

29

THE PROGRAM EVALUATION FUNCTION

Uncertain governance and effects

Robert P. Shepherd

True genius resides in the capacity for evaluation of uncertain, hazardous, and conflicting information.

Winston Churchill¹

Introduction

To suggest that western bureaucracies are experiencing tectonic shifts in the way they operate would be trite. As this handbook demonstrates, continuous change in epistemological terms is occurring on several fronts with respect to the expectations of public sector performance, responsibilities and characteristics of leadership, ethical management, and what constitutes appropriate and effective oversight of policies, programs, systems and processes. Most importantly, however, questions abound regarding the role of government in light of such changes. The entire ontological framework of the Westminster system is being questioned even by those working within it, as significant proportions of public servants have “mentally opted out” (Hubbard and Paquet 2015: 3). Indeed, citizens are losing confidence in their governments to resolve the complex problems they face, and have also mentally checked out of exercising their franchise to vote, or participate in the governing process.

The Westminster system of bureaucracy itself has come under much attack for failing to keep its promises of flexibility, resilience, and innovation in light of persistent global policy challenges, and shifting citizen expectations for participation. As New Zealand’s Fulton Report on civil service reform assessed in the 1960s, “The Home Civil Service today is still fundamentally the product of the nineteenth century philosophy of the Northcote-Trevelyan Report. The tasks it faces are those of the second half of the twentieth century. This is what we have found; it is what we seek to remedy” (New Zealand Committee on the Civil Service 1968: 9). In Canada, civil service reform has been a familiar story with more than 20 reforms since 1867, few of which have led to satisfying change in its ability to wrestle with complex problems, confront its own inefficiencies in management, and engage the potential of its human resources to repair systemic coordinative issues and mediocre performance (Hubbard and Paquet 2015; Savoie 2013; Rhodes, Wanna, and Weller 2009; Dwivedi and Gow 1999).

Situated within this larger reform effort has been the sometimes schizophrenic responsibilities and ineffectual results associated with governmental program evaluation functions. Savoie comes to the damning conclusion in a recent book that “evaluations were very costly and that contributions to the government’s policy making and decision making was negligible” (Savoie 2013: 149). Although such a conclusion could be considered severe, it nonetheless raises important questions about the role program evaluation plays in informing policy and decision-making as these relate to budget-making and controlling expenditures, and whether the expectations placed on that function are actually appropriate and realistic.

The chapter begins with the conventional rationales for program evaluation as a function and the promises it has made, followed by a discussion on whether the function is positioned to speak truth, given recent shifts in the role of evidence in policy making. Subsequently, the chapter explores the complexity of contemporary objects of evaluation (evaluands), and whether evaluation is able through methods to understand that complexity. It then gives a comparative review of where the function is situated typically in the Westminster public management as this constrains or facilitates speaking truth to power (Wildavsky 1987). Finally, the chapter concludes with some insights on the potential of the function to inform better policies and programs. Although this potential varies according to country contexts and conditions, there are some common themes that provide a way forward for this beleaguered function.

Program evaluation traditions: the debate

There are two broad traditions regarding the approach to evaluation: the rationalistic and the argumentative. The rationalistic tradition maintains that neutral advice and objective assessment will insulate the evaluation function from political pressures. Under this perspective, evaluation research is applied social research (Alkin 2004: 127). As an input to decision-making, it consists of the application of various social research methods to provide objective, credible and reliable information to support decisions regarding the design of programs, and assessing the effectiveness and efficiency of those programs. This view is consistent with Rossi and Freeman’s proposition that “evaluation is the systematic collection and analysis of evidence on the outcomes of programs to make judgments about their relevance, performance and alternative ways to deliver them or to achieve the same results” (World Bank 2014; Rossi 2004: 4). This perspective of evaluation is also consistent with several country understandings as stated in centralized policies, including that of Canada’s “Evaluation Policy” (Canada Treasury Board Secretariat 2009). At the root of each of these definitions is the idea that evaluation serves to understand whether policies and programs are actually delivering the effects expected to solve a defined public problem.

