

# Linking Evaluation and Spending Reviews: Challenges and Prospects

---

**Jim McDavid and Astrid Brousselle**

*University of Victoria*

**Robert P. Shepherd**

*Carleton University*

**David Zussman**

*University of Ottawa*

The global financial crisis in 2008 was a significant watershed for governments everywhere. Diminished prospects for growth coupled with continuing demands for government interventions and chronic constraints on resources, prompted in part by the widespread adoption of variants of neo-liberalism (constrain resources to limit spending and shrink governments), have created fiscal environments where rationing expenditures among programs and policies is chronic and even acute.

One response to these situations is to systematically assess government expenditures, in effect, looking for savings by conducting broad-based spending reviews. The OECD has advocated for conducting spending reviews (OECD, 2011). Spending reviews can be used to address several objectives, including to tackle deficits and debt, to reallocate funds for better policy prioritization and efficacy, and to create fiscal space for new initiatives (Hawkesworth & Klepsvik, 2013; IMF, 2015; Moynihan & Beazley, 2016; OECD, 2011, 2016). In times of significant fiscal restraint, other possible objectives of spending reviews can be symbolic or justificatory. For example, they can be used as a communications tool, to support or bolster government legitimacy, to bring about line ministry fiscal discipline, or to publicly justify difficult budget decisions.

Our special issue of the *Canadian Journal of Program Evaluation* explores ways that evaluation can be linked to systematic, evidence-based assessments of government expenditures. We include articles that offer international, national (Canadian), and provincial (Quebec) perspectives. In this introduction, we summarize key themes in the articles offered by our contributors and, where appropriate, knit them together.

In the synthesis that follows the six articles in this volume, we reflect on key themes and offer an overall assessment of the prospects for linking evaluation and spending reviews.

Marc Robinson's article, "The Role of Evaluation in Spending Review" offers us an international overview that explores the current and prospective links between evaluations and spending review. He defines key terms, including what

spending review is: “. . . a process for systematically scrutinizing baseline expenditure to identify and implement savings measures.” Baseline expenditures are existing expenditures, so spending review is aimed at assessing what is currently being spent with a view to finding ways of reducing/reallocating. Reducing expenditures can occur via strategic cuts (doing less) or efficiency cuts (improving productivity and operating efficiency). Typically, governments make budget decisions that are intended to do both.

Robinson suggests six possible criteria for constructing the terms of reference for spending review: relevance, duplication, effectiveness, efficiency, equity, and market failure (the question of whether the private sector or other entities could provide particular programs). Among those, effectiveness is a key criterion, since it speaks to a core question in most program evaluations: Did the program or policy achieve (or contribute to) its intended results?

A significant point he makes is that the international financial crisis of 2008 has altered the fiscal picture for governments, globally. Demands for expenditures continue and arguably will grow as the impacts of climate change are factored into government programs and expenditures. But public resources are scarce, and this trend has accelerated since 2008. This pressure, with some exceptions, does not appear to be alleviated by deficit spending—that is a norm that is traceable to new public management and, more broadly, neo-liberalism internationally.

Robinson sees evaluation as a prospective contributor to the rational information bases that comprise the foundation for spending review. At the same time, he acknowledges that the record of actual uses of evaluation-related lines of information in spending review processes internationally is very limited. In our volume, this gap between expectations for evaluation and what has actually been accomplished is an important theme. Broadly, the gap between promise and performance is one that extends to all evaluation-related endeavours, including performance measurement (McDavid & Huse, 2012).

Robinson suggests six possible factors (some of which are closely related) that explain this gap between what evaluation could do for spending review and what has actually happened. Among those reasons are themes that are picked up in other articles in our volume: evaluations are focused on program improvement, are intended to help managers, and tend to make positive program improvement recommendations (Mayne’s article, among others, picks these themes up); evaluation functions in countries tend to be decentralized to spending departments (Shepherd, Mayne, Dobell and Zussman, and Bourgeois and Whynot all mention this issue); judgments about program effectiveness are rarely “yes” and “no” and are based on fallible lines of evidence (Dobell and Zussman, among other contributors, make this point); and the tendency for evaluation terms of reference to not give a lot of weight to questions focused on budget savings (Mayne, Shepherd, and Dobell and Zussman all comment on the effects of the decentralized institutional home for the Canadian federal evaluation function [evaluation units report to their respective deputy ministers] on its potential for delivering strategic evaluation-related information).

