

# Expenditure Reviews and the Federal Experience: Program Evaluation and Its Contribution to Assurance Provision

---

Robert P. Shepherd  
Carleton University

Citizens are always the real heroes of public sector reform, because it is they who must accept the sacrifice, and it is they who pay the price of failure. (Bourgon, 2009, p. 24)

**Abstract:** *Various forms of assurance are being demanded by different constituencies in the federal public administration. One form of assurance is that of financial accountability, and spending reviews are an essential input to processes that contribute to federal budgetary and expenditure management decisions. Program evaluation has also been an important contributor, but it may be the case that this federal function is overextended in that contribution. It may be time to consider removing this responsibility and attaching it to other functions, thereby affording the function to better focus on what it does best: contribute to program improvement, including effectiveness. This would also mean a shift in evaluation culture to one of learning, rather than accountability.*

**Keywords:** *assurance, accountability, decision criteria, evaluation results, federal context, public management, spending reviews*

**Résumé :** *Diverses formes d'imputabilité sont exigées des différents groupes de l'administration publique fédérale, dont la responsabilité financière. Les exercices de révision budgétaire sont une composante essentielle contribuant aux décisions du gouvernement fédéral en matière de gestion budgétaire et de dépenses. L'évaluation de programme a joué un rôle important à ce niveau, mais il se peut que le fédéral ait des attentes trop élevées concernant sa contribution. Il serait peut-être temps d'envisager d'enlever cette responsabilité à l'évaluation et de la rattacher à d'autres fonctions, ce qui lui permettrait de se concentrer sur ce qu'elle fait le mieux : contribuer à l'amélioration et l'efficacité des programmes. Cela signifierait aussi que la culture de l'évaluation évolue vers une culture d'apprentissage plutôt que vers la reddition de comptes.*

**Mots clés :** *assurance, responsabilisation, critères de décision, résultats de l'évaluation, contexte fédéral, gestion publique, examen des dépenses*

**Corresponding author:** Robert P. Shepherd, School of Public Policy and Administration, Carleton University; [RobertPShepherd@Cunet.Carleton.Ca](mailto:RobertPShepherd@Cunet.Carleton.Ca)

## INTRODUCTION

There is little new about expenditure reviews or the role of program evaluations to support them. Both are essential instruments and systems supporting deputy ministers in their efforts as gatekeepers who must ensure financial scrutiny and accountability in their departments/agencies under the Financial Administration Act. Such instruments and systems provide assurance that such scrutiny is adequate and effective by collecting data, monitoring activities, and providing assurance that resources are being used in ways that meet the expectations of Parliament and the executive. The challenge today is that different forms of assurance are being demanded, different systems have been instituted to provide forms of assurance, and there are several constituencies that demand one or more forms of it. There is a multiplicity of policies, systems, and processes in place at the federal level that serve this expanding set of expectations for assurance. For program evaluation, however, its roles and responsibilities in the federal system have changed many times since it centralized in 1977, including supporting expenditure management. The result of both the reforms and the addition of responsibilities is a function increasingly without a clear sense of itself, despite the many iterations of evaluation policies.

To be precise, federal program evaluation has not lived up to the high expectations placed on it, and there are many reasons for this, not the least of which is that it serves too many purposes, including supporting expenditure reviews, and it has many constituencies (Borys, Gauthier, Kishchuk, & Roy, 2005). Its target of small-*p* programs also limits its ability to contribute to government-wide decision-making, such as budgeting and expenditure management. That is, the function was given the difficult task of providing assurance that programs are effective, efficient, and delivered in a cost-efficient manner. However, there is an expectation among decision makers that the function can do more than enlighten programmatic decisions. The function is expected to make an evidence-based link between the effectiveness of program activities and budget efficiency (Aucoin, 2005, p. 2). In effect, the targets of study from a program focus to budget focus do not line up. Such confusion of responsibilities has led commentators such as Savoie (2013) and others to come to the conclusion that the program evaluation function and its products have had no noticeable effect on government-wide budgetary decision-making, including supporting decisions regarding the expenditure budget. He argues that “little has changed over the past twenty years, apart from the fact that more money and staff are now being dedicated to the evaluation industry” (Savoie, 2013, p. 149).

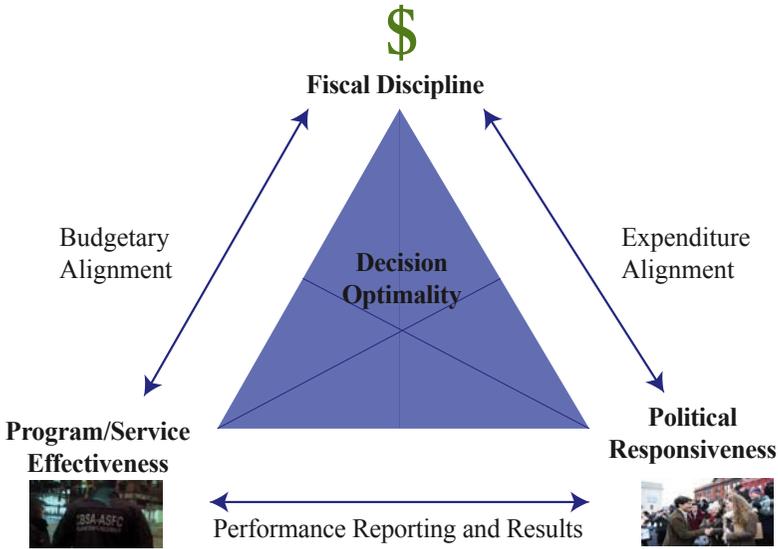
Others in this volume comment on the limited role that program evaluation has played in expenditure reviews (Mayne, 2018; Dobell & Zussman, 2018; Robinson, 2018). However, the particular reasons for this conclusion should not go unstated. This paper argues that program evaluation has a role to play in federal policy and management decision-making, and that it is time for the function

to be reformed and expectations from various constituencies to be placed into proper alignment. Given the various federal reforms underway in financial management, this article focuses on the role of program evaluation within the larger public administration to support senior decision-makers in their efforts for better expenditure management.

It addresses the aims of the article in three parts. First, it explores decision-making at the federal level, and where program evaluation and expenditure reviews fit in that context. Second, it then highlights in historical context the forms of assurance to which evaluation contributes, without overly replicating the narrative provided by Dobell and Zussman. Third, it draws on 12 interviews with senior managers (mainly assistant deputy ministers) and heads of evaluation over the months of May and June 2017 regarding possible areas of weakness in the function and how it can be better aligned with expenditure reviews in this era of deliverology. Respondents were asked to consider the role of evaluation in assisting with budget making, human resources allocations decisions, programmatic improvements, and supporting the Treasury Board Secretariat (TBS) expenditure review criteria. The essential purpose of the interviews was to determine the strengths and weaknesses of evaluation to support senior management decision-making and where they thought the function was performing well. Respondents were also asked the extent to which the function might be stretched to the point where it is not or cannot serve a useful purpose. The perspectives of both senior managers and heads of evaluation were remarkably homogenous, many of which are reflected in various parts of the article.