Conversely, the argumentative tradition suggests that evaluation is a contributor to public debate, which incorporates competing interests, and explicitly accounts for politics in the ex post assessment of policy performance (Shillabeer, Buss, and Rousseau 2011: 3–16; Mabry 2002). DeLeon (1998) refers to this approach as “consensus through deliberation,” which is rooted in the deliberative democracy literature (Bohman and Rehg 1997).

Contemporary policy and program evaluation approaches emanate from positivism, and reside in the rationalistic tradition. Evaluators contend, not incorrectly, that contributions and judgments about policies and programs must be rooted in empirical evidence (Scriven 1997). For Berk and Rossi (1999: 3), evaluation research is “essentially about providing the most accurate information practically possible in an even-handed manner.” According to Chelimsky (1987), positivism assumes that policies have clearly defined and measurable goals and objectives, expected results, and sound information upon which to draw rational judgments of effect. With the increasing complexity of public problems, such assumptions rarely hold as decision-makers

wrestle with limited resources, capacity, skills shortages, and poor coordination. In addition, governments are also realizing that they are often unable to resolve these problems alone, which means that partners or other contracted agents are being leveraged for their expertise or services. Non-profit groups, for example, are contracted to work with governmental partners, who may have different visions or approaches to resolving particular problems than do governments or other non-profit agents. As the complexity of social issues increases, so too do the arrangements to address them. Rationalistic approaches to understanding effectiveness, for example, are not straightforward, especially as attribution to specific interventions may not always be possible, nor will they always be aligned with policy objectives (Forss, Marra, and Schwartz 2011).

This raises the value of the argumentative tradition, which asserts that positivist approaches are distorted in their view of separating empirical facts from argumentation based on social virtues. Policy examples, such as climate change, despite the appearance of issues with repairs to be found in science, are actually value-laden and are inherently prone to bias given the multiplicity of perspectives, information, and political will that can be brought to bear on them. In this respect, the post-positivist would maintain that climate change is not merely a set of physical objects to be measured, but the interpretation of the scientists that matters (Fischer 1995; Guba and Lincoln 1989). This tradition is rooted in realistic evaluation, whereby facts are considered dependent on underlying assumptions that give meaning to the reality as we collectively understand it. The advantage of this tradition is that it helps decision-makers to understand ethical and epistemological differences in disagreements, rather than prioritizing facts provided by the rational tradition. The promise of the argumentative tradition is that assessments inform consensus building based on beliefs, rather than on detached and contestable prioritization of facts and figures. In this respect, this tradition places evaluation at the front end of the policy analysis, whereas the rationalistic tradition pays attention to the outcomes, or end, of the assessment narrative.

In essence, governmental evaluation functions have tended to rely almost exclusively on rationalistic and (post-)positivist approaches in the contestable belief that analysis is factual and impartial. As the line between policy and intervention becomes increasingly blurred, the greater is the need to find alternative ways of understanding the complexity of public problems (Forss, Marra, and Schwartz 2011). Whether the issue is climate change, health of Indigenous communities, or encouraging technological innovation, more fulsome assessments are required that take into account context and conditions, and the felt reality of all direct participants (Mertens 2012). Alkin (2012) maintains that the field of evaluation is attempting to bridge the two traditions, especially with respect to realistic and transformative approaches (Mertens 2012). The transformative approach, for example, maintains that evaluation must engage and advocate for public discussion, and live the reality of individuals and groups directly affected by public policy decisions, whether implicit or explicit.

Program evaluation and speaking truth

Speaking truth has always been difficult for the evaluation function given that evaluation is not a politically neutral enterprise. For example, when Canada's Auditor General, Sheila Fraser, pronounced on a particular program in 2004 and concluded that "Parliament was not informed of the [program's] true objectives," there was a "weak control environment," and a "lack of transparency in decision-making" (Canada Office of the Auditor General 2004: ch. 3), it caused a firestorm that cast the government under then Prime Minister Paul Martin into an election. That this was an Auditor General giving this report gave it credibility unparalleled anywhere else in government. And, increasingly as public services become ever more politicized, there is

pressure to turn to agents of parliament to conduct such assessments. Even then, the most upright and impartial of actors may find themselves in the middle of controversy (Pawson and Tilley 1997; Stone 1997). In Westminster systems, loyal and impartial service is a cornerstone convention. Telling ministers that their ideas are bad or questionable is not regarded as prudent. To be safe then, the bureaucracy focuses on evaluating the interventions rather than the ideas underlying them, on the premise that governments decide, bureaucracies execute.

