opera House is an example of the reverse, originally lambasted as an expensive planning disaster (Hall 1982), but now one of the most iconic buildings in the world and a monument for the Australian tourist industry.

In addition, there are likely to be clear patterns in type of the timeframes that different actors choose. So, governments tend to use shorter timeframes which reflect their concerns with immediate political and electoral concerns. As an example, a number of studies of policy transfer have pointed out that governments can transfer policy from another jurisdiction when they have limited information about the effectiveness of the policy in that other jurisdiction, because there has been insufficient time, or political

will to adequately assess if the policy has been 'successful' there. Indeed, Mossberger (2000) in her study of 'enterprise zones', found that the mere idea itself travelled from Britain to the United States, before the appropriate legislation had even been passed in the UK. Similarly, the media and consultants tend to operate with a shorter timeframe. In contrast, academics are more likely, or at least should be more likely, to use a longer timeframe because their concerns are broader and not so driven by the immediacies of politics. In the US, for example, there is a veritable cottage industry of academics who examine retrospectively the 'successes' (and failures) of past presidents and their policies.

There is also a spatial dimension involved in assessing success, particularly if we are assessing comparative policy success across different countries. We define space not in terms of geography, but in terms of different polities, each with its own culture, values, political system and socio-economic conditions. So, what is regarded as success in one political system or political culture will not necessarily be regarded as a success in another. The entire policy making process in the EU is an exemplar. The preponderance of directives (where national autonomy is needed in order to achieve broad EU goals) is indicative of a political system where EU success, necessitates national adaptation (and even opt outs) from common policies. EU social policy in areas such as working time and parental leave is a case in point. As Falkner et al. (2005, p. 1) argue in their comprehensive study of 'soft law' in the EU: 'one-size-fits-all solutions are often neither politically feasible or normatively desirable'.

### **Method**

There are also a variety of methodological problems involved with assessing policy success. In the first instance, we need to acknowledge that success is not all or nothing. Clearly, a policy can be more or less successful. It may achieve some of its objectives and not others. For example, in 2001 when the British Government finally resolved its oil blockade dispute by negotiating with fuel protestors engaged in civil disobedience, it did so in direct opposition to its own earlier portrayal of protestors as 'enemies of the state' who would be dealt with by the forces of law and order rather than dialogue (Robinson 2003). In this case, as in others such as the sanctioning of the private provision of public services or going to war in the knowledge that there will be significant casualties, determining political success depends on how we weight the differing imperatives of dispute resolution and non-negotiation with those prepared to break the law in order to influence the policy process. Similarly, some policy objectives may be partially achieved, creating space for protagonists and commentators to focus on those dimensions which accord with their beliefs and/or purposes. For example, when governments throughout the world attempt to eliminate budget deficits, but only partially meet their targets, they typically attract plaudits from their supporters for their 'success' in reducing the deficit, and criticism from opponents who focus on their 'failure' to produce a balanced budget.

At the same time, while it is usual to see policy success and policy failure as two ends of a continuum, Bovens, 't Hart and Peters (2001), 596) raise the possibility that there is a third category, non-failure, which it would be difficult to locate on such a continuum. Their analysis of the restructuring of the steel industries in France, England and, to some extent, Germany, identifies the reforms as successful in so far as they led to rationalisation and increased competitiveness, but they also produced painful job losses and political volatility. They argue, therefore, that: 'It may be more

accurate to label steel industry restructuring as a matter of non-failure than to speak of straightforward success' (Bovens, 't Hart and Peters 2001, 596).

Undermining the 'all vs. nothing' distinction even further is the possibility that success may be greater than planned. The EU's Common Agricultural Policy is a classic, and high profile example. In the mid-1960s it accounted for roughly 13% of the then EEC budget, but such was its success in providing production incentives to farmers, that by the early-1980s it commanded almost 80% of the budget. In the process, it nearly bankrupted the Community and created the infamous waste associated with butter mountains and wine lakes. In effect, an 'excess' of success can produce failure.

Overall, we can say that there may be contradictions between what is achieved in relation to the process, programmatic and political dimensions of success. So, a government faced with an urgent problem may introduce a policy which is successfully passed with minor amendment, because there is strong party discipline and the government has a workable parliamentary majority; as such, it is a process success. It may also be politically, and electorally, successful, in the short-term at least, because the government is seen as strong, taking decisive action. However, in the medium to long term it may be a programmatic failure, because it fails to achieve its objectives.