## FEDERAL DECISION-MAKING CRITERIA

Governments have long struggled to find innovative ways to improve certainty or the probability that decisions on governmental programs and expenditures are optimal. Although decision optimality comes in many different forms, and the criteria used to assess it are multiple, every government wants to see that public resources are being assigned in a way that maximize effect so that a reasoned defence can be made of their record in office. Most often, governments will make decisions that optimize fiscal discipline or rigour, program/service effectiveness, or political responsiveness to citizen demands (Aucoin, 2005). In particular, much has been written on the value and uses of evidence in decision-making regarding governmental policy priorities, how public resources are allocated and used, and shifts in the role decision makers play in using evidence (S. P. Young, 2013). One way that the federal government has ensured that expenditure decisions align with performance is expenditure reviews, or what the Canadian federal government refers to most recently as resource allocation reviews (RARs). Expenditure reviews can be one-time or ongoing. The current *Policy on Results* (Canada, TBS, 2016b) and Management Accountability Framework (Canada, TBS, 2017), among other



**Figure 1.** Federal Decision-Making Priorities

policies, formalizes the ongoing review of all direct program and departmental spending every 4 years.

Governmental decision-making regarding the determination of priorities and the allocation of resources to achieve them is a complex process in which evidence, collected from structured or systematic data collection and dynamic learning or experience, combines with prevailing ideas at the time, interests, epistemologies, institutions, and thought leaders. Each of these factors operates at both the political and administrative levels, whereby some forms of evidence may take precedence at different times over others (Buss & Shillabeer, 2011). This does not imply that one form of evidence is necessarily more rational than the other, but that decisions are driven by circumstances in response to planned and emergent expectations by governments and citizens (Mintzberg & Waters, 1985).

Figure 1 shows the interplay of the different decision criteria and where concerns of budget and resource allocation, expenditure management, and reporting and results tend to interact with such criteria. Fiscal discipline refers to managing both the allocation of scarce resources according to needs and demands, and expenditure control within the public services. Both forms of discipline rely on effective systems of stewardship. For example, budget allocation requires systems that consult various interests, determine revenue levels, make determinations on allocations, and adjust budgets where appropriate. Likewise, expenditure management involves internal expenditure management and monitoring systems, and data gathering and accountability systems such as internal audit and evaluation.

Program and service effectiveness, as a criterion, concerns the extent to which expected programmatic effects are being observed and realized. Governments

want sufficient assurance that what they are doing is leading to expected results. Sometimes such decisions take priority, as in the case of Indigenous youth suicide and general programs aimed at Indigenous reconciliation. Spending decisions will be guided by the extent to which programs are observed to be yielding results that align with internal and external expectations.

Political responsiveness concerns both the reality and perception that government is responding to internal and external demands for action, or that it is being seen to represent various interests, ideas, institutions, and individuals (Howlett, 2010). Of the three decision-making priorities, this is the most dynamic, and it is a necessary part of governmental action. It moves actors and institutions to act, for better or worse. As Marland observes in his recent book, *Brand Command*, sometimes a government's brand can make a government behave to respect its own ideas, and at other times it can cause it to ignore or reject rational ways forward (Marland, 2017).

The interplay of these decision criteria is variable and complex. On any given priority or decision, one or more of these may take precedence at any given time. However, some broad considerations can be drawn between them. When fiscal discipline is combined with political responsiveness, usually there is a concern that expenditures be controlled, as has been observed during election cycles. When fiscal discipline is combined with program effectiveness, there is usually a concern that budgets are sufficient to provide assurance that government activities will yield desired effects. One often sees these priorities play out during governmental budget deliberations and supplemental debates. Finally, when program/service effectiveness is combined with political responsiveness, there is often a concern for program performance and results. This concern may be observed during times when the public is engaged, including current efforts at Indigenous reconciliation or foreign trade, or when governments provide reports on their performance, or when others such as the auditor general assess programs and services.

The specific role of program evaluation was regarded as an innovation in the 1970s, but its role supporting these different decision criteria has changed many times since. New public management (NPM) aims for results-based management and reporting have guided the objectives of the function, but even these purposes within NPM have shifted between fiscal discipline, performance monitoring, and program effectiveness. The extent of evaluation's contribution to governmental decision-making and expenditure management has proved to be less than stellar (see Bourgeois & Whynot, 2018), due mainly to confused responsibilities as evaluation contributes to governmental decision-making priorities (Segsworth, 1990; Shepherd, 2011). The sources of this confusion are important to consider to reorient the function.

## **ASSURANCE AND THE FEDERAL POLICY CONTEXT**

To achieve the aims of fiscal discipline, program effectiveness, and political responsiveness, various forms of assurance are needed to support political

decision-making. The present pattern is that cabinet and departmental decision-making is highly centralized and concentrated in the hands of the prime minister and senior advisors (Savoie, 1999, 2010; Dunn, 2010). Decisions regarding the connections between public policy and programmatic decisions related to design, implementation and resource utilization, are fueled by shifting political responsiveness to public needs and wants, implementation of election promises, and personal priorities of the prime minister and, to some extent, of cabinet. Typically, the Canadian system is highly centralized and concentrated, because there are fewer checks on central power (Savoie, 2010). Such concentration has allowed for much greater emphasis on fiscal discipline, a decision criterion that voters demand as a matter of routine. The key to success for the Harper government was the near obsession with fiscal balance to the exclusion of other priorities at times (Evans & Albo, 2011). Greater attention to fiscal discipline is attributed to heightened public demands for financial transparency, results achievement, and value-for-money. Likewise, various media are becoming vigilant to check government financial decisions, and the extent to which duplication with provincial/territorial activities is kept to a minimum (Aucoin, 2012). Such pressures have required careful attention to ongoing expenditure review and resource allocation in response to changing priorities. Interestingly, however, it has also meant more attention to program effectiveness and results achievement as governments are regarded increasingly as incapable of addressing more complex problems (Noveck, 2015).

These changes have meant that governments rely on various forms of assurance to give evidence that fiscal discipline is being maintained and that programs are performing optimally. Table 1 summarizes the various forms of assurance provided by internal and external systems, and the policy frameworks that support them. More importantly, these forms of assurance speak to different purposes and target audiences. At a minimum, there are at least five forms of assurance that are demanded by different decision actors: policy and program coherence, systems improvement, internal program performance, government-wide performance, and public accountability. Each form of assurance can individually or collectively support one or more decision criteria in Figure 1.

The role of policy and program evaluation can be observed in each of these various assurance forms, oftentimes with overlapping and layering of purposes. Such assurance speaks to the interaction and complexity of internal assurance and accountability functions, such as program evaluation as it contributes to several policy frameworks and instruments. To be clear, it is no longer the case that program evaluation focuses exclusively, if it ever did, on summative concerns for results. It has also been concerned with formative questions of program design and implementation, value for money, system and service gaps, public accountability, and, in rare cases, policy coherence and rationale (Shepherd, 2011). Evaluations serve a multiplicity of purposes, and the function itself is asked to contribute to several departmental and central agency projects, including ongoing expenditure reviews. With so many competing and overlapping

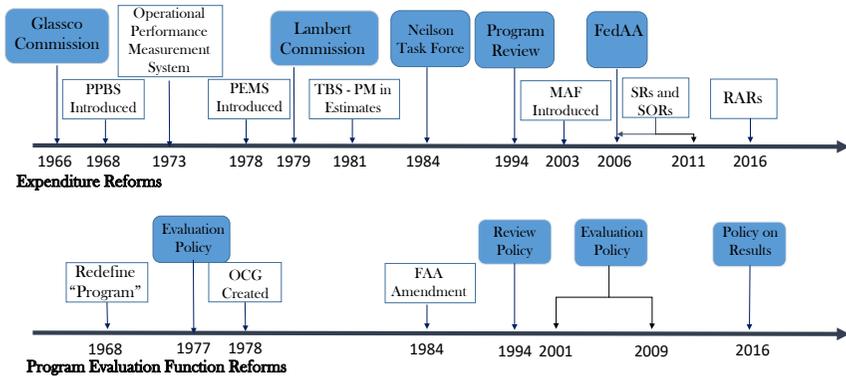
**Table 1.** Forms of Assurance and Policy Frameworks