One of the several benchmarks of ideal governmental evaluation functions is the extent to which these inform ex post analyses of policies and programs. Although there is an enormous normative literature on ex ante policy analysis conducted by evaluation functions, this will not be addressed in this short chapter as the philosophical debates are voluminous (Alkin 2012; Dunn 2004). Although speaking truth to power is the ideal objective behind any form of policy analysis (Wildavsky 1987), the fact is that like any internal governmental function, evaluation responds to political and bureaucratic imperatives. Ideally, both policy and program evaluation would serve to inform choices around interventions in public problems, provide valuable learning and feedback from past activities, and suggest ways to improve (Funnell and Rogers 2011: 3–13). However, evaluation is subject to many influences that can either enhance its effect, or reduce it to narrow examinations of process and outputs (Shepherd 2012).

The ideal structure of a formal evaluation process is agreed in the evaluation literature: the evaluating organization may initiate investigations independent of intervention with agreed scope and depth of inquiry (i.e., policies, initiatives, programs, outputs, outcomes); it draws on centrally derived or implicit evaluation criteria; it uses an array of tools to assess data; it comes to findings based on evidence, and makes recommendations for the future; and it reports according to accepted standards of practice. There are several recipes prescribed on evaluation systems based on these accepted elements, but the aim is the same: to contribute meaningfully to decisions (McDavid, Huse, and Hawthorn 2013; Funnell and Rogers 2011; Dunn 2004; Weiss 1998; Vedung 1997). The internal evaluation function is highlighted here as the largest contributor to most formal governmental evaluation assessment systems. However, this contribution is indeed shifting, especially at the level of policy analysis, given reduced efficacy of governmental program evaluations (Shepherd 2012).

There is no consensus on how to cope with the political dimension of assessments. This idea is combined with the idea that governments rely on many forms of evidence, and evaluation is not as privileged as it once was (if it ever was). With the increasing influence of think-tanks, lobbyists, and political staff, evaluation that is dependent on rationalistic methods supposedly removed from politics has declined in prominence (Shillabeer, Buss, and Rousseau 2011). This observation is not new, as policy and program evaluation has seen this decline since at least the mid-1980s, leading some to conclude that the field is experiencing a prolonged identity crisis. Given the lack of consensus in the field about its contribution (Savoie 2013: 149; Aucoin 2005), political actors have defined the scope of assessment in the absence of field advice: to examine value for money, rather than support budget making.

Situating the program evaluation function in government operations

Institutions, and the way they operate matters. Where functions such as program evaluation reside in the public administration, these frame to a large extent what responsibilities they will carry out and how they will carry them out. Many Westminster countries have created centralized policies or functions that guide their conduct, whether implicit or explicit. These policies and their forms vary based on country preferences regarding governance and methods, institutions, approaches, and the degree to which they are held in priority or relevance by

decision-makers. The extent to which these policies and forms can be translated into the regular public management have been debated in an attempt to find common insights into how to make these evaluation policies and their resultant frames more effective (Mayne 2006; Pollitt and Bouckaert 2004; Furubo, Rist, and Sandahl 2002).

Centralized (as opposed to dispersed) evaluation functions are created for three, generally but not necessarily, overlapping reasons: to supply political and/or management decision-makers with the information they need to make better budgetary or policy decisions on their programs; to demonstrate fiscal prudence, efficiency and accountability (Good 2008); and to provide decision-makers with information as to whether the right programs are in place to address the problems effectively. In many countries, attempts have been made at different times to find a balance among these. However, as is often the case, one or more of these rationales generally predominates at any given time or is held in higher priority than the others, often leading to some confusion about the ongoing role and value of evaluation in the overall public management (Aucoin 2005). In Canada, Australia and New Zealand, for example, fiscal austerity has tended to drive the nature of evaluation information sought in the last few decades that supports political concerns for financial accountability, and the need to reduce expenditures (Savoie 2013: 150–153).

