

# Strategic Evaluation Utilization in the Canadian Federal Government

---

**Isabelle Bourgeois**

*École nationale d'administration publique*

**Jane Whynot**

*University of Ottawa*

**Abstract:** *Given the potential of the federal program evaluation function to inform decision-making at the highest levels of government, this project sought to investigate the nature and extent to which program evaluation findings are used as part of spending reviews and other reallocation exercises in selected government organizations. The multiple case study design used in this investigation included a qualitative content analysis of evaluation reports published between 2010 and 2013, as well as a series of key informant interviews conducted with evaluation staff and program managers. The findings show very little evidence of strategic evaluation utilization by organizational leaders. This is thought to be due to a few key factors: (a) the requirements of the 2009 Policy on Evaluation that was in effect at the time of the study; (b) the program-level focus of the evaluations; and (c) the public nature of the evaluation reports.*

**Keywords:** *Canadian federal government, evaluation utilization, organizational decision-making, organizational evaluation capacity*

**Résumé :** *Cet article a pour but d'examiner l'utilisation des évaluations fédérales lors des exercices de révision budgétaire et de réaffectation des ressources financières au gouvernement fédéral. Notre étude de cas multiples comprenait à la fois une analyse de contenu qualitative de rapports d'évaluation publiés entre 2010 et 2013 ainsi que des entretiens semi-dirigés auprès d'évaluateurs et de gestionnaires de programmes. Nos résultats démontrent qu'en général, les évaluations n'ont pas servi à la prise de décisions stratégiques par la haute direction. Nous croyons que trois éléments sont en cause : a) les exigences de la Politique sur l'évaluation (2009) en vigueur pendant notre étude ; b) la portée des évaluations fédérales, qui ont tendance à évaluer un seul programme ; et c) la diffusion publique des rapports d'évaluation.*

**Mots clés :** *gouvernement fédéral canadien, utilisation de l'évaluation, prise de décision organisationnelle, capacité d'évaluation organisationnelle*

**Corresponding author:** Isabelle Bourgeois, École nationale d'administration publique, 283 Alexandre-Taché, Gatineau, QC, J8X 3W7; [Isabelle.bourgeois@enap.ca](mailto:Isabelle.bourgeois@enap.ca)

## INTRODUCTION

In recent years, the government of Canada has implemented various exercises meant to reduce and manage its expenditures (i.e., Strategic Reviews, conducted in 2007–2010, and the Deficit Reduction Action Plan, conducted in 2012). Such measures aim to “reallocate funding from low-priority, low-performing programs to higher priorities” (Government of Canada, 2012). These initiatives have contributed, over time, to a complex reallocation system that relies on information from several different sources, including periodic program evaluations. This confers a different role to the evaluation function, which has traditionally been focused on providing managers with recommendations related to the improvement of program design and delivery. However, the *Policy on Evaluation* (Government of Canada, Treasury Board Secretariat, 2009), which highlights expectations and requirements related to the conduct and dissemination of evaluation in the federal government, clearly states that evaluation must also contribute to high-level, strategic decision-making.<sup>1</sup> The information contained in federal program evaluation reports is meant to be used at three levels: first, for strategic uses (including budgetary reallocation within and between departments); second, to improve program design and delivery; and third, to report on organizational accountability. Although evaluation utilization has been an important area of research since the 1980s, few studies have focused on how evaluation findings are actually used by Canadian government organizations, especially for funding reallocations (Office of the Auditor General of Canada, 2013). Given the potential of the program evaluation function to inform decision-making at the highest levels of government, and the considerable resources invested in evaluation (in 2012–2013, 459 full-time equivalents were dedicated to evaluation activities in the federal government, and expenditures related to evaluation totalled \$56.2 million; Government of Canada, Treasury Board Secretariat, 2015), this study sought to investigate the nature and extent to which program evaluation findings are currently used in government organizations.

## LITERATURE REVIEW AND RESEARCH PROBLEM

Program evaluation has been a core function of the Canadian federal government for over 40 years. Even though the contributions of evaluation to improvements in program design and delivery have long been acknowledged (e.g., see Patton, 2008), questions remain regarding its ability to inform strategic and budgetary decisions (Dumaine, 2012; McDavid & Huse, 2012). The importance of evidence gathered through evaluation, performance measurement, and other means in budgeting exercises, forecasting, and spending reviews is recognized internationally. Shand (1996) specifies that government budgeting efforts require “macro” objectives of fiscal consolidation, while also seeking to improve “micro” performance at the program level. To this end, he states that “many [OECD] countries have . . . taken steps to formalize requirements for program evaluation usually as part of reforms to the budgetary process” (p. 62).

Indeed, in the absence of effectiveness and efficiency information pertaining to specific policies or programs, senior decision makers find it difficult to justify spending cuts or changes in government priorities (Grizzle, 1987). However, decision-making is a complex undertaking, especially within the context of budget planning or spending reviews. Several factors and types of evidence mix with experience, stakeholder input, ideas, ideologies, and institutional structures; this means that evaluation is one component of the overall decision-making process, and its relative positioning against other sources of information has not been constant over time in the federal government (Aucoin, 2005; Heinrich, 2007). In an analysis of evaluation utilization within a policymaking context, Weiss (1988) specifies that evaluation is one piece of the political game, where decision makers are required to consider many different voices and positions. In many cases, evaluation may not even be foremost in decision makers' minds, but can influence them indirectly: "The major use of social science research is not the application of specific data to specific decisions. Rather, government decision makers tend to use research indirectly, as a source of ideas, information, and orientations to the world. Although the process is not easily discernable, over time it may have profound effects on policy" (Weiss, 1979). Contandriopoulos and Brousselle (2012) refer to "collective systems" situated at the organizational or policy level, which act as knowledge exchange mechanisms and include, among others, sense-making, coalition building, and persuasion and rhetoric. Evaluation, according to these authors, is one component of such systems, and evidence gathered as part of evaluations must be embedded within specific "action proposals" or policy options.