Form of assurance	Purpose of assurance	Policy frameworks	Target for assurance
Policy and program coherence	<ul style="list-style-type: none"> <li>• Relevance and coherence of policy ideas</li> <li>• Assess program relevance against ideas</li> </ul>	<ul style="list-style-type: none"> <li>• External audit (OAG)</li> <li>• Evaluation/results</li> <li>• Internal audit</li> </ul>	<ul style="list-style-type: none"> <li>• Parliamentarians</li> <li>• Senior departmental and central executives</li> </ul>
Internal program performance	<ul style="list-style-type: none"> <li>• Monitor program implementation</li> <li>• Monitor expenditures</li> </ul>	<ul style="list-style-type: none"> <li>• Evaluation/results (PIPs/performance Measurement)</li> <li>• Expenditure review</li> </ul>	<ul style="list-style-type: none"> <li>• Senior departmental and central executives</li> </ul>
Government performance	<ul style="list-style-type: none"> <li>• Governmental election targets</li> <li>• Program coordination</li> </ul>	<ul style="list-style-type: none"> <li>• Evaluation/results</li> <li>• Expenditure management</li> </ul>	<ul style="list-style-type: none"> <li>• Ministers</li> <li>• Central executives</li> </ul>
Systems improvement	<ul style="list-style-type: none"> <li>• Quality improvement</li> <li>• Performance improvement</li> <li>• Assess system gaps</li> </ul>	<ul style="list-style-type: none"> <li>• MAF</li> <li>• Evaluation/results</li> <li>• Internal audit</li> </ul>	<ul style="list-style-type: none"> <li>• Senior departmental executives</li> <li>• Senior corporate services managers</li> </ul>
Public accountability	<ul style="list-style-type: none"> <li>• Governmental accountability for performance</li> </ul>	<ul style="list-style-type: none"> <li>• CIO/CHRO/CFO</li> <li>• Evaluation/results</li> </ul>	<ul style="list-style-type: none"> <li>• Parliament</li> <li>• Public</li> </ul>

responsibilities, it is small wonder that federal evaluation has become cumbersome and finds itself in difficulty managing a plethora of demands and expectations from various constituencies. One can also argue that its less than stellar performance can be attributed to doing too many things, under conditions of constantly shifting demands and within a policy framework that is not clearly linked to decision-making. In effect, the relevance of the function is constantly being questioned by increasingly more governmental actors ([Shepherd, 2011](#); [Savoie, 2013, 2015](#), pp. 182–186).

## THE CONTRIBUTION OF PROGRAM EVALUATION TO PROVIDING ASSURANCE

The following provides a series of narratives on the contribution of program evaluation to the various forms of assurance, including, most notably, decision-making



**Figure 2.** Timeline of Reforms: Expenditure Reviews and Evaluation

with respect to fiscal discipline and program effectiveness. The narratives on each of the forms of assurance show the multiple and sometimes competing purposes of the function, and that evaluation is not ideally suited to supporting expenditure reviews. [Figure 2](#) provides key turning points in the overall narratives of expenditure reforms and reviews, and illustrates how the program evaluation function and its several reforms have played out with respect to serving expenditure reviews. Using the various forms of assurance, these turning points can be understood from several perspectives, including changes to government policy on managing transparency, accountability, and program effectiveness.

### *Policy and Program Coherence*

Cabinets of the pre-1970s period from the Depression onward examined policy and program proposals collectively through an elaborate array of cabinet committees intended at the time to bring coherence to the diffused system of decentralized ministry decision-making that served the country well through the war years. In a system where powerful individual ministers held sway in cabinet, the value of evidence to convince colleagues of their policy ideas was critical ([Aucoin & French, 1974](#); [Dunn, 2010](#)). Cabinets invariably gave way to powerful ministers. Federally, “there appeared to be a symbiotic relationship between effectiveness evaluation and Cabinet decision-making. This symbiosis was based, to some extent, on confidence in a highly technocratic and professional public service” ([Shepherd, 2011](#), p. 6).

It made sense, then, that under this form of decision-making the idea or definition of “programs” was any activity of government on a small or large scale, referred to as big-*P* programs. Cabinet was able to weave together a vision of its overall policy program or government-wide initiatives (big-*P* programs) expressed through the Speech from the Throne. Evaluations were focused on

examining matters of relevance at the level of policy ideas and the extent to which big-*P* programs served to meet policy expectations of cabinet. One challenge at the time, however, was that line budgeting conducted at the centre of government determined ministry allocations that were out of step with policy-level thinking. That is, budgeting was conceived on the basis of the number of desks, people, and pads of paper, rather than as a coordinated “program” to resolve a policy problem. The combination of the Glassco Commission recommendations to “let the managers manage” and decentralize financial management and planning authority to ministries, and the introduction of the Planning, Programming Budgeting System (PPBS) in 1968 and the Policy and Expenditure Management System (PEMS) in 1978, shifted the conception of “program” radically. The *Planning Programming and Budgeting Guide* now defined program to mean “a collection of activities having the same objective or set of objectives” (Canada, TBS, 1969). That is, programs were considered a collection of ministry or departmental micro-activities, rather than a government-level big-*P* initiative.

When the *Evaluation Policy* was first instituted in 1977, its focus was to be on small-*p* programs resident within individual ministries or departments and on prescribed questions about micro-program accountability and performance, rather than questions of policy relevance and higher-order outcomes orientation (Jordan & Sutherland, 1979). This was eventually reinforced by directives from the newly created Office of the Comptroller General in 1978, despite recommendations from the Royal Commission on Financial Management and Accountability (Lambert Commission) in 1979 that evaluations contribute to big-*P* programs and report on all programs on a 5-year cycle. The introduction of the Operational Performance Measurement System (OPM) in 1973 further shifted the focus of evaluation from program effectiveness to program performance. Between the PPBS and OPM, the role of internal evaluation was solidified to enable “adequate and reliable means, wherever feasible, for performance measurement” (Canada, TBS, 1976, p. 5). At the time, the function did not believe it could properly assess program effectiveness, program performance, and fiscal accountability simultaneously (Shepherd, 2011, p. 7). The problem was, and remains, that evaluations geared to assess micro-programs are not placed well to assess budget allocations or expenditures regarding big-*P* programmatic or summative issues: the targets of analysis are at different levels of perspective.

The 1977 *Evaluation Policy* had cast the function in a way that manifested in the decentralized governance structure that remains in place to this day. Evaluation was regarded as a management tool to support departmental decision makers in improving essentially the efficiency of their activities. In 1981, TBS directed departments to collect performance information to include in their annual estimates to Parliament. However, the type of performance information to include was left to the judgement of program managers, without tying such information to policy-level outcomes. This further reinforced the role of internal evaluation to support small-*p* programs, and to assist program managers in their efforts to

collect output-level data, rather than outcomes or effects of programs. Amendments to the Financial Administration Act in 1984 further required departments to gauge value for money of small-*p* programs, reinforcing the evaluation's emphasis on programmatic activities and outputs.