With respect to physical location of evaluation functions, most Westminster countries have tended to locate these in the executive branch. Some countries have clear lines that distinguish evaluation from audit, while others have tended to regard the functions as similar, or seen the line as unimportant. Always used as a point of comparison, the United States houses its evaluation function mainly within the Government Accountability Office (GAO), which is part of the legislative branch, although there are other institutions that conduct evaluation research. The scope and approaches used by the US function have been debated for decades, especially around the appropriate limits of its reach, the types of studies it can initiate, and the skills it can bring to bear on the function. Unlike many Westminster countries, the GAO has shown a willingness to conduct policy evaluation (Derlien 1999: 148–150). As a part of the legislative branch, the GAO has been asked by legislators to understand the effects of interventions so as to support decision-making on budgets. In an ideal sense, all program evaluation functions aspire to informing sound policy decisions, and the US is considered a role model in this respect.

The location of the program evaluation function in Westminster countries is varied. In the United Kingdom, Canada, Australia, and New Zealand, program evaluation is housed mainly in the executive branch, associated often with Treasury or finance departments, whose mandates relate mainly to expenditures control. The creation of the Program Analysis Review (PAR) in the United Kingdom in 1970, which was coordinated by the Central Policy Review Staff (CPRS) and the Treasury, was an attempt to coordinate program evaluation efforts across all departments with the aim of coming to holistic understandings on the effects of cross-departmental policies, but failed by the mid-1980s as departments exercised ministerial control over studies. It was not until the Thatcher government that evaluation was reactivated with an emphasis on value for money (Derlien 1999). Most recently, efforts aimed at policy evaluation have tended to reside with the National Audit Office, something increasingly common to other external audit functions, with variations on the extent to which such offices can examine programmatic outcomes. Departments continue to produce evaluation studies as a regular part of overall departmental management with some prescriptive guidance from the Comptroller General.

In Australia, formal and centralized evaluation has been evident since the 1950s, mainly with respect to education programs. However, the function did not become a major priority until the Coombs Report, *Report of the Royal Commission on Australian Government Administration*

(Australian RCAGA 1976) and Braume Report, *Through a Glass Darkly* (Australian Senate Standing Committee on Social Welfare 1979). The Coombs Report aimed to introduce organizational diagnosis, and benchmarking (performance measurement) as a way to understand departmental performance. The Braume Report recommended a whole-of-government approach to evaluation for the purpose of “reviewing the efficiency, effectiveness and appropriateness of any program or group of programs” (SSCSW 1979a: 5). Despite efforts at the federal and state levels, however, and Australia’s national government investments in evaluation capacity in the 1970s, it did not become mandatory as a whole-of-government approach to accountability until the 1980s, and even then was evident only in a handful of policy areas (Sharp 2004: 9).

Australia has been shown to be a leader since 1984, incorporating performance measurement into its program evaluations as a tool of management improvement. Even this innovation has proved problematic. According to a recent examination of performance measurement in departments, they “continue to find it challenging to develop and implement performance indicators, in particular effectiveness indicators that provide quantitative and measurable information, allowing for an informed and comprehensive assessment and reporting of achievements against stated objectives” (Australian Audit Office 2011: 12). Performance measurement has proved to be less of a panacea than hoped, and Australia’s experience in this respect is not unlike that of other countries that have moved in this direction, including the UK and Canada.

The Canadian experience is also not unique. Canada’s system, the second oldest evaluation function next to the US, was centralized in 1977 under the direction of the Treasury Board Secretariat, and framed under a new Results Policy, introduced in July 2016. This replaces the previous Evaluation Policy 2009, which maintained a separate evaluation function. Program evaluation is now subsumed under a larger results framework to “improve the achievement of results across government” (s.3.1.1). Evaluations are also used to support resource allocation decisions based on performance (s.3.2.3). The argument is that departments will better understand success by linking programs to centralized policy priorities. This approach appears to build on the Australian and British experience.