In their account of the role of program evaluation in government decision-making, Dobell and Zussman (1981) highlight three reasons why evaluations are not often considered by decision makers: first, evaluation as a function and a practice is not always clearly understood by senior managers and program staff—their expectations are sometimes misaligned with what evaluators are able to contribute. Second, the authors cite tacit resistance to evaluation on the part of program managers. More recent research on organizational evaluation capacity building confirms that organization-wide evaluation literacy tends to promote evaluation utilization (e.g., see Bourgeois & Cousins, 2013; Cousins & Bourgeois, 2014; Newcomer, 2007). Third, and most interestingly in our current context, "... the attempt to establish comprehensive evaluation systems, or even comprehensive evaluation studies, to serve cabinet and Parliament, failed to consider the true nature and information needs of the user" (Dobell & Zussman, 1981, p. 407). This is still relevant today and bears further examination, especially given the 2016 release of the *Policy on Results*. Dobell and Zussman clearly delineate the informational needs of program managers (such as program implementation issues, output production, early outcomes) from those of organizational decision makers (such as program impact, relevance, alignment with other programs and policies, value for money). This distinction is key to understanding evaluation use for budgetary decision-making.

From these analyses, it is possible to surmise that the Canadian evaluation function has not traditionally been positioned to influence budgetary decision-making and policy setting. Other countries, such as the Netherlands, have observed similar issues and have attempted to devise an evaluation system to mitigate the problem. [Schoch and den Broeder \(2013\)](#) describe the Dutch evaluation system, where evaluation is meant to support decision-making through different mechanisms: ex ante evaluations that take the shape of social cost–benefit analyses for projected programs, ex post policy reviews that focus on the effectiveness and efficiencies of policy suites, and spending reviews, which are conducted by the Ministry of Finance and focus on reallocation of financial resources. The latter include the findings of the ex ante and ex post studies conducted by the ministries themselves. Although the Dutch evaluation system takes into account the different types of users and evaluation uses, the authors specify that ex ante and ex post evaluations often do not contribute meaningfully to spending reviews, as may be the case in Canada. Based on these and the other papers reviewed here, we assume that evaluation results do not often find their way into either organizational budgetary decision-making or government-wide spending reviews in the federal government, or are cast aside in favour of other factors (e.g., political, social, anecdotal information). However, aside from audits of the evaluation function and internal work done by the Treasury Board Secretariat (such as the *Evaluation of the 2009 Policy on Evaluation* or the *Annual Health of the Evaluation Function* reports), very little external empirical research has been conducted on program evaluation utilization in the federal government; therefore, aside from anecdotal reports, we don't yet know if and how evaluation is actually used to make high-level, budgetary decisions. Furthermore, even though studies of evaluation utilization in other jurisdictions abound in the evaluation research literature, most of these studies focus on the perspectives of the evaluators (i.e., how evaluators perceive use by others) rather than on the perspectives of evaluation users (e.g., [Fleischer & Christie, 2009](#)). Our study therefore sought to address both of these issues and contribute empirical data on evaluation use, obtained from both evaluators and evaluation users.

## CONCEPTUAL FRAMEWORK

This study essentially focuses on the conditions or factors that support the utilization of program evaluation in organizations. This research domain is called *organizational evaluation capacity*, or EC, and has gained considerable attention in the last decade ([Cousins, Goh, Elliott, & Bourgeois, 2014](#)). EC has been linked to organizational learning and improvement, and is thought to be essential to sustainable evaluation utilization ([Alkin & Taut, 2002](#); [Torres & Preskill, 1999](#); [Preskill & Boyle, 2008](#)). The concept of EC, therefore, constitutes an important element of our study's theoretical base. Even though several theoretical and practical frameworks have been developed to describe the various components of organizational evaluation capacity, we elected to apply our own framework ([Bourgeois](#)

& Cousins, 2013) as the conceptual backdrop to our work; this framework was developed through an empirical study conducted with four federal government organizations and thus reflects the language, practices, and structures for evaluation found in these organizations, while other models of EC apply to other types of organizations. The framework, shown in its entirety by Bourgeois and Cousins (2013), essentially identifies six dimensions of EC: three of these dimensions illustrate an organization's *capacity to do* evaluation (Human Resources, Organizational Resources, and Evaluation Planning and Activities), while the other three dimensions focus on an organization's *capacity to use* evaluation (Evaluation Literacy, Organizational Decision-Making, and Learning Benefits). The three dimensions that are of particular conceptual interest for the present study are those related to organizational capacity to use evaluation: 1) *Evaluation Literacy* refers to the broad knowledge of evaluation across the organization; 2) *Organizational Decision-Making* focuses on the instrumental and conceptual application of evaluation findings in broader processes, such as program development or budget reallocation exercises; and 3) *Learning Benefits* focuses on applying knowledge derived from evaluations into everyday activities and organizational learning (including instrumental and process uses). It should be noted that most of the research conducted in the area of evaluation capacity focuses on evaluation's supply side (capacity to do evaluation), and little attention has been paid to its demand side (capacity to use evaluation): for example, models developed by Preskill and Boyle (2008) and Labin, Duffy, Meyers, Wandersman, and Lesesne (2012), as well as others, mainly identify practices and behaviours associated with organizational capacity to do evaluation rather than use (Bourgeois, 2016). This study, therefore, offers a unique perspective focused primarily on organizational capacity to use evaluation, and how this capacity effectively translates into concrete utilization.

