

IN SEARCH OF A BALANCED CANADIAN FEDERAL EVALUATION FUNCTION: GETTING BACK TO THE FUTURE

Jim McDavid
University of Victoria

██████████ Dobell and Zussman (1981), in their trenchant assessment of the federal evaluation function in Canada at that time, offered this view of the prospects for program evaluation in the federal government: “Clearly there has been no lack of good intention. But realization seems always to dangle tantalizingly in the future” (p. 405).

Robert Shepherd’s article in this issue of the *Canadian Journal of Program Evaluation* offers us an equally trenchant assessment of the evaluation function today. Shepherd’s article is not optimistic. In fact, he references Zussman (2010) when he says, “this time the buzz from the community is that this is likely the last chance evaluation will get to demonstrate its value to political decision-makers” (p. 3). When we read the article, we are reminded repeatedly that the evaluation function has not lived up to expectations. Shepherd suggests that over time the federal evaluation function “has gradually moved away from the strategic uses of evaluation for program effectiveness and political responsiveness to central agency concerns for essentially fiscal prudence and accountability” (p. 6).

Shepherd suggests in his article the need to rebalance the evaluation function so that the current emphasis (his view) on requirements to apply a narrow one-size-fits-all evaluation policy that requires cyclical evaluations of all programs is leavened with a more flexible approach that facilitates strategic evaluations that can be politically responsive. In different words, he wants to see evaluations being more relevant to (and presumably being used by) political decision-makers.

But when were program evaluations more relevant strategically? Dobell and Zussman (1981) are clear in their assessment—program

evaluations that were intended to be summative, that is, offer recommendations about the future of programs to political decision-makers, were not getting used that way. In fact, the federal evaluation policy of that time, although it made claims to being about summative program evaluations, was not meeting expectations. Shepherd implies that before centralization of the evaluation function began in 1977, things were different, but his article does not offer us an image of evaluation before it was institutionalized in the late 1970s.

What Dobell and Zussman advocated in their paper is instructive today. They did not believe that the evaluation function could deliver summative advice to political decision-makers. The nature of evaluation, grounded as it is in a rational view of administration and government, is different from political rationality. Analysis can inform the (political) policy-making process, but evaluation reports rarely play a role as an input into that process. Parenthetically, this “two worlds” view is supported by empirical research that has looked at whether political decision-makers use public performance reports (McDavid & Huse, 2012).

Dobell and Zussman (1981) advocated formative evaluations. They outlined an evaluation function that was aimed primarily at serving departmental priorities, for both managers and deputy heads. A robust evaluation function needed to be grounded in departmental priorities, addressing questions of relevance to managers and executives, and producing reports and recommendations that were intended to improve program performance.

Let us look at the federal evaluation function for a moment. The first thing we can say is that it has persisted since its inception in the late 1970s. This appears to be in spite of the odds. Shepherd chronicles the less than complimentary reports by Auditors General, the merger with the internal audit function in the 1990s, and the periodic reductions in the number of evaluation professionals and resources in departments and agencies. But evaluations are still being done, resources are still being allocated to the function, and in fact there is now a government-wide requirement to evaluate.

Why has the function survived? In my view, a key reason is that evaluation has continued to play the formative role that Dobell and Zussman sketched out in 1981. Evaluation may not be doing well from a strategic decision-making perspective, but it is well embedded in departments and agencies and, over time and many trials, has

established its credibility as a contributor to the (primarily) internal performance dialogue that happens across government.

What are we to make of the 2009 Evaluation Policy? Shepherd suggests that the scope is too narrow, “[using] a particular set of value criteria that supports central agency concerns for fiscal prudence and accountability” (p. 4). The policy focuses on evaluating the performance of programs (economy, efficiency, and effectiveness), and these criteria can easily be traced to performance auditing (Barzelay, 1997).