The efficacy of evaluation in Westminster countries to inform sound policy, or even program decisions as a benchmark of effectiveness has been less than stellar (Shepherd 2012; Canada Office of the Auditor General 2009; Mulgan 2008; UK Committee of Public Accounts 2001). Resident within the executive branch, internal evaluation functions are concerned with management and accountability, and questions of austerity, limiting the scope of programs, and reducing government. Although this has not always been the case, much of its history as an internal function can be characterized in this way. Indeed, in Canada several decades of Auditor General’s reports on the effectiveness of the evaluation function have cited concerns for lack of effectiveness information on programs. The repair for legislators has been to turn to agents of parliament for greater attention to program effects, thereby blurring the lines between audit and evaluation. Although there remains a role for internal program evaluation, it is highly bounded by political imperatives to tell good news stories, and not to embarrass the government. And, with increasing concern for financial accountability, finding reports that provide rigorous conclusions on the effect of programs has been increasingly challenging.

The way forward: a different model?

Any discussion on the future of internal or accountability-focused evaluation functions is wrought with epistemological differences in perspective, and competing value and worth claims

about the internal evaluation function as a governmental enterprise. This is indeed highly contestable territory, as the entire evaluation function (dependent admittedly on rationalistic approaches) is struggling to rediscover its place in the public sphere. The field continues with great rigour to debate the merits of objectivity in evaluative analysis, validity and reliability of data, flavour-of-the-month analytic and data-gathering approaches, and establishing attribution and contribution of interventions. These debates are attempts at insulating the field and the rationalistic tradition from epistemological relativism, and understanding the increasing primacy of political involvement in the assessment process no doubt related to current movements toward new political governance.

One could argue that such debates are to miss the point, as political and other forces and their desire for realistic analysis are not going away anytime soon, as shown in other contributions to this volume. Such an argument could suggest that the merits of the argumentative approach are worthy of consideration, and one would be advised to look in this direction at least for part of the way forward for improving the relevance of evaluation within the public sector. Given the growing number, complexity and globalization of many public problems today, it makes intuitive sense that rationalistic approaches to policy making are not enough. In fact, it is reasonable to suggest that our knowledge about social outcomes, including public policy making, is increasingly based (if this was not always the case) on social constructions, which assumes limited information and understanding of those constructions. Such constructivist understandings necessarily move assessments into depending on multiple evidentiary sources, including argumentative or deliberative approaches. Of course, this is the ideal state of evaluation research. It assumes that governmental actors would welcome such approaches, and would be willing and committed to policy debates. Unfortunately, as shown in other contributions here, secrecy is actually increasing in many public sectors, and there is a movement toward politicizing public services in such a way as to restrict further any information that would act as an obstacle to re-election or government messaging.

The way forward, therefore, for policy analysts and evaluators is not straightforward. However, as Majone (1989: 182) explains:

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.

Majone goes on to argue that the role of policy analysts and evaluators is to create a dialogue among “advocates of different criteria” (Majone 1989: 183). This perspective aligns well with that of Mertens and Wilson (2012), who suggest that only by engaging in dialogue among competing interests can evaluation be relevant and useful. In effect, they promote the idea of advocacy of policy ideas rather than relying on traditional notions of objective analysis. In their view, this *one can lead a horse to water, but cannot force it to drink* approach to evaluation use is no longer effective, nor even widely assumed to be true any longer (Shillabeer, Buss, and Rousseau 2011).

The problem with rationalistic evaluation is that it is premised on the simple view that governments should be held to account for their promises. We know from decades of evaluation work that this view is rife with faulty assumptions, including that governments even want to be held to account, despite their rhetoric to the contrary (Shillabeer, Buss, and Rousseau 2011). This is not to suggest that the rationalistic approach be discarded, but that it qualify discussions

on effectiveness with organizational learning considerations (Pawson and Tilley 1997). This is not a new idea: in addition to holding governments to account and understanding effect, the evaluator is asked to consider the organization's ability and capacity to adapt to changes in its internal and external environments over time, and to assess its ability to control costs of the intervention being evaluated (Bovens, 't Hart and Kuipers 2006: 330). This learning approach to evaluation would examine all three criteria, and make an assessment about the relative merit of each in program implementation. Bovens et al. (2006: 330) suggest, as one possible avenue forward, combining this rational logic with the argumentative approach that understands how policy and other decision-makers are involved, represented and assessed in the political arena through an understanding of "symbols, emotions, political ideology, and power relationships." For them, it is not the social consequences of policy that matter, "but the political construction of these consequences, which might be driven by institutional logics and political considerations of wholly different kinds."