In addition to the central concept of organizational evaluation capacity, the concept of evaluation utilization has been examined extensively in the literature (e.g., Cousins & Leithwood, 1993; K. Johnson et al., 2009) and, although we do not wish here to enumerate all possible uses of evaluation, this concept provides key insights for our study. As we have stated elsewhere (Bourgeois & Naré, 2015), there are four main types of evaluation use recognized in the majority of texts on the subject: *instrumental*, which refers to the direct application of evaluation recommendations; *conceptual*, which occurs when an evaluation influences stakeholders' understanding and attitudes about a program; *symbolic*, which occurs when individuals use evaluation information for political self-interest (R. B. Johnson, 1998); and *process use*, which is observed when behavioural and cognitive changes occur in evaluation participants by virtue of their proximity to the evaluation process (Alkin & Taut, 2002; Amo & Cousins, 2007; R. B. Johnson, 1998; Leviton, 2003; Patton, 2008). Although these four broad types of evaluation utilization are helpful in providing context for how evaluation findings can be integrated into various types of decisions, they do not highlight specifically the mechanisms by which evaluation contributes to high-level organizational

decision-making. Rather, all of these uses could apply to policymaking or decisions on budgeting and program allocations. Instrumental use, for instance, applies to both direct program improvements and budget reallocations—the scope and depth of evaluation findings, however, will vary accordingly. The same principle could be applied for the other three types of uses, as well as the often-cited *evaluation influence* term coined by Kirkhart (2000), which refers to an indirect application of evaluation akin to conceptual use. Therefore, while this typology is helpful in understanding how evaluation findings can be applied, we feel that a different conceptual lens is needed here, that of *strategic* use of evaluation. By strategic use, we refer to the application of evaluation results to broader, organizational-level decisions, which often include budgetary considerations. Our definition of strategic use therefore encompasses spending reviews but does not include individual program improvement. The previous *Policy on Evaluation* (Government of Canada, Treasury Board Secretariat, 2009) indirectly refers to strategic use by stating that deputy heads must apply the findings of program evaluations, but tends to focus on individual programs (rather than policy suites). Strategic use therefore seeks to relate to Weiss's conceptualization of evaluation use (Weiss, 1979), which focuses on the needs of policymakers and organizational leaders.

## RESEARCH QUESTIONS

Based on the EC framework described above, as well as the four types of uses outlined previously, the specific research questions that guided this study are the following: 1) Have program evaluation findings produced between 2010 and 2013 been used in the context of federal budgetary reallocation exercise (also called spending reviews)? 2) If so, how were the findings used? 3) Are program evaluation findings routinely applied to ongoing program design and delivery improvements? 4) What organizational systems and structures exist to support the use of program evaluation results?

## METHODOLOGY

The methodology for this study consisted of an exploratory case study conducted in two federal organizations.<sup>2</sup> These organizations were selected on the basis of prior research experiences with our team; both had participated in other studies on evaluation capacity building (ECB) and were identified as having an intermediate-level evaluation capacity across most dimensions on the Organizational Evaluation Capacity Self-Assessment Instrument (Bourgeois, Toews, Whynot, & Lamarche, 2013).

The study was conducted in two phases. The first was a content analysis undertaken of all published evaluation reports by both organizations between 2010 and 2013. This specific period was selected for two reasons: first, to allow sufficient time to have passed since the publication of each report to detect instances of

use for decision-making and reallocation at an organizational level; and second, because this period coincided with two federal spending reviews, the 2010 Strategic Review exercise and the 2012 Deficit Reduction Action Plan. Through these initiatives, federal departments and agencies were specifically asked to consider the findings of evaluations, as well as other sources of information, to reduce program spending and allocate resources where they were most likely to foster organizational outcomes. Therefore, we felt that the organizations studied would be most likely to show evidence of evaluation use in this time period, due to these additional reallocation initiatives. Seventeen reports were analyzed, representing 23 programs. The content analysis focused on whether the published reports reflected accepted evaluation practices in the Canadian federal government (i.e., quality and credibility are essential components of use), and whether they produced actionable recommendations. The qualitative data obtained in these reports were analyzed using both Microsoft Excel and NVivo. The detailed methods and findings of this first phase of the study are reported in [Bourgeois and Naré \(2015\)](#). The following section on findings highlights key results of the content analysis, to set the stage for the second phase of the study, reported here.

Based on the preliminary findings obtained through the report review, we organized semi-structured interviews with both the evaluation managers responsible for the 17 reports analyzed previously, as well as with program personnel (i.e., program managers and staff) who were the primary contacts for each of the evaluations. The interviews were meant to explore the evaluation lifecycle and highlight the ways in which each evaluation had contributed to higher-level decision-making in their respective organizations. We obtained a contact list through the evaluation directorate of each organization and proceeded to invite the potential interviewees through both email and telephone outreach. Individual consent forms were provided to each contact, and these outlined the project, inherent risks, and procedures. Out of 7 potential evaluators, 6 chose to participate in the study (the remaining evaluator is no longer with the organization and elected not to participate). Out of 17 potential program personnel, 6 also chose to participate in the study. Reasons cited for not participating include retirement, transition to the private sector, and transition within government. Three potential participants did not respond to multiple emails and phone calls.

Two interview guides were developed: one for evaluation managers, and one for program personnel. There was significant overlap between the two interview guides, to enable a level of comparison between the two groups. The interview guides were largely based on the EC and evaluation utilization conceptual frameworks and, as stated previously, generally followed the evaluation lifecycle. The interviews were meant to help us gather information about how the evaluations were conducted and who was involved at each step (planning, data collection, reporting). Specific questions were included on how the evaluation findings were used and at what stage of the evaluation they had been made available to stakeholders. Other questions focused on the involvement of senior decision makers throughout the evaluation process, as well as the overall evaluation capacity of each organization.

Each interview was recorded and transcribed by a team member. A research assistant then organized the interview data in a Microsoft Excel spreadsheet according to interviewee type (evaluator or program personnel) and question. Two team members were then charged with the analysis of the interview data by group (evaluators vs. program personnel) and by interview question. Once this first analytical phase was completed, the two analysts compared their summaries by question and by group—the themes identified through this first round of analysis were highly similar in both cases. The analysts then interpreted these findings together in light of the conceptual framework highlighted above, to produce the final study findings.