One oft-cited example here is the UK 1991 Dangerous Dogs Act (Lodge and Hood 2002). Deaths from dog attacks were, and are, rare in the UK (just over one year). However, in the Spring and Summer of 1991 there were a number of dog attacks which resulted in severe injuries. These received massive attention in the press in what Lodge and Hood term a 'feeding frenzy', with the media demanding immediate government action to deal with 'devil dogs,' particularly pit bull terriers. At the time, the Conservative Home Secretary, Kenneth Baker, was already under attack as a result of a major prison riot and needed to restore his political credibility. The 1991 Dangerous Dogs Act was quickly drafted and welcomed by the tabloid press. It made it illegal to have a dog which was dangerously out of control in a public place and introduced much tougher controls on pit bulls and a few other 'luxury' fighting breeds. The Bill was supported by all the major political parties and passed through all its stages in the House of Commons in one sitting. As such, it was a success in process terms and, initially at least initially, a success in political terms, as it was welcomed by the media and the public; the Government was seen to be acting to deal with an 'important' problem.

However, the law soon became seen as, in the words of the tabloids, 'barking mad' (Lodge 2001). It was patchily enforced and also attracted criticism from magistrates and judges. The Act was amended in 1997 to remove the obligation on courts to sentence to death the dogs of owners who failed to comply with the law. Certainly, this does not seem to be a story of programmatic success, but, in light of the earlier process success (swift and decisive action in the face of public safety concerns and media frenzy), it typifies the capacity of policy to produce multiple and conflicting outcomes. Of course, this poses methodological difficulties in weighting success and failures.

A different aspect of this dilemma occurs when a policy is a success in programmatic terms but not in political terms because of the way the policy is portrayed in the media or by its opponents. So, to again take a UK example, official figures for crime in the British Crime Survey have been falling, but surveys suggest that people think that the situation is worse. Indeed, figures from *Social Trends* (Office for National Statistics 2008, 127) reveal that 65% of adults in England and Wales feel that there is more crime than there was in the country. Of course, this disparity between successful programmatic outcomes and the political failures arising from people's perceptions may be explained by the fact that alcohol-fuelled violent crime is rising (Barber 2007; Office for National Statistics 2008, 125-126) and that confidence in the criminal justice system is low, with 63% feeling that the system was 'not very' or 'not at all' effective in reducing crime (Office for National Statistics 2008, 133-134). However, our main point here is that there are major methodological difficulties in assessing the relative importance of programmatic 'success' and 'political' failure.

Policies also have unintended consequences and it is important to assess their impact and importance and how far (if at all) they undermine the original policy objectives. For example, Cornelius (2001) examines the efficacy and unintended consequences of US Government immigration border control policy since 1993. He argues:

The available data suggest that the current strategy of border enforcement has resulted in rechannelling flows of unauthorized migrants to more hazardous areas (resulting in more death), raising fees by people smugglers, and discouraging migrants already in the USF from returning to their places of origin. However, there is no evidence that the strategy is deterring or preventing significant numbers of illegal entries... (Cornelius 2001, 661).

Such unintended consequences may have a negative effect on the political standing of the government (Martin 2007). As just one example, the atrocities in Abu Grahib prison, which resulted from the decision to go to war in Iraq, had a negative effect on US public support for the Bush Administration. However, unintended consequences may also produce success in at least three ways. First, unexpected negative consequences may spur positive change. As an example, Ontario's light touch water regulatory regime led to the Walkerton contamination disaster, but the government's reaction was a quite dramatic, and popular, upgrading of regulatory processes (see Schwartz and McConnell 2008). Second, success may simply be an unexpected spin-off from a 'successful' program (such as government-funded scientific research in one area producing breakthroughs for another). Third, as a variation on the last point, a government, may hope for, and even anticipate, that an unintended consequence will be positive. One example here would be the classic neo-functionalist view of European integration which anticipated that successful cooperation in the coal and steel sectors would 'spillover' and prompt successful cooperation in related sectors, such as transport.

**[Table 2 about here]**

Identifying and assessing the importance of multiple objectives and complex mixes of 'whole' and 'partial' success is complicated even further because ascertaining success depends on the availability of relevant information. For example, a national urban

policy initiative in France was aimed at the 44 most deprived zones in the country, but a requirement that 20% of new jobs in such areas be given to local residents could not be evaluated because the program did not build in any administrative requirement that this information be gathered (Green, Trache and Blanchard 2001).