Shepherd suggests, very briefly, that “understanding the appropriateness of the program theory of change” (p. 34) is also important, but he does not explain what that would mean for evaluations. In my view, program appropriateness has some promise as an additional evaluation criterion. Although public auditors introduced appropriateness as one of their 12 dimensions of effectiveness (Canadian Comprehensive Audit Foundation, 1987), for evaluators, examining program appropriateness opens the door to comparing and assessing the means by which policy and program objectives are achieved. There is a rich and growing literature on policy instruments (Howlett, 2011) and, in the evaluation field, a burgeoning interest in program theories (Funnel & Rogers, 2011) and realist evaluation mechanisms (Sridharan & Nakaima, 2011) as ways of grappling with complexity in programs and policies. Assessing and comparing program appropriateness has some potential to produce some of the long-sought-after strategic advice to policy-makers.

The risk in the 2009 policy is that it goes too far in emphasizing the accountability face of program evaluation in the federal government. Shepherd chronicles the growth of accountability expectations for federal departments and agencies—from a program manager, executive, or evaluator perspective there is a veritable thicket of accountability requirements that must be navigated to get any work done. Finding a sustainable balance in the evaluation function principally involves program managers, department executives, Treasury Board, and federal evaluators. Forcing movement away from the working relationships between evaluators and program managers that are essential to doing credible evaluation work threatens to undermine the foundation of the whole enterprise. Federal program evaluators are internal evaluators (Volkov, 2011). They are not auditors, nor will they ever be. They have opportunities to work with their internal constituents to build evaluative cultures (Mayne, 2008; Mayne &

Rist, 2006), which in turn have some promise as drivers of improved performance.

It is well and good to have more independence for federal evaluators, but fundamentally evaluation is not and should not be about offering strategic advice to political decision-makers. Notwithstanding claims that this is what evaluation should do, it has rarely been successful in doing so. Further efforts to push the function in that direction may actually undermine the long-term viability of program evaluation in the federal government of Canada.

REFERENCES

- Barzelay, M. (1997). Central audit institutions and performance auditing: A comparative analysis of organizational strategies in the OECD. *Governance: An International Journal of Policy and Administration*, 10(3), 235–260.
- Canadian Comprehensive Audit Foundation. (1987). *Effectiveness reporting and auditing in the public sector*. Ottawa, ON: Author.
- 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.
- Funnell, S., & Rogers, P. (2011). *Purposeful program theory: Effective use of theories of change and logic models*. San Francisco, CA: Jossey-Bass.
- Howlett, M. (2011). *Designing public policies: Principles and instruments*. New York, NY: Routledge.
- Mayne, J. (2008, November). *Building an evaluative culture for effective evaluation and results management*. ILAC Brief 20. Amsterdam: Institutional Learning and Change Initiative. Retrieved from www.cgiar-ilac.org
- Mayne, J., & Rist, R. (2006). Studies are not enough: The necessary transformation of evaluation. *Canadian Journal of Program Evaluation*, 21(3), 93–120.
- McDavid, J., & Huse, I. (2012). Legislator uses of public performance reports: Findings from a five-year study. *American Journal of Evaluation*, 23(1), 7–25.

- Shepherd, R. P. (2012). In search of a balanced Canadian federal evaluation function: Getting to relevance. *Canadian Journal of Program Evaluation*, 26(2), 1–45.
- Sridharan, S., & Nakaima, A. (2011). Ten steps to making evaluation matter. *Evaluation and Program Planning*, 34, 135–146.
- Volkov, B. (2011). Internal evaluation a quarter-century later: A conversation with Arnold J. Love. In B. B. Volkov & M. E. Baron (Eds.), *Internal evaluation in the 21st century. New Directions for Evaluation*, 131 (Winter; pp. 5–12). San Francisco, CA: Jossey-Bass and the American Evaluation Association.
- Zussman, D. (2010). *Whatever happened to program evaluation? IT in Canada*. Retrieved from <http://www.itincanada.ca/index.php?cid=0&id=11513&np=8>

Jim McDavid is a Professor and Co-director of the Local Government Institute in the School of Public Administration, University of Victoria. He teaches and does research in performance management, program evaluation, and performance measurement in public sector and nonprofit organizations. He has lived and worked in Canada, the United States, and New Zealand, and his recent publications focus on legislator uses of public performance reports. The second edition of *Program Evaluation and Performance Measurement: An Introduction to Practice* will be published by Sage in October 2012.