The most important limitation of this phase of the study was the low number of respondents on the program side. Clearly, our study would have benefitted greatly from the views of other program managers and staff. For this reason, we chose to combine the interview results from both organizations in our analysis, rather than treat them separately—we felt that combining the views of all program personnel interviewed provided a more thorough perspective of evaluation use than disaggregated data by organization. The one downside to this, however, is that we may have gotten a slightly different picture of use in one organization versus the other if we had kept the interview data separate. In addition, several program managers and staff were responsible for more than one program evaluated during our target period; therefore, some of their comments were more general in nature, and applied to all of the evaluations in which they had been involved, rather than the specific evaluations that occurred during our study's time frame. The other significant limitation is the exclusion of senior managers from the interview process. This was a deliberate choice on our part, based on our assumption that program staff and managers would be best positioned to know if and how the evaluation reports were used as part of spending reviews. Interviewees were indeed able to address this issue; however, we also wonder whether senior executives may have shed some light on the spending review processes undertaken and the general role of evaluation within these processes.

## FINDINGS

The content analysis conducted as part of this first phase enabled us to detect some common themes among the reports studied. Overall, the reports provide a clear overview of the evaluation questions and methods used to produce findings, as well as recommendations meant to confirm program renewal and improve design and delivery (but not reallocation). These are the key components of instrumental utilization: first, the evaluations are shown to be credible sources of information, based on the descriptions provided of the methods used and their limitations; in all cases, triangulation of data sources and methods provides support for the findings, as does the integration of key stakeholders to the evaluation process (e.g., program staff, program clients). In general, these reports provide useful information—the high rate of agreement found in the management responses



suggests that the recommendations are considered relevant and appropriate by program managers, and their implementation will lead directly to instrumental use (Bourgeois & Naré, 2015). Of course, the reports only tell part of the utilization story. The way in which results were shared with program managers and what happened once each evaluation was completed were explored through the semi-structured interviews. Our interview findings are presented in detail in the remainder of this section, organized by theme/research question.

### ***Involvement in Evaluation***

Respondents were asked to comment on their own involvement in the evaluation project to gauge whether key stakeholders were able to influence the evaluation design. All of the evaluations were conducted in part by external consultants (as reported in Bourgeois & Naré, 2015). In a few cases, the internal evaluators were also responsible for some lines of evidence (e.g., analysis of program administrative data). The main role of evaluators, therefore, was to oversee the evaluation project and liaise with the program representatives. In many cases, evaluation planning was done internally, before hiring external evaluators, and involved extensive consultations with key stakeholders. However, program representatives expressed some dissatisfaction with the nature of some of the evaluation questions retained. This issue is related to the application of the core issues required by the *Policy on Evaluation* (Government of Canada, Treasury Board Secretariat, 2009) and is addressed in more detail in a later section.

Evaluation advisory committees (EACs) were constituted for each study and generally included program managers and some staff, internal evaluators, the director of evaluation, and sometimes, higher-level managers (i.e., directors general). The role of the EAC is to provide advice and context to frame the evaluators' work, as well as to provide feedback on various deliverables (for instance, evaluation questions, evaluation framework, recommendations). Beyond their role on the EAC, program respondents also played a more involved role in the evaluations (often by sitting on an operational evaluation working group), and most had previously been exposed to evaluation. This required them to provide extensive feedback on the evaluation questions and the recommendations; both of these components often required nuancing and rewording to better reflect the program context but generally did not involve significant modifications.

### ***Ongoing Sharing of Results***

One way to foster utilization of evaluation findings is by sharing them throughout the evaluation lifecycle, rather than waiting until the production of the final report. Both sets of stakeholders agreed that some findings were provided to the EAC during the evaluation; however, in most cases these were presented as preliminary results once all data were collected. This was thought to provide a more balanced view of the program and the evaluation findings to stakeholders, even if it meant waiting until all data were analyzed. The evaluation results were mainly presented in meetings; evaluators felt that emails or other written

communications may compete with other priorities and therefore wouldn't be the best vehicle for preliminary findings, which often require contextualization and discussion. In some cases the data collection instruments and technical reports were also shared with a broader group of stakeholders. One key limitation identified by program personnel regarding the sharing of results is that they did not have dedicated time to reflect upon the emerging results and act on them.

Generally, emerging findings did not result in substantive actions taken by program personnel. According to these respondents, issues requiring a significant response or program change are identified well before an evaluation takes place and are dealt with accordingly. Most of the evaluations conducted in the time period selected did not reveal surprising findings that required immediate attention, but they did provide a mechanism through which program personnel could validate certain data or anecdotal impressions on the program.

Once the recommendations were drafted, however, they were often implemented before the approval and publication of the final evaluation report. In one case, the program was changed significantly during the evaluation, but according to the evaluator responsible for the study, this was mostly due to political factors. In another case, significant issues were raised regarding the availability and quality of performance measurement information, and changes were made to the program before the evaluation was completed.

### ***Credibility of Evaluation Process and Findings***

Overall, the evaluation process and findings were thought to be credible. The two organizations studied have a long history of evaluation and a stable complement of evaluation personnel. Program staff and managers understand the challenges faced by evaluators (in terms of scope, mandate, methods, data, etc.) and also understand the limitations of evaluation. They expressed confidence in the expertise of the internal evaluators to develop sound evaluation plans and to oversee the work of consultants. However, some respondents expressed some doubts as to the expertise of some consultants, as well as their experience in evaluating programs in the policy areas covered by both organizations. In some instances, the program personnel questioned the interpretation of some of the findings.

### ***Timeliness of Evaluation Reports***

Interview questions regarding the timeliness of evaluation reports obtained different responses from evaluators and program personnel. Evaluators stated that, in general, evaluation findings could be considered timely given the program lifecycle. However, they did state that no real decisions were identified before the start of the evaluation (which was conducted to meet the requirements of the *Policy on Evaluation*) and that, in several cases, the evaluations ran behind schedule and were completed later than anticipated. Program personnel, however, felt that the evaluations should be timed to coincide with Treasury Board submissions but that it wasn't currently the case. Others stated that the planned timing of the evaluations does not generally align well with management processes, and

explained that any information required outside of the five-year evaluation cycle is obtained through other means (e.g., commissioning studies or engaging with stakeholders directly).