John Mayne's article, "Linking Evaluation to Expenditure Reviews: Neither Realistic nor a Good Idea," argues (picking up on a theme in Robinson's article) that not only have evaluations not contributed to spending reviews (Mayne calls them expenditure reviews), but, in the current federal government of Canada evaluation function, they should not. In Mayne's view, the existing program evaluation function in the federal government is generally well-positioned to contribute through periodic formative evaluations of programs to the ongoing program management-related dialogue in departments and agencies among program managers, evaluators, and even departmental executives (to whom evaluators report). The federal program evaluation function has persisted for over 40 years and although it has taken some significant knocks, as Mayne and other contributors to this volume acknowledge, it continues to be an important part of the analytical capacity of departments and agencies.

Mayne reprises some of the history of federal program evaluation, pointing out that in 1981 there was a clearly stated expectation that evaluations would "... produce credible, timely, useful, and objective findings on programs appropriate for resource allocation, program improvement and accountability" (Mayne, quoting the [Treasury Board Secretariat's \[1981, p. 3\] \*Guide on the Evaluation Function\*](#)). Evaluation would be both formative and summative and would be able to cover a wide range of expectations for the function.

As Mayne and others in this volume note, this range of expectations has not been fulfilled. Dobell and Zussman point out that evaluation's contributions to spending reviews in Canada starting with the 1984–1986 Nielsen task force have been desultory. But, unlike some other commentators, Mayne is clear that having the existing decentralized evaluation function is an asset—not for all the purposes adduced in 1981, but for formative assessments of programs with a view to improving their effectiveness and cost-effectiveness.

If evaluation is going to make a contribution to expenditure reviews, Mayne, like Robinson and Shepherd in this volume, suggests that a different evaluation function needs to be created that would be housed in a central agency (he suggests Treasury Board). This is not a new idea (Dobell and Zussman reference publications and reports that speak to creating an evaluator general for the federal government) to mirror the mandate of the auditor general and address the current asymmetry between evaluation and audit functions, wherein audit is both a decentralized (internal audit units) and a centralized function, whereas evaluation is not.

The article by Isabelle Bourgeois and Jane Whynot, "Strategic Evaluation Utilization in the Canadian Federal Government," is the first in our volume that examines evaluation use empirically in Canadian federal departments and connects uses with strategic decision-making. They have done a comparative case-study analysis that focuses on all program evaluation reports completed between 2010 and 2013 in two federal departments. Their two lines of evidence are a content analysis of the program evaluation reports and qualitative interviews with program managers and evaluators in both departments.

They acknowledge the importance of the literature on evaluation use—using evaluations is an enduring theme in both the scholarship and practice in the evaluation field—and point out that—aside from the now standard typology of four uses: instrumental (contributing to program-related decisions); conceptual (contributing the background knowledge of stakeholders with respect to a program or set of programs); symbolic (rationalizing program or policy-related decisions by pointing, post hoc, to particular evaluation findings or even the fact that an evaluation has been done); and process (the fact that an evaluation is being done influences program or even organizational behaviours)—there are strategic uses. They define strategic uses this way, “we refer to the application of evaluation results to broader, organizational-level decisions, which often include budgetary considerations.” Such uses could intersect with any of the other four types of uses—what distinguishes them is the focus on organizational-level influences of evaluation products and processes.

What they find from assessing the evaluation reports and from their interviews is that there are no evident strategic uses of the program evaluations produced between 2010 and 2013 in those two departments. The significance of this finding is that, during that time, there was a program review in 2010 (one of a series mandated annually by Treasury Board starting in 2007) and the Deficit Reduction Action Plan in 2012 (mandated by the Conservative government to identify four billion dollars in savings by 2015). Their findings suggest that those review processes did not make use of the stock of available program evaluations.

Unlike Robinson and Mayne in this volume, Bourgeois and Whynot suggest that it might be possible to work within the existing institutional structure for the evaluation function to get more purchase for strategic-level uses of program evaluations. They recommend focusing on portfolios of programs as the main unit of analysis to create more opportunities to compare performance (this has been done selectively since the 2009 *Evaluation Policy* was implemented); refocusing the core evaluation questions away from the mandatory “relevance and performance” set (these reappear in the directive that accompanies the 2016 Treasury Board *Policy on Results*) to better reflect the concerns of senior departmental decision-makers; and making the timing of evaluations more flexible to better reflect the information needs of decision-makers (this is possible given the risk-based approach to evaluation planning that is part of the 2016 *Policy on Results*).

Robert Shepherd’s article, “Expenditure Reviews and the Federal Experience: Program Evaluation and Its Contribution to Assurance Provision,” offers us a comprehensive public administrative overview of what he calls the assurance-related policies, instruments to implement those policies, and the roles that evaluation plays in each of five different forms of federal government-based assurance: policy and program coherence, internal program performance, government performance, systems improvement, and public accountability. At the same time that he is pointing to the different roles that evaluation has played or is expected to play across a wide range of purposes for assurance and a wide array of stakeholders, he is suggesting that federal evaluation has not lived up to its expectations.