In Canada, an inquiry was called in 2000 to examine the events surrounding the contamination of water at a plant in Walkerton, a town of approximately 5,000 people in central Ontario. The inquiry looked into an outbreak of *E. coli* that resulted in the deaths of seven people. The inquiry, led by the Honourable Dennis R. O'Connor, did not simply want to understand the series of events that resulted in the deaths, but the systemic policy and implementation problems as well. He examined the role of various governmental institutions, both operational and regulatory, and followed the evidence on the effects of systemic failures in other jurisdictions in the province. He sought out experts, ministry and local operations officials, municipal councils, health authorities, and citizens to figure out not simply who to blame, but how to recognize and address the contributing factors that led to the outbreak. He actually sought out political and bureaucratic perspectives to understand the decisions, contexts and conditions that framed the tragedy. In the end, the report did find culpability, but it also made and insisted on changes to Ontario's safe water legislation (Ontario Walkerton Inquiry 2002). Such assessments take time, but the objective of the study was to identify different perspectives, and generate a common understanding of the problems through dialogue. Unfortunately, this is but one of very few examples of such an approach in the Canadian context.

Concluding thoughts

The field of governmental evaluation has not fully resolved conceptual challenges in the field, thereby clinging to limited insights provided by studies buttressing expenditure control decisions and management decisions around implementation. Little question is given in internal evaluation to the ideas that support programs, leaving some to doubt the relative worth of the function (Savoie 2013; Mulgan 2003: 87–90). As shown, the dependence on rationalistic methods has lost its sheen once enjoyed (maybe) in the 1960s and 1970s. Again, these arguments are not new: legislators are unconcerned by bureaucratic efficiency, and auditors and evaluators are reticent to question government policy.

Another challenge for internal evaluation, and public services more broadly is a reluctance on the part of politicians to trust that bureaucrats will come to reasonable expenditure budgets. In Australia, UK and Canada, evaluations have been tasked alongside audits to address matters of value for money to rectify this problem. Again, this is not a new development as it has been suggested many times in various forms with limited success that bureaucrats will be able to come to reasoned ideas about controlling spending. It is no wonder that politicians believe they have little choice but to implement wide-sweeping operating reviews in order to curb spending. Internal evaluators are reluctant to suggest cutting or trimming programs, even inefficient or

irrelevant ones. It may make better sense for evaluation to support budgetary decisions that recommend calibrating or re-allocating of resources, rather than to review expenses alone. The fact is that even as a management tool, internal evaluation has failed to live up to expectations.

The repair appears to be more reflexive approaches to evaluation. These are sensitive to testing assumptions in policy and program theories of change and action, contexts and conditions of interventions, and depend on open systems of debate. As shown elsewhere in this volume, these are high ideals indeed, but they nonetheless set the stage for useful evaluation that enables public sector reform efforts. Debates abound in many Westminster countries as to evaluation approaches, criteria, scope, and institutional arrangements, and these are important as it defines the realities of many public sector evaluators. Discussions about instituting independent and impartial evaluators general or other similar models, while meritorious, neglect ongoing epistemological deficiencies in the field. Human beings are social creatures, and as such, are rooted in contexts. We want to understand causality as established in these methods, but we also want to know that the findings from research are relevant and useful to informing better policy choices, and their resultant interventions.

The promise of evaluation to understand policy and program effects remains intact especially as more and more jurisdictions regard it as a useful decision tool. The challenge for the field is to re-stake its ground as the voice for holistic assessment of public interventions and the ideas that inform these. This means that it must embrace both rationalistic and argumentative approaches, especially as public decisions and the information that surrounds such decisions is removed increasingly from the view of citizens. Perhaps there is a place for evaluators to become advocates and promoters of change, rather than remain insulated in scientific objectivity.