### ***Utilization of Evaluation for Program Improvement***

The interviews revealed that evaluations are systematically used in both organizations for instrumental and conceptual purposes at the program level. They are used, notably, to validate existing program management concerns and to support incremental changes in program design and delivery. Some examples cited by respondents include improvements in processes, adjusted grant and contribution amounts, clarification of program theory, and learning about target audiences. These findings correspond to those of the first phase of the study. A few respondents also cited some process use, where program personnel involved closely in an evaluation then tended to ask evaluative questions and focus on program results, even after the end of the evaluation.

Overall, program respondents were in agreement with the final evaluation recommendations, which typically focus on instrumental uses. However, for various reasons, not all recommendations were implemented. For instance, in one case, a decision about the program was made outside of the evaluation process and had an impact on some of the recommendations. Interviewees also noted that the short time frame for the production of the management response (as per the *Policy on Evaluation*) affected the degree to which they could review the recommendations and make adjustments to their programs.

### ***Utilization of Evaluation for Budgetary/Strategic Decisions***

Our analysis reveals that evaluations were not used for budgetary reallocations conducted as part of spending reviews in the two organizations studied. In one instance, a radical change made to a program portfolio resulted from political influence, according to respondents, not evaluation findings. Where budgetary decisions were made concerning programs, these focused mainly on specific activities within programs (e.g., changing the amount of an available grant or contribution) rather than allocations between programs.

In some cases, the evaluations were used successfully to make cases internally or externally for increased funding or program expansion. However, opinions on this point were mixed—none of the stakeholders clearly stated that the evaluations were used for either Strategic Review or DRAP. Some respondents stated that evaluations do not provide enough between-program comparability to enable such decisions to be made. One other respondent felt that evaluations simply cannot provide sufficient information on program impacts to warrant budgetary reallocations on their own. In some cases, the evaluations provided data to justify some decisions already made (symbolic use).

To be more useful at a strategic level, respondents felt that evaluations should provide information on program and organizational expenditures, as well as how the programs respond to specific government priorities. They also felt that that the

informational needs of decision makers should first and foremost orient program evaluation planning, in terms of both timing and content.

A few other ways in which evaluations are used include the following: the production of program statistics, promotion, website, communications with external stakeholders, communications within government, contribution of data to a government-wide review of a specific policy area, program renewal, and the development of an organization-wide partnership strategy.

### ***Influencing Factors on Budgetary/Strategic Decisions***

When asked why they thought that evaluations had not been used for high-level decision-making within their organizations, respondents in both groups explained that, in many cases, the timing of the evaluations did not meet program needs (as noted previously). Another factor cited by respondents was the distance of senior management from the evaluation process. While senior managers were once involved more closely with evaluation activities, respondents said that they no longer sit on evaluation advisory committees. This means that they are no longer kept informed of evaluation findings on an ongoing basis, and tend to only see the final draft of the evaluation report before approval. Less contact with the evaluation process, according to some respondents, meant less discussion on the findings and inevitably led to non-use. However, despite the lack of involvement of senior leadership in the evaluation process, respondents felt that in cases where policy analysts were invited to sit on the EAC, the evaluation questions and findings were more closely aligned with the needs of senior decision makers. Interestingly, neither the evaluators nor the program managers identified the departmental evaluation committee as a significant intersection of evaluation and decision-making.

Two important observations were made by respondents concerning the level of analysis of most federal evaluations. First, the core issues required by the *Policy on Evaluation* (Government of Canada, Treasury Board Secretariat, 2009) seem to preclude the development of evaluation questions that are of direct interest to program personnel. Coupled with the policy's coverage requirement (100% of direct program spending must be evaluated every 5 years), both evaluators and program personnel felt that their respective organizations did not have sufficient evaluation resources to conduct studies that would both meet Treasury Board requirements and internal decision-making interests: "Accountability and utility can be at loggerheads." In addition, program personnel especially felt that the core issues result in a retrospective assessment of the program only; although they acknowledged that this could be helpful in better understanding what works well, they would appreciate advice on prospective investments and reallocations from evaluators. Second, respondents felt that focusing on individual programs as a unit of analysis provided useful information to direct program-level changes and improvements, but did not allow for a cross-sectional or portfolio analysis of programs. Respondents mentioned that this higher-level focus would provide a more strategic view to senior managers, and could then influence decisions regarding the reallocation of resources across programs. It should be noted, however, that the policy does

not specifically preclude portfolio-wide evaluations. The comments made by respondents regarding the constraining nature of the policy may, therefore, stem more from internal interpretations of the policy than its actual content.

When asked if evaluation could feed into a high-level decision-making process, evaluators felt that it might be possible but that evaluations would have to focus on prospective decision-making rather than just a retrospective assessment. This could be done, for example, by including environmental scanning in the evaluation process or by including specific questions that focus on what decisions will be required in the near future. Evaluators also felt that meta-analyses conducted across programs could also likely be useful and could fall within the purview of their directorate. Program personnel, however, felt that evaluations could generally not feed into strategic decisions, due to their timing (i.e., the set 5-year evaluation cycle required by the policy) and their program-specific focus. Given that, in both of these organizations, programs benefit from stable funding and respond to an ongoing need, evaluations that move beyond the core issues and toward portfolio-wide reviews could be envisaged. Program personnel also felt that, at a minimum, evaluation could potentially be used as one input to a broader decision-making process by identifying poor or underperforming programs.