The situation is complicated even further when we accept that, not only are many policy objectives unstated, but also, from the perspectives of elected policy makers, a policy fulfils a 'political' function (such as seeking to secure votes, boost leadership or appease disgruntled stakeholders, see Prasser 2006) which they cannot publically articulate for fear of being accused of not acting in the public interest. As a consequence, it a highly subjective task to match the unstated policy goals (or political goals) against the political spin-offs of any particular policy.

The problems of ascertaining policy success are further compounded because, ideally, we would need to isolate and ascertain the effect of the policy on the outcome, controlling for other potential causal factors such as media coverage, the broader economic climate, external shocks, interest group activity, the role of private sector pressure, particularly if public/private partnerships are involved, the actions of other jurisdictions whether national or international and even other linked policy sectors. For example, most observers would accept that government AIDS/HIV campaigns in the 1980s (such as the Grim Reaper ads in Australia) led to transformations in sexual behaviour and a reduction in infections. However, we also need to acknowledge that it is difficult to separate the effect of these camapaigns on people's educational awareness, from other influences, such as learning/pressure from peers, newspapers, TV and films etc. Of course, in addition, these latter influences might themselves be influenced to some degree by government campaigns in an iterative process, and this, again, illustrates the immense methodological difficult of attributing positive outcomes to a 'policy effect'. Indeed, complexity abounds because we would also need to make a judgement about whether, or not, extraneous factors, particularly shocks like 9/11, were knowable and predictable. A policy knocked off course by a completely unpredictable shock could be viewed as a success, because policy makers 'did everything right', or labelled as a failure, because programme outcomes were not achieved.

## CONCLUSION

Political, practitioner and academic discourse on policy success has not yet been matched by a framework which allows us to explore the nature of 'success' in a comprehensive way. There have been some attempts to do so, which we have built upon here. Boyne's work is interesting, but, because of his focus on *performance improvement*, he excludes other forms of success. The work of Bovens, 't Hart and collaborators broadens the subject of study to policy and governance and distinguishes between programmatic and political success. We have considerable sympathy with this approach, but think it is important to add a process dimension.

Our contribution here is to blend programme, political and process in a way which allows analysts and practitioners to approach the issue of success within a useful heuristic framework, as well as helping develop a dialogue about this, remarkably underexplored, aspect of public policy. However, we also recognise that, whatever

dimensions of policy are being considered, there are significant complexities involved in assessing success (see Table 2). What constitutes success can differ according to the perspective and/or interests of the participant in, or observer of, the policy process. At the same time, there are significant methodological difficulties posed by lack of information and the problem of attempting to identify the causal effect of a policy, compared to other independent variables, such as overlapping policies, media influences, economic forces etc. When we study policy success, we are immediately faced with classic political science issues of power, and contestation over interpretation of context and outcome. Nevertheless, in our view, the framework outlined in this article provides a useful heuristic, which allows us to think more systematically about a key, indeed, in some ways, the key, issue in public policy.

## BIBLIOGRAPHY

Barber, M. 2007. *Instruction to Deliver: Tony Blair, Public Services and the Challenge of Achieving Targets*. London: Politico's.

Baumgartner, F.R. and B.D. Jones. 1993. *Agendas and Instability in American Politics*. Chicago: University of Chicago Press..

Bebbington, A. and W. McCourt. eds. 2007. *Development Success: Statecraft in the South*. Basingstoke: Palgrave Macmillan.

Bovens, M. and P. 't Hart 1996. *Understanding Policy Fiascos*. New Brunswick, Transaction.

Bovens, M., P. 't Hart and B.G. Peters eds. 2001. *Success and Failure in Public Governance: A Comparative Analysis*. Cheltenham, Edward Elgar.

Bovens, M., P. 't Hart, and S. Kuipers.(2006. 'The Politics of Policy Evaluation' in M. Moran, M. Rein and R.E. Goodin. eds. *The Oxford Handbook of Public Policy*, Oxford: Oxford University Press, pp. 319-335.

Boyne, G.A. 2003. 'What is Public Sector Improvement?', *Public Administration* 81 (2): 211-227.

Boyne, G.A. 2004. 'Explaining Public Service Improvement: Does Management Matter?', *Public Policy and Administration* 19(4): 100-117.