In 1993, the auditor general concluded in a scathing audit of the evaluation function that “priority has been given to meeting the needs of departmental managers. As a result, evaluations examine smaller program units or lower-budget activities and focus on operational performance. They are less likely to challenge the existence of a program or to evaluate its cost-effectiveness” (Canada, Office of the Auditor General [OAG], 1993, sec. 9.2). The same audit recommended that evaluations should be subject to external review to improve objectivity (sec. 10.2), that timeliness and relevance could be linked to activities such as ongoing expenditure reviews (sec. 10.4), and that the function could be better linked to decision-maker needs (sec. 10.6). In effect, it was concluded that, for program evaluation to be more effective, it must have an effective entry into decision-making. This conclusion has been reached many times since by other commentators (Shepherd, 2011; Savoie, 2013), including other articles in this volume.

The focus on evaluation as a management function was reinforced with the 2001 *Evaluation Policy*, which emanated from the Results for Canadians initiative in 2000, and manifested in a results-based management orientation (G. Young, 2006). The Results for Canadians policy called for better management of programs and services, and integrating departmental activities into policy decisions. The challenge, however, was that the function could not evaluate all departmental programs and services, thereby building a comprehensive picture of overall departmental performance (Dumaine, 2012, p. 68). With the introduction of the 2009 *Evaluation Policy*, departments and agencies were required to assess all direct program spending, and the administrative aspects of all statutory spending on a 5-year cycle. This built on previous efforts emanating, from the 2007 budget, that all grants and contributions programs also be evaluated on a 5-year cycle. Ultimately, however, the coverage requirements proved not to work. In the fall of 2009, the auditor general released a chapter, “Evaluating the Effectiveness of Programs” (Canada, OAG, 2009), concluding that departmental evaluations covered a relatively low proportion of its program expenses—between five and thirteen percent annually across the six departments [and that] the audited departments do not regularly identify and address weaknesses in effectiveness evaluation” (sec. 2.11). This coverage requirement has since been removed in the *Policy on Results* 2016, with a shift to risk-based assessment of programs to be evaluated in the “program inventory” (secs. 4.3, 15.2). The challenge now is that many evaluations are focusing on multiple programs in the same report with no coherent general picture for assessing departmental spending and performance. That said, the new policy holds the promise of aligning program results with governmental policy priorities. This might assist with creating evidence that better supports expenditure reviews, according to respondents.

## Internal Program Performance

Program evaluation has long focused on internal small-*p* program effectiveness and performance as a form of assurance (Shepherd, 2011). However, it has not always been clear how such evaluations inform departmental expenditure reviews. The main challenge, as noted, was that the targets of assessment are misaligned. Whereas evaluations assess program effects (either formatively or summatively), expenditure reviews aim to reduce overall departmental spending, and, by extension, government-wide expenditures. In addition, evaluations have not typically tracked the connections between program performance and effects, given that there has often been a dearth of performance measurement information. To fill this information gap, the 2001 *Evaluation Policy* required departments to establish ongoing performance monitoring and measurement, and many efforts since the 2007 budget have been made to improve performance information supporting expenditure reviews. During the 2001 policy period, evaluation was driven by central agency concerns for accountability and, to a much lesser extent, by program improvement and renewal (Borys et al., 2005). It required that TBS submissions for new programs include a Results-based Management and Accountability Framework (RMAF), which was essentially an evaluation plan that included the construction of logic models which had two purposes: to show the connection between program activities and spending; and, to demonstrate the connection between ultimate results and governmental level policy objectives.

The objective of the 2009 *Evaluation Policy* was “to create a comprehensive and reliable base of evaluation evidence that is used to support policy and program improvement, expenditure management, cabinet decision-making, and public reporting” (Canada, TBS, 2009, sec. 5.1). The policy further defines evaluation as “the systematic collection and analysis of evidence on the outcomes of programs to make judgements about their relevance, performance, and alternative ways to deliver them or to achieve the same results” (Canada, TBS, 2009, sec. 3.1). The combination of both sections of the policy demonstrates the scope of the challenge for evaluation: there are multiple purposes dependent on performance measurement and several forms of assurance demanded for different audiences. The ambitions for the function are simply too high and destined to fail. If this were not enough, the policy makes a shift toward value-for-money as a principal aim in an effort to better support expenditure reviews. The challenge of the 100% coverage requirement was difficult to satisfy, making the function’s overall contribution to departmental expenditure reviews questionable at the very least.

The new *Policy on Results* 2016 places increased emphasis on department-level performance measurement and less on program effectiveness, thereby further limiting the evaluation function’s efficacy. Again, the theory is that, by building a comprehensive picture of program performance, departmental performance can be assessed. This has proved elusive to date, and it will be interesting as to whether the issue will be resolved in the new policy.

## **Whole-of-Government Performance**

Several efforts have been made to improve assurance for governmental performance, including the new *Policy on Results* 2016 as the latest attempt. The managerial alignment of spending with priorities was observed most notably with the Lambert Commission, which recommended in 1979 that central controls over departmental spending be fortified in response to a blistering auditor general report in 1976 accusing the government of “losing control over the public purse” (Canada, OAG, 1976, sec. 1.2). Lambert proposed the creation of a fiscal plan for government, covering 5-year periods, which, when supported by data gathered through the OPM, would see the federal government allocate resources according to central priorities within budgetary limitations. Fiscal plans were set by the Treasury Board, Privy Council Office, and the Department of Finance with the effect that spending limits were now placed on departmental spending (Sutherland, 1986, pp. 119–124). In 1985, the Conservative government of Brian Mulroney established a taskforce led by Finance Minister Erik Nielsen to reduce the public service by 15,000 positions within 3 to 5 years and make programs and services more efficient and effective. The evaluation function was called upon to provide relevant assessments of program performance and spending that supported government-level decisions. However, its input was regarded largely as ineffectual, as the function could not support efforts to provide information on strategic-level objectives (Mayne, 1986, pp. 98–100).

In 1993, an important study by former Clerk of the Privy Council Gordon Osbaldeston concluded that, despite important shifts in management resulting from PPBS and PEMS to improve departmental financial management, the role of TBS should be shifted from ensuring accountability and control to one that creates the conditions necessary for better management. This led to increased political direction and control over financial management, resulting in more centralization of decision-making (Osbaldeston, 1989). A second one-time program review was conducted under Jean Chrétien in 1994 to improve the management of government, which acknowledged that any policy expenditure management system must be fully integrated with systems of budgetary review that accounted for government priorities that drew on evidence about program effectiveness. The program review exercise was successful in ensuring that spending aligned with policy priorities, and there is evidence to suggest that program effectiveness was considered in spending allocations, but mainly it was an exercise in fiscal discipline (Paquet & Shepherd, 1996). TBS played a major role to set budget reduction targets, departments carried out their own reviews and reduction recommendations to the centre, and cabinet along with senior public service officials played a challenge role in the recommendations posed by departments (Paquet & Shepherd, 1996). The role of evaluation was once again questioned in this process, as it failed to deliver the kind of information needed to support fiscal discipline decisions (Müller-Clemm & Barnes, 1997).

Following the experiences of that program review exercise, a new Expenditure Management System (EMS) was instituted in 1995 to replace PEMS with

a greater role for the finance minister and prime minister in financial management decisions. There were three main goals of the EMS. First, programs would continue to be reviewed on an ongoing basis, especially where departments were requesting new money. Second, annual targets were set for deficit reduction, and achieved through A-based reviews. Lastly, all new expenditure proposals for new programs would be considered by the minister of finance, and accommodated in the budget and fiscal framework.