### ***Policy Environment for Evaluation Use***

As mentioned above, the *Policy on Evaluation* ([Government of Canada, Treasury Board Secretariat, 2009](#)) tends to be perceived as a hindrance, rather than a support, to strategic evaluation use. Evaluation is viewed by program personnel as a cost of doing business under the auspices of accountability rather than as a management support function. Program personnel felt that the evaluation advisory committees on which they generally sit target only the program under study and do not provide the needed space to discuss more strategic issues. However, they felt that the departmental evaluation committee might enable a broader conversation about program design and outcomes, and thus contribute to more strategic decision-making in the organization, even though they were uncertain as to whether this was currently the case.

In addition, the publication requirements outlined in the policy were felt to hinder the development of action plans (evaluation reports must be published on the departmental website no later than 120 days after approval and must include a management response and action plan) based on in-depth discussions with senior management and stakeholders. Although both sets of respondents described the recommendation development process as highly collaborative and iterative, the structures and processes of each organization limit the ability of evaluators to engage with senior management about foundational issues related to program expenditures and reallocations.

### ***Organizational Evaluation Capacity***

Given that organizational evaluation capacity is known to be strongly connected to all types of evaluation use (e.g., [Bourgeois & Cousins, 2013](#)), respondents

were asked to comment on the overall knowledge of program staff and managers regarding evaluation. In general, respondents in both organizations felt that organizational knowledge about evaluation is high, particularly among program managers. Contributing factors highlighted by respondents include strong evaluation leadership, expertise/competencies of evaluators, longevity, and the organizational climate which strongly supports the production of evidence. Evaluations are generally perceived to be credible and reliable, as stated earlier, and evaluations that employed a particular methodology requiring high levels of expertise (for instance, cost-benefit analysis) were lauded as good practice by respondents. However, evaluation resourcing continues to be problematic in both organizations and influences the ability of the evaluators to produce evaluations that meet both Treasury Board requirements and the informational needs of program managers. Furthermore, some respondents expressed concerns with the fact that evaluators have never managed programs themselves and thus may not have a firm grasp on the programs that they evaluate. The time and workload required to conduct evaluations and provide feedback or act on recommendations were also thought to exceed the benefits of evaluation (one respondent used the “love/hate” analogy to represent the relationship with evaluation).

### ***Factors That Facilitate and Inhibit Evaluation Use***

Taking the previous findings into account, respondents were asked to highlight the factors that facilitate and inhibit evaluation use (including strategic use). Factors thought to encourage use include relationship building between evaluators and program managers through multiple projects, the use of a collaborative approach, sharing information throughout the evaluation, building staff capacity in evaluation (e.g., through secondments), and a stable cadre of program and evaluation directors. Factors thought to inhibit evaluation use include the time required to conduct evaluation and disseminate findings, the evaluation “jargon” used in the reports, the type of reporting used, staff turnover during evaluation projects, issues related to performance measurement data, the generally positive tone of the findings, perceived bias in terms of evaluation participants, risk aversion due to the *Access to Information Policy*, and evaluation staff being pulled into other initiatives and activities during a project. Evaluators felt that more could be done to foster use once an evaluation is completed (lunch and learn presentations, evaluation findings summaries and brochures, etc.) but cautioned that, currently, project workloads do not lend themselves to any additional commitments.

## **DISCUSSION**

Overall, our findings point to the fact that evaluations are not used for strategic purposes in the two organizations studied, despite a pan-governmental climate that encouraged such uses (i.e., requirements associated with the 2010 Strategic Review Exercise and the 2012 DRAP). From a capacity perspective, our findings show that, although the essential precursors to use are present in both

organizations (e.g., stable evaluation and program leaders, widespread knowledge of evaluation among program managers, credible evaluation processes), the actual content of the 17 evaluations reviewed does not lend itself to strategic use.

Recommendations focus largely on incremental program changes meant to improve implementation and early outcomes, and cross-program comparisons or meta-analyses do not seem possible based on the information collected. So, although both organizations had scored relatively high in previous studies on the Organizational Evaluation Capacity Self-Assessment Instrument (Bourgeois et al., 2013), it appears as though there is still some room for improvement in terms of crafting evaluations that will contribute to spending reviews in both organizations. These findings are somewhat different from those of the *Evaluation of the Policy on Evaluation*, conducted by the Treasury Board Secretariat in 2015. The evaluation found that most evaluation studies meet the strategic and decision-making needs of senior administrators; however, these results mainly stem from case studies selected by departments to demonstrate actual evaluation use and may not be representative of all evaluations conducted over the life of the policy. They also do not directly address the use of evaluation in spending reviews. Nonetheless, some of the stakeholder consultations held as part of that evaluation with senior administrators reveal that evaluations are thought generally to be useful. Therefore, further systematic investigations of stakeholder views on utilization may be warranted at this point to identify specific instances of strategic uses of evaluation and learn from them. Most notably, consultations with senior executives on the specific issue of spending reviews might highlight the ways in which evaluation can best serve their needs.

The role of evaluation in government decision-making, as highlighted in the literature review, varies depending on several factors. Governments do not always consider evidence in policymaking and budgetary reallocation, and evaluation is not generally seen as the best source of evidence for spending decisions (Zussman, 2003). To foster use, evaluators must consider the needs of potential users of evaluation and the context within which decisions are made in the organization (Mohan & Sullivan, 2006). Decision makers also have an important role to play in facilitating the use of evaluation. According to Aucoin (2005), they need to create an environment in which public servants are encouraged to provide evidence of program success or failures; they need to become “consumers of research”; and, among other things, they need to ask for more systematic and rigorous evaluations to ensure that they have access to the best possible information on which to base sometimes difficult decisions. To do this, however, evaluation may need to shift its priorities, or at least attempt to meet the needs of two different audiences: program managers and senior decision makers. By eschewing the distinction between formative and summative evaluation, we believe that the most recent iteration of the policy failed to recognize the needs of different evaluation users. Further, by establishing the program as the main unit of analysis, and requiring 100% coverage of all direct program spending in a 5-year cycle, the policy failed to address how evaluations could be used more strategically by senior managers.