Note

- 1 Churchill, Winston. BrainyQuote.com. Xplore Inc, 2015. Available at: www.brainyquote.com/quotes/quotes/w/winstonchu144998.html.

References

- Alkin, M. 2004. *Evaluation Roots: Tracing Theorists? Views and Influences*. Thousand Oaks, CA: Sage Publications.
- Alkin, M. 2012. *Evaluation Roots: A Wider Perspective of Theorists? Views and Influence*. 2nd ed. Thousand Oaks, CA: Sage Publications.
- Aucoin, P. 2005. "Decision-Making in Government: The Role of Program Evaluation." Discussion Paper for Treasury Board Secretariat. Ottawa: Supply and Services.
- Australian National Audit Office. 2011. *Development and Implementation of Key Performance Indicators to Support the Outcomes and Program Framework*. Canberra: National Audit Office.
- Australian Senate Standing Committee on Social Welfare. 1979a. *Through a Glass Darkly: Evaluation in Australian Health and Welfare Services*, Volume 1. The Report. Canberra, Australian Government Publishing Service.
- Australian Senate Standing Committee on Social Welfare. 1979b. *Through a Glass Darkly: Evaluation in Australian Health and Welfare Services*, Volume 2. Papers. Canberra, Australian Government Publishing Service.
- Berk, R.A. and Rossi, P.H. 1999. *Thinking about Program Evaluation*. Thousand Oaks, CA: Sage.
- Bohman, J. and R. William, eds. 1997. *Deliberative Democracy: Essays on Reason and Politics*. Cambridge, MA: MIT Press.
- Bovens, M., 't Hart, P., and Kuipers. S. 2006. "The Politics of Policy Evaluation." In *The Oxford Handbook of Public Policy*, edited by Michael, M., Martin, R., and Goodin, R.E, 310–335. London: Oxford.
- Canada Office of the Auditor General. 2009. "Report of the Auditor General of Canada 2009." Chap. 1 in *Evaluating the Effectiveness of Programs*. Ottawa: Supply and Services.