The challenge for this new system, however, was that it did not provide for a continuous process of expenditure review and reallocation for the whole of government, a role assumed directly by the prime minister and finance minister. Program evaluation under this system was reduced to simply providing performance measurement evidentiary support for budget reviews, solidified under a new review policy instituted in 1994. However, a significant change was that all new funding proposals were to be accompanied by an evaluation plan, as required in the 1994 *Evaluation and Review Policy*. Submissions were to address seven program review questions and to demonstrate how the program would achieve its objectives (Shepherd, 2011, pp. 9–11), which further reinforced evaluation's role supporting mainly fiscal discipline.

In 2007, so as to support a whole-of-government approach to decision-making, the Treasury Board instituted the Management, Resources, and Results Structure Policy (MRRS), which stipulated the performance effectiveness and evidence orientation to demonstrate departmental results (Canada, TBS, 2007). The policy required that departments identify clear and measurable strategic-level outcomes, an effective governance structure to achieve such results, and a detailed "Program Activity [Alignment] Architecture," which would ensure a results focus. Departmental evaluations were to serve this framework and provide input to outcomes-level decision-making. The challenge under this approach, which still persists today, is that programmatic activities might address more than one strategic outcome, making it difficult to separate expenditures supporting any one outcome. The level of program assurance needed to support government-wide reporting was and remains limited.

Also in 2007, the Harper government announced the formal ongoing strategic review initiative in its budget as a way to re-establish the policy and expenditure decision-making connection observed in the 1994 program review. The initiative sought to find cost savings of approximately \$2.8 billion. The Privy Council Office was afforded the main policy challenge responsibility, observed also in the program review. In short, strategic review was a mandatory process for all federal departments and agencies to assess all of their direct program spending, including the operational costs of statutory programs on a 4-year cycle. Direct program spending also included financial contributions made to third-party delivery organizations such as First Nations governments on-reserve, sports granting bodies, research organizations, such as universities, and health organizations carrying out drug tests, to name a few. It also included ongoing operating and capital spending to deliver programs such as Veterans services and regional economic

development. What was not included in strategic reviews was transfers to individuals (tax expenditure programs, for example) and other orders of government, which represents approximately 45%–50% of all federal spending.

Departments and agencies were required to respect two key conditions surrounding the operation of strategic reviews. First, they were to address the program review's seven questions in a cascading scale of priorities. Second, departments and agencies "are required to identify reallocation options totaling 5 percent from their lowest-priority, lowest performing program spending," and channeling the "savings" to high-priority and high-performing programs identified by central decision makers. The result of the 2007–2010 series of strategic reviews was less about reinvestment, and more about deficit reduction similar to that of the 1994 program review exercise (Dumaine, 2012, p. 67). The role of program evaluation was to contribute as one of several sources of information. Other sources included reports from the Management Accountability Framework instituted in 2003 to provide ongoing performance information of internal departmental and agency operating systems and management strength. Additional sources included audit reports, program performance reports, and external reports, such as those from the auditor general.

The strategic reviews initiative was expanded in the 2011 federal budget and renamed the Strategic and Operating Review (SOR) initiative, which was intended as a 1-year process led by the Treasury Board to identify efficiencies in the public service, including assessments of salaries, benefits, and outsourcing of work. As opposed to strategic reviews, SORs

will examine direct program spending, as appropriated by Parliament. About \$80 billion of direct program spending will be reviewed with the objective of achieving at least \$4 billion in ongoing annual savings by 2014–15 or 5 percent of the review base. The review will place particular emphasis on generating savings from operating expenses and improving productivity, while also examining the relevance and effectiveness of programs. (Government of Canada, 2011, p. 182)

Like the strategic review exercise, SORs did not include transfers to individuals and other governments. Whereas the strategic review exercise did not lead to any major reductions in people and program spending, SORs did produce significant reductions between 2013 and 2015 in planned operating spending, and public service positions (Rounce & Levasseur, 2015, pp. 29–35).

In both cases of strategic reviews and SORs, program evaluation has been observed as not living up to expectations on serving expenditure reviews (Lester, 2015). The latest Policy on Results 2016, which replaces the 2009 *Evaluation Policy*, aims to "improve the achievement of results across government; and [to] enhance the understanding of the results government seeks to achieve, does achieve, and the resources used to achieve them" (Canada, TBS, 2016b, sec. 3.1). The policy is premised on the work of Sir Michael Barber (2010), and his concept of *deliverology*, which is a manifestation of results-based management that highlights political results and their measurement from the centre of government,

and emphasizes a whole-of-government approach. Government-level results are gathered through delivery units that coordinate with other results-measurement functions, such as internal audit and program evaluation. The focus of the policy is performance measurement, and is much less about program effectiveness, which accords with the UK model of program assessment (Schacter, 2016). The entire focus of the policy is to improve government-wide performance results against centrally contrived targets, which is aided by a new role for the Treasury Board to carry out government-wide evaluations (Canada, TBS, 2016b, sec. 5.2.3). The question is whether program evaluation can contribute effectively in this model, which is addressed in the next section and by Dobell and Zussman (2018). There are those, including Montague, who believe a move to greater performance measurement with a proper focus on targets can improve government-wide performance reporting (Montague, 2016). The real question is whether program effectiveness will be considered a priority as well.

### ***Systems Improvement***

Also arising from the Results for Canadians initiative was attention to ideal management systems and practices. This culminated in the creation of the Management Accountability Framework (MAF) “to develop a comprehensive system that would attempt to gauge and report on the quality of management of departments and agencies, and encourage improvement every year” (Lindquist, 2010, p. 51). It was designed to improve management performance in 10 areas of management including accountability, results, and performance. The MAF has undergone several amendments, with the most recent update in 2016, focusing on leadership and strategic direction, and results and accountability. Each year selected departments and agencies undergo MAF assessment, culminating in an annual government-wide report on the state of management practices and performance (Canada, TBS, 2016a). Such assessments are tied directly to the *Policy on Results*, and the evaluation and internal audit functions are assessed on their ability to support effective departmental management and expenditure decisions. In addition, these functions are also expected to contribute evidence that supports MAF assessments at the department level. Again, the same challenge exists that these functions tend to operate at the micro-program level and are not well positioned to contribute to larger questions of management practice beyond program management.

### ***Public Accountability***

A final form of assurance to which evaluation contributes is public accountability. The principal target audience of public accountability reporting, including departmental performance reports, for example, are Parliament and the public. Evaluation contributes to preparing public reports that are typically posted on departmental and central agency websites. However, the clear link between programmatic activities and departmental performance is often unclear (Savoie, 2013). Various evidentiary forms, such as evaluation reports, are submitted to

chief information officers (CIOs), chief human resources officers (CHROs), and chief financial officers (CFOs) to glean what they can on higher-level performance findings. At best, evaluation reports can highlight programmatic conclusions, but they contribute in a cursory way to departmental performance. That said, some departments are able to use evaluation findings to build a performance and expenditure management picture more effectively than others, but these are a minority.

## **EXPENDITURE REVIEWS: REALIGNING THE PROGRAM EVALUATION FUNCTION**

There have been multiple academic and expert papers, academic conferences, and government symposia regarding how to adjust the federal evaluation function since 1977, and this article is but one more in a long line. The particular twist in this article is how to make program evaluation more relevant for purposes of expenditure reviews and results-based management in general. Evaluation to this point has been in place mainly to serve the accountability requirements of government (Turpin, 2009, p. 7), and this is both an asset and limitation. It is an asset because it focuses the function on specific objectives. It is a limitation because the function is capable of more than contributing to fiscal decision-making, which has been observed as less than stellar to date. The following provides general thoughts, collected from respondents, on where the function might better focus its energies and competencies.