Three main recommendations stem from our work. First, if evaluations are to be used at a strategic level, the unit of analysis must shift from individual programs to suites or portfolios, to provide senior managers with a bird's eye view of program performance in a given domain. This would better enable strategic decisions to be made across programs rather than focus more narrowly on individual programs. Second, evaluations should directly address the specific questions and needs of senior decision makers, rather than focus exclusively on Treasury Board Secretariat core issues. Although in theory evaluation questions can be geared toward both purposes, in practice this does not seem possible in the two organizations studied. Therefore, a greater consideration of upcoming decisions or program realignment must be integrated into evaluation plans. And third, as is the case elsewhere, the timeliness of evaluations should match organizational demands for information. This is certainly difficult to achieve, and evaluation managers already conduct a considerable amount of forecasting as part of the quinquennial evaluation planning process, but more responsive approaches to evaluation and reporting may enable the two organizations to conduct evaluations with higher utilization potential. One potential support here may be the integration of performance measurement data in evaluations; although this has been recommended for quite some time (Mohan, Tikoo, Capela, & Bernstein, 2006; and more recently, in the *Evaluation of the Policy on Evaluation* [Government of Canada, Treasury Board Secretariat, 2015]), in practice this has been a challenge for federal departments and agencies. The 2016 *Policy on Results* addresses these concerns specifically with new requirements related to the creation of a head of performance measurement and the integration of performance measurement into departmental performance measurement and evaluation committees; although we have arrived at these recommendations through our empirical analysis, these have been voiced by many stakeholders in the federal evaluation community in recent years.

## CONCLUSION

This study sought to examine the extent to which evaluations are used in two federal government organizations. The literature in this area points to the fact that evaluations are not generally used in the context of spending reviews and other strategic decisions; however, many of the articles published on this topic are dated, and there are very few recent empirical studies examining this question. Given the recent spending reviews undertaken in the federal government, we felt that this question should be studied empirically through a qualitative content analysis and series of semi-structured interviews. We especially felt that, unlike in many studies on evaluation use, program managers or evaluation users should be directly asked about utilization. Our findings are generally consistent with those of previous studies and reveal that, although instrumental and conceptual use tends to be quite high at a program level, there is little or no strategic use of evaluation beyond program improvement, and that evaluation is not used



to inform spending reviews. Various factors influence this non-use: the unit of analysis, timeliness of the evaluation, and the types of questions asked were all found to be particularly critical in inhibiting use. Further evaluation capacity building in terms of evaluation literacy and the application of evaluation to organizational decision-making may help the two organizations studied move beyond the implementation of instrumental, program-level evaluation findings, toward a higher-level, strategic use.

Our results also highlight the need for more research in this area, especially in light of the findings of the *Evaluation of the 2009 Policy on Evaluation*, which as mentioned previously were positive about strategic uses of evaluation. Next steps, for our team, will include the development of a government-wide survey focusing specifically on strategic uses of evaluation and organizational evaluation capacity.

## NOTES

- 1 The *Policy on Evaluation* (Government of Canada, Treasury Board Secretariat, 2009) was replaced by the *Policy on Results* in July 2016. Although many of the changes made to the newest iteration of the policy apply to strategic evaluation use, the policy was not yet in place at the time of our data collection and analysis, and, therefore, we will mainly refer to the 2009 policy in this paper.
- 2 Both organizations requested anonymity; therefore, all possible identifying information has been removed from this paper.

## ACKNOWLEDGEMENTS

The authors are grateful to the Social Sciences and Humanities Research Council for funding this project (ref #430-2014-1091).

## REFERENCES

- Alkin, M. C., & Taut, S. M. (2002). Unbundling evaluation use. *Studies in Educational Evaluation*, 29(1), 1–12. [https://doi.org/10.1016/S0191-491X\(03\)90001-0](https://doi.org/10.1016/S0191-491X(03)90001-0)
- Amo, C., & Cousins, J. B. (2007). Going through the process: An examination of the operationalization of process use in empirical research on evaluation. *New Directions for Evaluation*, 2007(116), 5–26. <https://doi.org/10.1002/ev.240>
- 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>
- Bourgeois, I. (2016). Performance measurement as precursor to organizational evaluation capacity building. *Evaluation Journal of Australasia*, 16(1), 11–18.
- Bourgeois, I., & Cousins, J. B. (2013). Understanding dimensions of organizational evaluation capacity. *American Journal of Evaluation*, 34(3), 299–319. <https://doi.org/10.1177/1098214013477235>