- Canada Office of the Auditor General. 2004. "November Report of the Auditor General of Canada 2004." Chap. 3 in *The Sponsorship Program*. Ottawa: Supply and Services.
- Canada. Treasury Board Secretariat. 2016. "Policy on Results." Ottawa: TBS.
- Chelmsky, E. 1987. "The Politics of Program Evaluation." *Society* 25: 24–32.
- DeLeon, P. 1998. "Introduction: The Evidentiary Base for Policy Analysis: Empiricist Versus Post-Positivist Positions." *Policy Studies Journal* 26: 109–113.
- Derlien, H.U. 1999. "Genesis and Structure of Evaluation Efforts in Comparative Perspective." In *Program Evaluation and the Management of Government: Patterns & Prospects across Eight Nations*, edited by Ray, R. London: Transaction Publishers.
- Dunn, W. 2004. *Public Policy Analysis: An Introduction*. 3rd ed. Upper Saddle River, NJ: Prentice-Hall.
- Dwivedi, O.P. and Gow, J.I. 1999. *From Bureaucracy to Public Management: The Administrative Culture of the Government of Canada*. Toronto: Broadview.
- Fischer, F. 1995. *Evaluating Public Policy*. Chicago: Nelson Hall.
- Forss, K., Marra, M., and Schwartz, R. 2011. *Evaluating the Complex: Attribution, Contribution, and Beyond*. 18 vols. London: Transaction Publishers.
- Funnell, S. and Rogers, P. 2011. *Purposeful Program Theory: Effective Use of Theories of Change and Logic Models*. San Francisco, CA: Jossey-Bass.
- Furubo, J.E., Rist, R., and R. Sandahl, eds. 2002. *International Atlas of Evaluation*. New Brunswick, NJ: Transaction Press.
- Good, D. 2008. *The Politics of Public Money: Spenders, Guardians, Priority Setters, and Financial Watchdogs Inside the Canadian Government*. Toronto: University of Toronto Press.
- Guba, E.G. and Lincoln, V.S. 1989. *Fourth Generation Evaluation*. Newbury Park, CA: Sage.
- Hubbard, R., and Paquet, G. 2015. "The Canadian Federal Public Service: Tinkering Can No Longer Suffice." *Optimum Online* 45 (3).
- Mabry, L. 2002. "Postmodern Evaluation – Or not?" *American Journal of Evaluation* 23 (2): 141–157.
- Majone, G. 1989. *Evidence, Argument and Persuasion in the Policy Process*. New Haven: CT: Yale University Press.
- Mayne, J. 2006. "Audit and Evaluation in Public Management: Challenges, Reforms, and Different Roles." *Canadian Journal of Program Evaluation* 21 (1): 11–45.
- McDavid, J.C., Huse, I. and Hawthorn, L. 2013. *Program Evaluation and Performance Measurement: An Introduction to Practice*. 2nd Edition. Thousand Oaks, CA: Sage.
- Mertens, D. and Wilson, A. 2012. *Program Evaluation Theory and Practice: A Comprehensive Guide*. New York: Guilford Press.
- Mulgan, R. 2008. "The Accountability Priorities of Australian Parliamentarians." *Australian Journal of Public Administration* 67 (4): 457–469.
- Mulgan, R. 2003. *Holding Power to Account*. New York: Palgrave Macmillan. New Zealand. 1968. Committee on the Civil Service (Fulton). *Final Report*, Cmnd. 3638. London: HMSO.
- Ontario Walkerton Inquiry. 2002. *Report of the Walkerton Inquiry: The Events of May 2000 and Related Issues*. Toronto: Ontario Ministry of the Attorney General, Part One.
- Pawson, R. and Tilley, N. 1997. *Realistic Evaluation*. London: Sage.
- Pollitt, C. and Bourcaert, G. 2004. *Public Management Reform: A Comparative Analysis*. 2nd Edition. New York: Oxford.
- Royal Commission on Australian Government Administration (RCAGA). 1977. *Report*. Canberra: Acting Commonwealth Government Printer.
- Rhodes, R.A.W., Wanna, J., and Weller, P. 2009. *Comparing Westminster*. New York: Oxford.
- Rossi, P.H., Lipsey, M., and Freeman, H. 2004. *Evaluation: A Systematic Approach*. 7th ed. Thousand Oaks, CA: Sage Publications.
- Savoie, D. 2013. *Whatever Happened to the Music Teacher?: How Government Decides and Why*. Montreal and Kingston: McGill-Queen's University Press.
- Scriven, M. 1997. "Truth and Objectivity in Evaluation." In *Evaluation for the 21st Century: A Handbook*, edited by Chelmsky, E. and Shadish, W.R., 477–500. Thousand Oaks, CA: Sage.
- Sharp, C. 2004. "Development of Program Evaluation in Australasia and the Australasian Evaluation Society: The Early Decades." *Evaluation Journal of Australasia* 3 (2): 6–16.
- Shepherd, R. 2012. "In Search of a Balanced Evaluation Function: Getting to Relevance." *Canadian Journal of Program Evaluation* 26 (2): 1–45.
- Shillabeer, A., Buss, T., and Rousseau, D. 2011. *Evidence-Based Public Management: Practices, Issues and Prospects*. London: M.E. Sharpe.

The program evaluation function

- Stone, D. 1997. *Policy Paradox: The Art of Political Decision-Making*. New York: Norton.
- United Kingdom Committee of Public Accounts. 2001. "Holding to Account." Report of the Review by Lord Sharman of Redlynch. Office of the Treasury.
- Vedung, E. 1997. *Public Policy and Program Evaluation*. New Brunswick, NJ: Transaction.
- Weiss, C.H. 1998. *Evaluation: Methods for Studying Programs and Policies*. Upper Saddle River, NJ: Prentice-Hall.
- Wildavsky, A. 1987. *Speaking Truth to Power: The Art and Craft of Policy Analysis*. New Brunswick, NJ: Transaction Publishers.
- World Bank. 2014. "What is Monitoring and Evaluation?" Available at: <http://ieg.worldbankgroup.org/what-monitoring-and-evaluation>. Accessed June 6, 2016.