In different words, evaluation, with its rich history of seeking relevance in a changing governmental and policy environment, has been saddled with expectations that are arguably incompatible. In his own assessment of the federal evaluation function (Shepherd, 2012), he pointed out that, given the (then) 2009 *Evaluation Policy* and its premium on accountability-related uses of program evaluations, the function was not delivering for key stakeholders. His observation then was that time was running out to align the evaluation function with government assurance-related expectations. Mayne, of course, would argue that the capacity of the function to resist these changing expectations while continuing to deliver formative evaluations that make a difference to program managers and other internal stakeholders is a sign of its enduring relevance.

Although Shepherd suggests (Table 1 in his article) that expenditure reviews intersect with “internal program performance” as a form of assurance, the focus of expenditure reviews on cross-program and cross-government assessments and comparisons does not sit comfortably with the existing program evaluation policy (in the 2016 *Policy on Results*, as Shepherd points out, there is no substantive mention of program effectiveness in the policy). Instead program evaluation is relegated to the accompanying directive, where the language is similar (relevance and performance) to that included in the 2009 *Evaluation Policy*.

What is to be done? Shepherd suggests two options. One is similar to what Mayne and Robinson have suggested—create and resource a central agency evaluation unit, the purpose of which would be to conduct expenditure review-related evaluations. That unit would, in effect, be expected to design and execute summative evaluations that are intended to meet the decision-making needs of stakeholders involved in budget-related reviews.

The second option is to create separate evaluation units within federal departments and give those units a mandate to conduct expenditure evaluations. Recalling the Bourgeois and Whynot article, these units would be focused on meeting the needs for strategic decision-making fora. If we look at this option for a moment, Shepherd suggests that building trust around the mission and operations of these units would be critical—in part because they would continue to rely on program managers and perhaps even their formative evaluation-focused colleagues for information. In effect, he appears to be signalling a need to reorient the evaluation-related cultures of departments and agencies.

Rod Dobell and David Zussman, “Sunshine, Scrutiny, and Spending Review in Canada, Trudeau to Trudeau: From Program Evaluation and Policy to Commitment and Results,” offer us an article that brackets the 50-plus years of Canadian federal evaluation policy and practice but goes well beyond that to include contemporary issues, such as freedom of information policy and its impacts on all the analytical (evidence-based) work that supports federal government decision-making. As well they include observations related to *deliverology*, as a current evaluation-related initiative of the Liberal government, and assess the prospects for program evaluation (and the evaluation function) generally, and, specifically, to make contributions to future expenditure reviews.

In 1981, Dobell and Zussman published an article in *Canadian Public Administration*, “An Evaluation System for Government: If Politics Is Theatre Then Evaluation Is (Mostly) Art” (Dobell & Zussman, 1981). That article pointed out that the (then) new program evaluation function, even at that time, was not meeting expectations. Mayne points out that in hindsight the expectations were not achievable given the institutional base of the program evaluation function.

A theme in their 1981 article that is taken up and expanded in their current article (this is their bottom line—without it other current and prospective analytical efforts will ultimately fail to make much of a difference) is the importance of “sunshine and scrutiny,” openness that requires access to information and the willingness of government (to modernize freedom of information legislation) to make available what is required to support competent evaluations and any other analytical work that assesses programs, policies, and operations. This theme is reflected among US evaluators—in 2008, Eleanor Chelimsky, (then) president of the American Evaluation Association, included these observations in her plenary address:

But we’re now seeing a metastasis of secrecy, far beyond intelligence and the military, into unrelated domains like environmental impacts, or hospital error rates, or drug side effects, or student test scores, where no national security interest can possibly be invoked. And we’re seeing extensive reclassification of materials that had already been declassified. This is a critical issue both for evaluation and for government, because it precludes the examination of all the facts, it inserts involuntary advocacy for agency programs and policies into “independent evaluations,” it makes a sham of executive accountability to the legislature, and it drastically distorts the public’s knowledge of what the government is doing. (p. 408)

Another theme in their earlier article that resonates in our current collection is the reality that program evaluations are not summative. Instead they are formative and, although they have uses in program adjustments within departments and agencies, they are not strategic, nor do they address questions that would be asked as cabinet decision-makers choose among programs and policies.

What are the prospects for program evaluations? Dobell and Zussman do not see evaluations having a role in spending review in the future. They even suggest (speculatively) that there may be a point where evaluations (similar to the way that medical diagnosis is going, for example) could be assisted with, if not conducted by, AIs.