- Bourgeois, I., & Naré, C. (2015). The “usability” of evaluation reports: A precursor to evaluation use in government organizations. *Journal of Multidisciplinary Evaluation*, 11(25), 60–67.
- Bourgeois, I., Toews, E., Whynot, J., & Lamarche, M. K. (2013). Measuring organizational evaluation capacity in the Canadian federal government. *Canadian Journal of Program Evaluation*, 28(2), 1–19.
- Contandriopoulos, D., & Brousselle, A. (2012). Evaluation models and evaluation use. *Evaluation*, 18(1), 61–77. <https://doi.org/10.1177/1356389011430371>
- Cousins, J. B., & Bourgeois, I. (2014). Cross-case analysis and implications for research, theory, and practice. *New Directions for Evaluation*, 2014(141), 101–119.
- Cousins, J. B., Goh, S. C., Elliott, C. J., & Bourgeois, I. (2014). Framing the capacity to do and use evaluation. *New Directions for Evaluation*, 2014(141), 7–23.
- Cousins, J. B., & Leithwood, K. A. (1993). Enhancing knowledge utilization as a strategy for school improvement. *Science Communication*, 14(3), 305–333. <https://doi.org/10.1177/107554709301400303>
- 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>
- 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>
- Fleischer, D. N., & Christie, C. A. (2009). Evaluation use: Results from a survey of U.S. American Evaluation Association members. *American Journal of Evaluation*, 30(2), 158–175. <https://doi.org/10.1177/1098214008331009>
- Government of Canada. (2012). *Budget 2012*. Retrieved from <http://www.budget.gc.ca/2012/home-accueil-eng.html>
- Government of Canada, Treasury Board Secretariat. (2009). *Policy on evaluation*. Retrieved from <http://www.tbs-sct.gc.ca/pol/doc-fra.aspx?section=text&id=15024>
- Government of Canada, Treasury Board Secretariat. (2015). *Evaluation of the 2009 Policy on Evaluation*. Retrieved from <https://www.tbs-sct.gc.ca/hgw-cgf/oversight-surveillance/ae-ve/cee/orp/2015/e09poe-epse09-eng.asp>
- Grizzle, G. A. (1987). Linking performance to funding decisions: What is the budgeter’s role? *Public Productivity Review*, 10(3), 33–44. <https://doi.org/10.2307/3380085>
- Heinrich, C. J. (2007). Evidence-based policy and performance management: Challenges and prospects in two parallel movements. *American Review of Public Administration*, 37(3), 255–277. <https://doi.org/10.1177/0275074007301957>
- Johnson, K., Greenesid, L. O., Toal, S. A., King, J. A., Lawrenz, F., & Volkov, B. (2009). Research on evaluation use: A review of the empirical literature from 1986 to 2005. *American Journal of Evaluation*, 30(3), 377–410. <https://doi.org/10.1177/1098214009341660>
- Johnson, R. B. (1998). Toward a theoretical model of evaluation utilization. *Evaluation and Program Planning*, 21(1), 93–110. [https://doi.org/10.1016/S0149-7189\(97\)00048-7](https://doi.org/10.1016/S0149-7189(97)00048-7)
- Kirkhart, K. E. (2000). Reconceptualizing evaluation use: An integrated theory of influence. *New Directions for Evaluation*, 2000(88), 5–23.

- Labin, S., Duffy, J. L., Meyers, D. C., Wandersman, A., & Lesesne, C. A. (2012). A research synthesis of the evaluation capacity building literature. *American Journal of Evaluation*, 33(3), 307–338. <https://doi.org/10.1177/1098214011434608>
- Leviton, L. C. (2003). Evaluation use: Advances, challenges and applications. *American Journal of Evaluation*, 24(4), 525–535. <https://doi.org/10.1177/109821400302400410>
- 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>
- Mohan, R., & Sullivan, K. (2006). Managing the politics of evaluation to achieve impact. *New Directions for Evaluation*, 2006(112), 7–23. <https://doi.org/10.1002/ev.204>
- Mohan, R., Tikoo, M., Capela, S., & Bernstein, D. J. (2006). Increasing evaluation use among policymakers through performance measurement. *New Directions for Evaluation*, 2006(112), 89–97. <https://doi.org/10.1002/ev.210>
- Newcomer, K. E. (2007). How does program performance assessment affect program management in the federal government? *Public Performance and Management Review*, 30(3), 332–350. <https://doi.org/10.2753/PMR1530-9576300302>
- Office of the Auditor General of Canada. (2013). Chapter 1 – Status report on evaluating the effectiveness of programs. In *Spring report of the Auditor General of Canada*. Retrieved from [http://www.oag-bvg.gc.ca/internet/English/parl\\_oag\\_201304\\_01\\_e\\_38186.html#hd5m](http://www.oag-bvg.gc.ca/internet/English/parl_oag_201304_01_e_38186.html#hd5m)
- Patton, M. Q. (2008). *Utilization-focused evaluation* (4th ed.). Thousand Oaks, CA: Sage.
- Preskill, H., & Boyle, S. (2008). A multidisciplinary model of evaluation capacity building. *American Journal of Evaluation*, 29(4), 443–459. <https://doi.org/10.1177/1098214008324182>
- Schoch, M., & den Broeder, C. (2013). Linking information on policy effectiveness and efficiency to budget decisions in the Netherlands. *OECD Journal on Budgeting*, 12(3), 1–22. <https://doi.org/10.1787/budget-12-5k468nqj2n28>
- Shand, D. (1996). Budgetary reforms in OECD member countries. *Journal of Public Budgeting, Accounting and Financial Management*, 10(1), 63–88. <https://doi.org/10.1108/JPAFM-10-01-1998-B003>
- Torres, R. T., & Preskill, H. (1999). Ethical dimensions of stakeholder participation and evaluation use. *New Directions for Evaluation*, 1999(82), 57–66. <https://doi.org/10.1002/ev.1137>
- Weiss, C. H. (1979). The many meanings of research utilization. *Public Administration Review*, 39(5), 426–431. <https://doi.org/10.2307/3109916>
- Weiss, C. H. (1988). Evaluation for decisions: Is anybody there? Does anybody care? *American Journal of Evaluation*, 9(1), 5–19. <https://doi.org/10.1177/109821408800900101>
- Zussman, D. (2003). Evidence-based policy making: Some observations of recent Canadian experience. *Social Policy Journal of New Zealand*, 20, 64–71.

## AUTHOR INFORMATION

**Isabelle Bourgeois**, PhD, is Associate Professor at l'École nationale d'administration publique (ENAP) in Gatineau, Québec. She teaches program evaluation and research methods and her research work focuses mainly on organizational evaluation capacity building and evaluation utilization. Prior to starting her academic career, Isabelle worked for more than 10 years as an evaluation manager in two federal government organizations. In 2017, she received the Karl Boudreault Award for Leadership in Evaluation from the National Capital Chapter of the Canadian Evaluation Society. She is currently the Editor of the Canadian Journal of Program Evaluation.

**Jane Whynot** has been a consultant providing performance expertise to the government, and voluntary sectors for approximately 2 decades. She currently serves as the Past President of the Canadian Evaluation Society's National Capital Chapter (CES-NCC), is an instructor in Carleton University's Graduate Certificate Program in Evaluation, and is currently a 4th year PhD candidate at the University of Ottawa with a research interest in exploring gender-based analysis in the federal government evaluation function.