Committee on Foreign Affairs. 2007. *Extraordinary Rendition in U.S. Counterterrorism Policy: The Impact on Transatlantic Relations*. Joint Hearing Before The Subcommittee On International Organizations, Human Rights, and Oversight and the Subcommittee on Europe of the Committee on Foreign Affairs, House Of Representatives, 17 April, Washington, D.C.

Cornelius, W. 2001. 'Death at the Border: Efficacy and Unintended Consequences of US Immigration Control Policy', *Population and Development Review* 27(4): 661-685.

Davidson, J.E. 2005. *Evaluation Methodology Basics: The Nuts and Bolts of Sound Evaluation*. Thousand Oaks, CA: Sage.

Dolowitz D., S. Greenwold and D. Marsh. 1999. 'Policy Transfer: Something Old, Something New, Something Borrowed, but why Red, White and Blue?', *Parliamentary Affairs* 52(4): 719-730.

Dolowitz, D.P. and D. Marsh. 2000. 'Learning from Abroad: The Role of Policy Transfer in Contemporary Policy-Making', *Governance* 13(1): 5-23.

Dryzek, J.S. 2006. 'Policy Analysis as Critique', in M. Moran, M. Rein, M. and R.E. Goodin, eds. *The Oxford Handbook of Public Policy*. Oxford: Oxford University Press, pp. 190-203.

Dye, T.R. 2005. *Understanding Public Policy*, 11<sup>th</sup> Edition, New Jersey: Pearson Prentice Hall.

Edelman, M. 1977. *Political Language: Words that Succeed and Policies that Fail*. New York, Academic Press.

Exworthy, M. and M. Powell, 2004. 'Big Windows and Little Windows: Implementation in the "Congested State"', *Public Administration* 82(2): 263-281.

Falkner, G., O. Teib, M. Hartlapp and S. Leiber. 2005. *Complying With Europe: EU Harmonisation and Soft Law in the Member States*. New York: Cambridge University Press.

Goldfinch, S. 2006. 'Rituals of Reform, Policy Transfer, and the National University Corporation Reforms of Japan', *Governance* 19(4): 585-604.

Green, H., H. Trache and D. Blanchard. 2001. 'An Experiment in French Urban Policy: Evaluation and Reflection on the Implementation of The Zones Franches Urbaines', *Planning Theory & Practice* 2(1): 53-66.

Hall, P.G. 1982. *Great Planning Disasters*. Berkeley: University of California Press.

Hay, C. 2002. *Political Analysis: A Critical Introduction*. Basingstoke: Palgrave Macmillan.

HC471 2001-2. *Better Public Services through Joint Working*, Select Committee on Public Accounts, House of Commons.

Head, B. W. 2008. 'Three Lenses of Evidence-based Policy', *Australian Journal of Public Administration* 67(1): 1-11.

Hodgson, L., C.M. Farrell and M. Connolly. 2007. 'Improving UK Public Services: A Review of the Evidence', *Public Administration*. 85(2), pp. 355–382.

Hogwood, B.W. and L.A. Gunn. 1984. *Policy Analysis for the Real World*. Oxford: Oxford University Press.

Home Office. 2008. *Efficiency & Productivity Strategy for The Police Service: 2008-11*, London: Home Office, URL < <http://police.homeoffice.gov.uk/news-and-publications/publication/human-resources/efficiency-productivity-strategy?view=Binary>> Consulted 15 April 2008.

Ingram, H. and A.L. Schneider. 2006. 'Policy Analysis for Democracy', in M. Moran, M. Rein, M. and R.E. Goodin, eds. *The Oxford Handbook of Public Policy*. Oxford: Oxford University Press, pp. 169-189.

Jas, P. and C. Skelcher. 2005. Performance Decline and Turnaround in Public Organizations: A Theoretical and Empirical Analysis, *British Journal of Management* 16 (3): 195–210.

Ling, T. 2002. 'Delivering Joined-up Government in the UK: Dimensions, Issues and Problems', *Public Administration* 80(4): 615-642.

Lodge, M. 2001. 'Barking Mad? Risk Regulation and the Control of Dangerous Dogs in Germany', *German Politics* 10(3): 65-82.

Lodge, M. and C. Hood. 2002. 'Pavlovian Policy Responses to Media Feeding Frenzies? Dangerous Dogs Regulation in Comparative Perspective', *Journal of Contingencies and Crisis Management* 10(1): 1-13.