Finally, in assessing deliverology, as promulgated by Sir Michael Barber and now the centrepiece of the 2016 federal *Policy on Results*, Dobell and Zussman are prepared to give this initiative more time. They point out that accompanying this focus on results is a focus on improving program and policy effectiveness via behavioural psychology and behavioural economics—several nudge units have been established to conduct experiments and quasi-experiments to improve delivery effectiveness. Whether these can deliver remains to be seen. There will be a point where this initiative should be evaluated.

Marie-Pier Marchand and Astrid Brousselle, “Regards sur l’expérience de la Commission de révision permanente des programmes au Québec,” contribute an article that focuses on Quebec and, more specifically, the recent experience, between 2014 and 2016, with the Commission de révision permanente des programmes (Commission for Permanent Program Review—CRPP) and the process undertaken by the Liberal government in Quebec to find ways to cut programs sufficient to balance the budget by 2015. Their article describes how this review commission was created and how it operated over its 2-year lifespan.

Quebec is one Canadian province that has recently implemented a government-wide evaluation function (as of 2013). The function is decentralized to administrative departments, and there is a general requirement to comply with Treasury Board evaluation policies including preparing a multi-year evaluation plan. As evaluation reports are completed, they are forwarded to Treasury Board.

Included among the lines of evidence/information sources used by the CRPP were completed program evaluations. It is not clear how those evaluations were used—it proved to be impossible to interview persons connected with the commission to get a better sense of how it operated, how it gathered information, even what sources of information were taken into account in assessing policy options and program cuts. Given the concerns expressed by Dobell and Zussman around access to information and the secrecy culture of governments more generally, this is noteworthy.

Marchand and Brousselle examined the reports that the commission produced, including their recommendations, to determine whether and to what extent the recommendations were reflected in subsequent budgets. The comparisons indicated very little correspondence between commission recommendations and the 2015 and 2016 budgets. The overall fiscal target was achieved (the Quebec government has balanced the budget in the last 2 years) but there is no direct (or even indirect) line from evaluations to budget decisions.

Looking ahead, the commission, before it was disbanded, recommended that the government create a permanent evaluation function that would be focused on conducting summative evaluations with a view to being able to recommend program revisions, including program cuts (presumably for those programs found wanting in some analytical sense). What is not clear at this time is whether this function will be centralized, or instead be layered onto the existing decentralized evaluation units—in effect repurposing them.

## REFERENCES

- Chelimsky, E. (2008). A clash of cultures: Improving the “fit” between evaluative independence and the political requirements of a democratic society. *American Journal of Evaluation*, 29(4), 400–415. <https://doi.org/10.1177/1098214008324465>
- Dobell, R., & Zussman, D. (1981). An evaluation system for government: If politics is theatre, then evaluation is (mostly) art. *Canadian Public Administration*, 24(3), 404–427. <https://doi.org/10.1111/j.1754-7121.1981.tb00341.x>

- Hawkesworth, I., & Klepsvik, K. (2013). Budgeting levers, strategic agility and the use of performance budgeting in 2011/12. *OECD Journal on Budgeting*, 13(1), 105–140. <https://doi.org/10.1787/budget-13-5k3ttg15bs31>
- IMF. (2015). *Republic of Slovenia: Technical assistance report—Establishing a spending review process* (Country Report No. 15/265). Retrieved from <https://www.imf.org/en/Publications/CR/Issues/2016/12/31/Republic-of-Slovenia-Technical-Assistance-Report-Establishing-a-Spending-Review-Process-43289>
- McDavid, J. C., & Huse, I. (2012). Legislator uses of public performance reports: Findings from a five-year study. *American Journal of Evaluation*, 33(1), 7–25. <https://doi.org/10.1177/1098214011405311>
- Moynihan, D., & Beazley, I. (Eds.). (2016). *Toward next-generation performance budgeting: Lessons from the experiences of seven reforming countries*. Washington, DC: World Bank. <https://doi.org/10.1596/978-1-4648-0954-5>.
- OECD. (2011, November). *Typology and implementation of spending reviews*. Discussion paper from the Senior Budget Officials Network on Performance and Results annual meeting, Paris, France.
- OECD. (2016). *2016 OECD performance budgeting survey: Integrating performance results and budgeting—Highlights*. Paris, France: OECD Publishing.
- Shepherd, R. (2012). In search of a balanced Canadian federal evaluation function: Getting to relevance. *Canadian Journal of Program Evaluation*, 26(2), 1–45.
- Treasury Board Secretariat. (1981). *Guide on the program evaluation function*. Ottawa, ON: Treasury Board of Canada, Office of the Comptroller General of Canada.