### ***Evaluation as a Core Function of Public Management: PM Is Not Enough***

Aucoin (2005) recognized that evaluation is a core function, “because it seeks to ascertain and assess the effectiveness of government programs in achieving desired results, impacts and outcomes.” He argues, rightfully, that a core function of any government is to be responsive to citizen needs and that programs operate to maximize benefit. He further argues that

results-based management [read *Policy on Results*] except insofar as it fully incorporates program evaluation, is no substitute for program evaluation, however useful it may be for management control and improvement. Performance measurement regimes do not seek to ascertain or assess program effectiveness. Rather, they seek to determine the extent to which departments achieve results or outcomes. They measure achievement against targets. They do not attempt to explain or account for the performance in question, let alone the effectiveness of their programs. . . . Program evaluation is not just another initiative: it is a core function of governance and management. (p. 6)

That “evaluation” has been removed in the new *Policy on Results* is telling in what may be argued is a diminished role. Equally telling is the emphasis on results, defined as stated results against prescribed targets, using deliverology-style

performance measurement as the preferred approach to assessing them. Although concerns for program effectiveness are identified in the policy (Canada, TBS, 2016b) and the glossary to the policy, such concerns are not stated anywhere else in the substance of the policy, despite promises in the mandate letter to Treasury Board Minister Scott Brison to research questions of program effects:

Take a leadership role to review policies to improve the use of evidence and data in program innovation and evaluation, more open data, and a more modern approach to comptrollership . . . [and to] work with the Minister of Finance and your colleagues to conduct a review of the expenditures and other spending to reduce poorly targeted and inefficient resources, wasteful spending, and government initiatives that are ineffective or have outlived their purpose. (Canada, Office of the Prime Minister, 2016)

The point may be semantic, but it appears evident that results units within the federal bureaucracy, accompanied by evaluators and other monitoring functions are to focus on performance measurement as the key to assessing results achievement. Respondents interviewed for this paper expressed concern that the “usefulness” of evaluation as a function has gone too far to follow the UK approach to performance measurement. More importantly, the centralization of priority setting, manifested as performance measures and targets, could mean that researching program effectiveness as a core function of government, and the capacity to understand it in ways other than through internally derived performance data, may be vastly diminished as time goes on.

### ***Evaluation Clients and Purposes Must Be Aligned: Trust Is Critical to Success***

As argued throughout this paper, there are several clients of evaluation. However, the function is designed mainly to serve program managers. This is revealed in the types of questions raised by evaluations. They tend often to be concerned with formative questions of program efficiency, cost-efficiency, and outputs achievement. Senior management was not included often in the scoping of evaluations in the past nor in the communication of results through departmental performance reports, which did not contribute much to the usefulness of the function. Technical advisory committees for evaluations more often than not comprise program officials, which limits significantly the scope of evaluation questions and the latitude of evaluators to ask tough or probing research questions. With greater use of outside beneficiaries and experts, technical advisory committees would be able to scope evaluations beyond simply a programmatic focus. This may not always be appropriate, but it would afford these committees the ability (and permission) to consider broader issues of policy ideas, user focus, and innovative methods that take into consideration other epistemological frameworks.

One obvious benefit of the *Policy on Results*, is that performance measurement has focused senior managers to consider results that are of interest to the Cabinet Committee on Results. In this regard, deputy ministers are concerned

with the information that is relayed to ministers, because that information must be pertinent to central decision-making and supported by performance data. There is now much greater potential to scope evaluation reports to serve a results focus, as envisioned in the policy. The next challenge would be to also reincorporate program effectiveness in the traditional sense (i.e., the extent to which programs are observed to cause or contribute to achieving expected effects).

With respect to expenditure reviews as a specific form of inquiry, senior departmental and central agency decision makers and evaluation units would have to create a trust relationship that allows for conversations not as much about costs as *aligning* spending with departmental objectives and central agency priorities and broadly defined meta-program objectives at the departmental and central government levels. One concern with the *Policy on Results* is that it focuses attention too heavily on centrally derived priorities to the exclusion of department-level objectives, which are based in legislative obligations. Central priorities get attention: ongoing departmental programs and services often do not. Or, it may not be obvious as to the connection between routine and ongoing activities and programs with central priorities. A few departments have included evaluation units in strategic-level decision-making on expenditure reviews, but this is not commonplace. It requires evaluation heads with a great deal of organizational confidence and moral authority to make this work, and to make their usefulness in such conversations crystal clear. Data sharing and expertise are essential ingredients in this trust relationship between those carrying out evaluations and those performing resource alignment reviews.

Building on this trust relationship means moving away from the accountability focus of evaluations that characterized both the 2001 and 2009 evaluation policies with a highly mechanized adherence to centrally derived questions. Evaluation units now would have to be reformed to be consultative and responsive to both program and strategic management needs. In this regard, evaluations have to be much more learning and problem-solving oriented than simply the ticking off of accountability boxes. Although coverage remains a requirement under Section 42.1 of the Financial Administration Act for grants and contributions programs, there is sufficient latitude in the coverage requirements in the *Policy on Results* to allow evaluation units to better calibrate direct spending coverage with programmatic and strategic needs, including expenditure management decision-making (Dumaine, 2012, pp. 71–72).

Equally important, a trust relationship has to be built between heads of evaluation and TBS. Given the compliance focus of the 2009 policy especially, TBS tended to be highly critical and prescriptive in its direction to departmental evaluation units. The role of evaluation in expenditure reviews was not clear from the centre, further eroding confidence in the TBS to create the conditions necessary for a useful contribution. The new policy provides an alternative approach: to place evaluation in the position of trusted advisor as indicated. The challenge for the centre of government will be to create the space needed for trust to be built with senior central decision makers. The challenge is this flexibility is not being used.

## ***Balancing Interests: Evaluation Use and Who Should Conduct Expenditure Reviews?***

Questions about whether program evaluation has a role to play in expenditure reviews have been swirling since the *Evaluation Policy* was instituted in 1977. As shown, evaluation units have been ill-equipped to contribute meaningfully in this respect despite attempts since the 2009 *Evaluation Policy* to better link evaluation with the strategic reviews initiative and its successor versions. It is fairly safe to conclude that there is significant structural resistance to contributing to expenditure reviews as long as the function is aligned to serving program managers and resident in departments. It may be time to accept the conclusion that [Mayne \(2018\)](#) has come to, which is that expenditure evaluations be adopted as a separate evaluation tool. The question then becomes, who may be better positioned to carry these out?

The usefulness of evaluations has long been discussed in the literature ([Shulha & Cousins, 1997](#); [Borys et al., 2005](#)). A major conclusion from the various literature is that there has been a significant expansion of conceptions of use from the individual to the organization levels. In short, evaluation uses and users have changed and expanded, leading to heightened expectations on the value of evaluation in decision-making. It would seem to make sense that evaluations cannot serve all of these expectations well. Building on the last observation, it may be time to apply the application of “full-meal deal” evaluations only when considerations of effects are warranted. There should also be “evaluation-lite” options, where very specific questions are being posed, such as value-for-money analyses. Creating specialized evaluation functions to address the several expectations related to timing, usefulness, scope, depth, and breadth may be an idea that has come of age in this time of “experimentation.” In other words, building a toolkit of evaluation types based on more than one epistemological framework is warranted to address different questions and expectations.