Lukes, S. 2005. *Power: A Radical View*. 2<sup>nd</sup> Edition, Basingstoke: Palgrave Macmillan.

Lundberg, T.C. 2007. 'Electoral System Reviews in New Zealand, Britain and Canada: A Critical Comparison', *Government and Opposition* 42(4): 471-490.

Marsh, D. and Furlong, P. 2002. 'A Skin Not a Sweater: Ontology and Epistemology in Political Science' in D. Marsh and G. Stoker eds. *Theory and Methods in Political Science*. 2<sup>nd</sup> Edition. Basingstoke: Palgrave Macmillan, pp. 17-41.

Marsh, D. (2008), 'Keeping Ideas in their Place: In Praise of Thin Constructivism', Copy available from [David.Marsh@anu.edu.au](mailto:David.Marsh@anu.edu.au)

Martin, B. 2007. *Justice Ignited: The Dynamics of Backfire*. Plymouth: Rowman and Littlefield.

Moore, M. H. 1995. *Creating Public Value: Strategic Management in Government*. Cambridge, MA: Harvard University Press.

Mossberger, K. 2000. *The Politics of Ideas and the Spread of Enterprise Zones*. Washington, D.C.: Georgetown University Press.

Nagel, S. 2002. *Public Policy Studies*. Hauppauge NY, Nova.

Office for National Statistics. 2008. *Social Trends 38*, Basingstoke, Palgrave Macmillan, URL: [http://www.statistics.gov.uk/downloads/theme\\_social/Social\\_Trends38/Social\\_Trends\\_38.pdf](http://www.statistics.gov.uk/downloads/theme_social/Social_Trends38/Social_Trends_38.pdf) > Consulted 16 April 2008.

Parsons, W. 2002. 'From Muddling Through To Muddling Up: Evidence Based Policy Making and the Modernization of British Government ', *Public Policy and Administration* 17(3): 43- 60.

Platzer, H. 2005. 'Best Value But Not Best Interests: Can Service Users Instruct Mental Health Advocates?' in D. Taylor and S. Balloch eds. *The Politics of Evaluation: Participation and Policy Implementation*. Bristol: The Policy Press, pp. 87-96.

Prasser, S. 2006. 'Aligning "Good Policy" With "Good Politics"', in H.K. Colebatch, ed. *Beyond the Policy Cycle: The Policy Process in Australia*. Crows Nest, NSW: Allen & Unwin, pp. 266-292.

Riker, W H. 1962. *The Theory of Political Coalitions*. New Haven: Yale University Press.

Robinson, N. 2003. 'Fuel Protests: Governing the Ungovernable', *Parliamentary Affairs* 56(3): 423-440.

Sabatier, P. A. and H. Jenkins-Smith, C. eds. 1993. *Policy Change and Learning: An Advocacy Coalition Approach*. Boulder, Westview Press.

Sanderson, I. 2002. 'Evaluation, Policy Learning and Evidence-Based Policy Making', *Public Administration* 80(1): 1-22.

Schwartz, R. and A. McConnell 2008. 'The Walkerton Water Tragedy and the Versailles Banquet Hall Collapse: Regulatory Failure and Policy Change', in A. Boin, A. McConnell and P. 't Hart eds. *Governing After Crisis: The Politics of Investigation, Accountability and Learning*. Cambridge: Cambridge University Press, pp. 208-231.

Squires, P. and L. Measor. 2005. 'Below Decks on the Youth Justice Flagship: The Politics of Evaluation' in D. Taylor and S. Balloch. Ed. *The Politics of Evaluation: Participation and Policy Implementation*. Bristol: The Policy Press, pp. 21-40.

Stone, D. 2002. *Policy Paradox: The Art of Political Decision Making*. 2<sup>nd</sup> Edition, New York: W.W. Norton and Co.

Tsebelis, G. 2002. *Veto Players: How Political Institutions Work*. Princeton, NJ: Princeton University Press.

Taylor, D. and S. Balloch. 2005. 'The Politics of Evaluation: An Overview' in D. Taylor, D. and S. Balloch. eds. *The Politics of Evaluation: Participation and Policy Implementation*. Bristol: The Policy Press, pp. 1-17.

Weimer, D.L. and A.R. Vining. 1989. *Policy Analysis: Concepts and Practice*. Englewood Cliffs, NJ: Prentice-Hall.