There may be justification for at least three possibilities in this regard. First, a unit could be created in the Treasury Board Secretariat that is trained to carry out these sorts of reviews ([Mayne, 2018](#)). Resources would have to be made available, given that TBS is in no way prepared to handle such a role at present. As Mayne suggests, there is merit to understanding the financial performance of programs from a central agency perspective. As there may be better positioned to provide a whole-of-government rationale for results. Although Mayne is referring to the effectiveness performance of programs, it is equally the case that financial performance of departmental and governmental activities may be better assessed from a holistic point of view. In addition, assessing governmental decision-making on the basis of go and no-go, using cost information, is not likely to be meaningful, as calculating value for money may be quite complicated, especially from a qualitative perspective. Rarely can programs be assessed mechanistically, despite the rhetoric of deliverology to the contrary. Also, TBS is better suited to get the timing right for linking expenditure evaluations to the budget cycle, rather than

relying on departmental program-based evaluations to somehow be timely with multiple planning and deadline expectations.

A second idea is to maintain a decentralized approach and house expenditure evaluation within departmental strategic planning or policy branches. Aside from the benefits of growing separate expertise, such units are better able to raise evaluation from a small-*p* focus to a higher-level of analysis and perspective. Although assuming responsibilities for comptrollership may be considered out of place for such units, it nonetheless makes better sense to unlink this type of evaluation from current structures. One could go so far as to argue that the experience of some departments, and small agencies in particular, is that some equivalent unit is at least coordinating these sorts of reviews at present anyway. The implementation challenge would be coordinating with TBS on a manageable template for review, and getting the timing right. The next challenge would then be the unit of analysis, whether this is to be programs or clusters of activities in the program inventory.

The benefit of actors other than evaluation units conducting expenditure reviews is that the burden of budgetary and economic considerations in program evaluations is removed. Although program evaluations are getting better at value-for-money analyses, the unit of measure is still out of alignment with programs being the centre of assessment. In addition, the users of expenditure review information tend to prefer quantitative or economic analysis of departmental activities, something again that program evaluations are not always able to provide definitively.

Regardless of which idea is embraced, the usefulness of program evaluations would continue to factor into expenditure reviews. One way to contribute that serves the *Policy on Results* would be to use evaluation expertise in the development of realistic and appropriate targets. As [Montague \(2016\)](#) suggests, program evaluators may be well-suited to identifying policy or program problems, gaps and opportunities, engaging stakeholder interests (both internal and external) to develop appropriate targets, tying them to behavioural change logic frameworks, and then linking them to life cycle plans at the program- or policy-level of change. In this regard, getting the targeting right leads to better measurement of outcomes, which may actually improve the allocation of resources.

## CONCLUDING THOUGHTS

The article highlights in a detailed way some of the reasons why federal program evaluation has not performed well in supporting expenditure reviews. By moving through such a diagnosis a discussion on realignment may be better annotated. In overall terms, it is hoped that a clear argument has been made that while, evaluators and evaluation units are well-trained, well-intentioned, and well-positioned within organizations to contribute to effective decision-making, there are structural challenges that have diminished their usefulness, especially in regard to matters of fiscal discipline and programmatic effectiveness. Most important among these is that program evaluation contributes to many forms of assurance, each demanding a particular focus, scope, depth, timing, and epistemological orientation.

The contribution of program evaluation to expenditure reviews given the *Policy on Results* requires a specific set of skills, a defined time frame where analyses become useful, and depth of analysis that cannot always be accommodated in traditional effectiveness evaluation studies. One repair is to accept that a separate type of evaluation (expenditure evaluation) with an independent and dedicated responsibility centre (Curran, 2016) could be created that offers the potential to improve evaluation use. A one-size-fits-all approach to program evaluation may be an idea well past its prime. Although there are several options considered to implement such an idea, ultimately it is likely that expenditure reviews will remain a basic public management function, requiring a specific tool to accommodate them. In essence, it is time to reorient and focus the function on what it does best: contribute to program improvement and organizational learning.

Regardless of whether one adopts a centralized or decentralized approach to expenditure evaluation, building a different trust relationship from traditional program evaluation will be essential. Such evaluators will need the confidence of program managers and department-level decision makers to use data from various sources and come to recommendations that make sense in the budget cycle. Building such relationships will clearly take time, but the incremental rewards may outweigh the investments of effort. As Bourgon (2009) observes, “citizens pay the price of failure when governments make the wrong choices, or make choices based on poor or misleading information. Evaluation continues to play a vital role in government, but it can do more and do better given proper aligning of capabilities with decision-makers’ expectations.”

## ACKNOWLEDGEMENTS

I am grateful for the expert research contribution of Benoît Gauthier, who assisted with much needed interviews to support this paper. I am also indebted to all those individuals across the federal public service who agreed to be interviewed and gave of their time to contribute and review this piece. All errors that remain are my own.

## REFERENCES

- Aucoin, P. (2005). *Decision-making in government: The role of program evaluation* (Discussion Paper). Retrieved from the Treasury Board of Canada Secretariat website: <https://www.tbs-sct.gc.ca/cee/tools-outils/aucoin-eng.asp>
- Aucoin, P. (2012). New political governance in Westminster systems: Impartial public administration and management performance at risk. *Governance*, 25(2), 177–199. <https://doi.org/10.1111/j.1468-0491.2012.01569.x>
- Aucoin, P., & French, R. (1974). *Knowledge, power and public policy*. Ottawa, ON: Science Council of Canada.
- Barber, M. (2010). *Deliverology 101: A field guide for educational leaders*. Thousand Oaks, CA: Corwin.

- Borys, S., Gauthier, B., Kishchuk, N., & Roy, S. N. (2005, October). *Survey of evaluation practice and issues in Canada*. Paper presented at the Joint CES/AEA Conference, Toronto, ON.
- Bourgeois, I., & Whynot, J. (2018). Strategic evaluation utilization in the Canadian federal government. *Canadian Journal of Program Evaluation*, 32(3), 327–346.
- Bourgon, J. (2009). *Program review: The government of Canada's experience eliminating the deficit, 1994–1999: A Canadian case study*. Waterloo, ON: Centre for International Governance Innovation.
- Buss, T. F., & Shillabeer, A. (2011). The emerging field of evidence-based public management. In A. Shillabeer, T. F. Buss, & D. M. Rousseau (Eds.), *Evidence-based public management: Practices, issues, and prospects* (pp. 3–16). New York, NY: Taylor and Francis.
- Canada, Office of the Auditor General. (1976). Report of the Auditor General of Canada. Ottawa, ON: Supply and Services.
- Canada, Office of the Auditor General. (1993). Report of the Auditor General of Canada: “The program evaluation system—Making it work. Ottawa, ON: Supply and Services.
- Canada, Office of the Auditor General. (2009). Evaluating the Effectiveness of Programs. In *Report of the Auditor General of Canada 2009* (Chapter 1). Ottawa, ON: Supply and Services.
- Canada, Office of the Prime Minister. (2016). *President of the Treasury Board of Canada mandate letter*. Retrieved from the Prime Minister's Office website: <http://pm.gc.ca/eng/president-treasury-board-canada-mandate-letter>
- Canada, Treasury Board Secretariat. (1969). *Planning programming and budgeting guide of the Government of Canada*. Ottawa, ON: Supply and Services.
- Canada, Treasury Board Secretariat. (1976). *Circular 1976–25: Measurement of the performance of government operations*. Ottawa, ON: Supply and Services.
- Canada, Treasury Board Secretariat. (2007). *MRRS policy*. Retrieved from <http://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=18218>
- Canada, Treasury Board Secretariat. (2009). *Evaluation policy 2009*. Ottawa, ON: CEE.
- Canada, Treasury Board Secretariat. (2016a). *Management accountability framework*. Retrieved from <https://www.canada.ca/en/treasury-board-secretariat/services/management-accountability-framework.html>
- Canada, Treasury Board Secretariat. (2016b). *Policy on results*. Retrieved from <https://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=31300>
- Canada, Treasury Board Secretariat. (2017). *Management accountability framework*. Retrieved from <https://www.canada.ca/en/treasury-board-secretariat/services/management-accountability-framework.html>
- Curran, R. (2016, July). Why Trudeau needs to embrace expenditure reviews. *Policy Options*, 1–6.
- Dobell, R., & Zussman, D. (2018). Sunshine, scrutiny, and spending review in Canada, Trudeau to Trudeau: From program evaluation and policy to commitment and results. *Canadian Journal of Program Evaluation*, 32(3), 371–393.