Winship, C. 2006. 'Policy Analysis as Puzzle Solving', in M. Moran, M., Rein and R.E. Goodin, eds. *The Oxford Handbook of Public Policy*. Oxford: Oxford University Press, pp. 109-123.

<b>Table 1: Dimensions Of Policy Success</b>		
<b>Dimensions</b>	<b>Indicators</b>	<b>Evidence</b>
<b>Process</b>	<p>Legitimacy in the formation of choices i.e. produced through due processes of constitutional and quasi-constitutional procedures and values of democracy, deliberation and accountability</p> <p>Passage of legislation - ie. was the legislation passed with no, or few, amendments?</p> <p>Political sustainability - i.e. did the policy have the support of a sufficient coalition</p> <p>Innovation and influence - i.e. was the policy based on new ideas or policy instruments, or did it involve the adoption of policy from elsewhere (policy transfer/diffusion)?</p>	<p>Legislative records, executive minutes, absence of legal challenges, absence of procedural challenges (e.g. Ombudsman), absence of significant criticism from stakeholders</p> <p>Analysis of legislative process, using legislative records, including identification of amendments and analysis of legislative voting patterns.</p> <p>Analysis of support from ministers, stakeholders, especially interest groups, media, public opinion.</p> <p>Government statements and reports (e.g. White/Green Papers), academic and practitioner conferences, interest group reports, think tank reports, media news and commentary, identification of similarities between legislation and that in other jurisdictions identification of form and content of cross-jurisdictional meetings/visits by politicians and/or public servants.</p>
<b>Programmatic</b>	<p>Operational - i.e. was it implemented as per objectives?</p> <p>Outcome - i.e. did it achieve the intended outcomes?</p> <p>Resource – ie. was it an efficient use of resources?</p> <p>Actor/interest – ie. did the policy/implementation benefit a particular class, interest group, alliance, political party, gender, race, religion, territorial community, institution, ideology etc?</p>	<p>Internal programme/policy evaluation, external evaluation (e.g. legislative committee reports, audit reports), review by stakeholders, absence of critical reports in media (including professional journals).</p> <p>Internal programme/policy evaluation, external evaluation (e.g. legislative committee reports, audit reports), review by stakeholders, absence of critical reports in media (including professional journals).</p> <p>Internal efficiency evaluations, external audit reports/assessments, absence of critical media reports.</p> <p>Party political speeches and press releases, legislative debates, legislative committee reports, ministerial briefings, interest group and other stakeholders’ speeches/press releases/reports, think tank reports, media commentary</p>
<b>Political</b>	<p>Government Popularity – e.g. is the policy politically popular? Did it help government’s re-election/election chances? Did it help secure or boost its credibility?</p>	<p>Opinion polls, both in relation to particular policy and government popularity, election results, media commentary.</p>

<b>Table 2: Critical Choices to Be Made in Assessing Policy Success</b>	
1. Form of Policy Success	Which form or forms of success is/are being assessed? Process? Programmatic? Political?
2. Time Frame	What Time Period(s) is/are being assessed? Short-term? Medium Term? Long Term?
3. Interests	In relation to whose interests is success being assessed e.g. target group? stakeholders? institution? interest group? individual? collective?
4. Reference Points	What is the standard by which success is to be judged? Compared to intentions? Compared to policy domain criteria e.g. efficiency and effectiveness? Compared to the past? Compared to ethical or moral principles? Compared to another jurisdiction?
5. Information	Is there sufficient and credible information to assess the extent of success?
6. Policy Isolation	With what degree of certainty and credibility it is possible to isolate and assess the impact of a policy from other factors such as other policies or media influences?
7. Conflict and Ambiguity	<p>What significance should be given to conflicts and ambiguities, and how should they be weighted in the overall judgement of success? For example:</p> <ul style="list-style-type: none"> <li>- process vs. programmatic vs. political</li> <li>- short-term vs. long term</li> <li>- interests benefiting vs. interests losing</li> <li>- one reference point vs. another e.g. moral principles and stated intentions</li> <li>- availability of information vs. lack of information</li> <li>- certainly in isolating the 'policy effect' vs. uncertainty in being able to do so</li> <li>- one formal objective vs. another formal objective</li> <li>- one informal objective vs. informal objective</li> <li>- one informal objective vs. another informal objective</li> <li>- unintended consequences vs. actual or intended consequences</li> <li>- foreseeable shocks vs. unforeseeable shocks</li> </ul>