- Dumaine, F. (2012). When one must go: The Canadian experience with Strategic Review and judging program value. *New Directions for Evaluation*, 2012(133), 65–75. <https://doi.org/10.1002/ev.20007>
- Dunn, C. (2010). The central executive in Canadian government: Searching for the holy grail. In C. Dunn (Ed.), *The handbook of Canadian public administration* (2nd ed., pp. 85–105). Don Mills, ON: Oxford University Press.
- Evans, B., & Albo, G. (2011). Permanent austerity: The politics of the Canadian exit strategy from fiscal stimulus. *Alternate Routes*, 22, 7–28.
- Government of Canada. (2011). *Budget 2011: The next phase of Canada's Economic Action Plan: A low-tax plan for jobs and growth*. Ottawa, ON: Supply and Services.
- Howlett, M. (2010). The policy process. In C. Dunn (Ed.), *The handbook of Canadian public administration* (pp. 383–400). Don Mills, ON: Oxford Press.
- Jordan, J. M., & Sutherland, S. L. (1979). Assessing the results of public expenditure: Program evaluation in the Canadian federal government. *Canadian Public Administration*, 22(4), 581–609. <https://doi.org/10.1111/j.1754-7121.1979.tb01840.x>
- Lester, J. (2015). Tax expenditures and government program spending. In G. B. Doern & C. Stoney (Eds.), *How Ottawa spends, 2014–2015: The Harper government—Good to go?* (pp. 95–107). Montreal, QC: McGill-Queen's University Press.
- Lindquist, E. (2010). How Ottawa assesses departmental/agency performance: Treasury Board's management accountability framework. In A. M. Maslove (Ed.), *How Ottawa spends, 2009–2010: Economic upheaval and political dysfunction* (pp. 47–88). Montreal, QC: McGill-Queen's University Press.
- Marland, A. (2017). *Brand command: Canadian politics and democracy in the age of message control*. Vancouver, BC: UBC Press.
- Mayne, J. (1986). In defense of program evaluation. *Canadian Journal of Program Evaluation*, 1(2), 97–102.
- Mayne, J. (2018). Linking evaluation to expenditure reviews: Neither realistic nor a good idea. *Canadian Journal of Program Evaluation*, 32(3), 316–326.
- Mintzberg, H., & Waters, J. A. (1985). Of strategies, deliberate and emergent. *Strategic Management Journal*, 6(3), 257–272. <https://doi.org/10.1002/smj.4250060306>
- Montague, S. (2016). *A response to Does "Deliverology" Matter? Targets can work if done properly*. Retrieved from <http://www.pmn.net/wp-content/uploads/Deliverology-Response-June-2016.pdf>
- Müller-Clemm, W. J., & Barnes, M. P. (1997). A historical perspective on federal program evaluation in Canada. *Canadian Journal of Program Evaluation*, 12(1), 47–70.
- Noveck, B. S. (2015). *Smart citizens, smarter state*. Cambridge, MA: Harvard University Press. <https://doi.org/10.4159/9780674915435>
- Osbaldeston, G. F. (1989). *Keeping deputy ministers accountable*. Toronto, ON: McGraw-Hill Ryerson.
- Paquet, G., & Shepherd, R. (1996). The program review process: A deconstruction. In G. Swimmer (Ed.), *How Ottawa spends, 1996–97: Life under the knife* (pp. 39–72). Ottawa, ON: Carleton University Press.

- Robinson, M. (2018). The role of evaluation in spending review. *Canadian Journal of Program Evaluation*, 32(3), 305–315.
- Rounce, A., & Levasseur, K. (2015). Government retrenchment and public service cuts: A tale of two processes. In G. B. Doern & C. Stoney (Eds.), *How Ottawa spends 2014–15: The Harper government—Good to go?* (pp. 25–39). Montreal, QC: McGill-Queen's University Press.
- Savoie, D. J. (1999). *Governing from the centre: The concentration of power in Canadian politics*. Toronto, ON: University of Toronto Press.
- Savoie, D. J. (2010). *Power: Where is it?* Montreal, QC: McGill-Queen's University Press.
- Savoie, D. J. (2013). *Whatever happened to the music teacher: How government decides and why*. Montreal, QC: McGill-Queen's University Press.
- Savoie, D. (2015). *What is government good at? A Canadian answer*. Montreal, QC: McGill-Queen's University Press.
- Schacter, M. (2016). Does “deliverology” matter? Retrieved from [http://docs.wixstatic.com/ugd/dadb01\\_67929f50ebaa46efa6e47df3526904a5.pdf](http://docs.wixstatic.com/ugd/dadb01_67929f50ebaa46efa6e47df3526904a5.pdf)
- Segsworth, R. V. (1990). Auditing and evaluation in the government of Canada: Some reflections. *Canadian Journal of Program Evaluation*, 5(1), 41–56.
- Shepherd, R. P. (2011). In search of a balanced Canadian federal evaluation function: Getting to relevance. *Canadian Journal of Program Evaluation*, 26(2), 1–45.
- Shulha, L. M., & Cousins, J. B. (1997). Evaluation use: Theory, research, and practice since 1986. *Evaluation Practice*, 18(3), 195–208. [https://doi.org/10.1016/S0886-1633\(97\)90027-1](https://doi.org/10.1016/S0886-1633(97)90027-1)
- Sutherland, S. (1986). The politics of audit: The federal Office of the Auditor General in comparative perspective. *Canadian Public Administration*, 29(1), 118–148. <https://doi.org/10.1111/j.1754-7121.1986.tb00397.x>
- Turpin, B. A. (2009). *An integrative evaluation approach in response to the Canadian evaluation policy of 2009: Enhancing federal public sector results-based management*. Unpublished manuscript.
- Young, G. (2006). Evaluation can cross the boundaries: The case of Transport Canada. *Canadian Journal of Program Evaluation*, 21(3), 73–92.
- Young, S. P. (2013). *Evidence-based policy-making in Canada*. Don Mills, ON: Oxford University Press.

## AUTHOR INFORMATION

**Robert P. Shepherd** is Associate Professor in the School of Public Policy and Administration at Carleton University. He is Supervisor of the graduate online Diploma in Public Policy and Program Evaluation. He is Past Chair of the Consortium of Universities for Evaluation Education, and the President of the Canadian Association of Programs in Public Administration. Dr. Shepherd joined the School in 2007, after serving in various federal public service and consulting company positions since 1987.