Journal of MultiDisciplinary Evaluation
Volume 11, Issue 25, 2015

ISSN 1556-8180
http://www.jmde.com

The "Usability" of Evaluation Reports: A Precursor to Evaluation Use in Government Organizations

Isabelle Bourgeois
École nationale d'administration publique

Clémence Naré
Ecole nationale d'administration publique

Research funded by SSHRC #430-2014-01091

Background: According to the Treasury Board of Canada's *Policy on Evaluation* (2009), evaluations produced by federal government departments must contribute to decision-making at an organizational level (mainly summative) as well as a program level (mainly formative). Previous research shows that although the formative objectives of evaluation are generally reached, the use of evaluation for broader, budgetary management is limited. However, little research has been conducted thus far on this issue.

Purpose: This study investigates the extent to which program evaluation is used in the Canadian federal government for budgetary management purposes.

Setting: This paper outlines the results obtained following the first component of a two-pronged research strategy focusing on evaluation use in Canadian federal government organizations.

Intervention: N/A

Research Design: Two federal agencies were recruited to participate in organizational case studies aiming to identify the factors that facilitate the use of evaluation for budgetary reallocation exercises.

Data Collection and Analysis: This report presents the findings from a detailed analysis of evaluation reports published by both agencies between 2010-2013. The data were collected from public evaluation reports and analyzed using NVivo.

Findings: The preliminary findings of the study show that instrumental use has occurred or can be expected to occur, based on the types of recommendations outlined in the reports reviewed and on the responses to the evaluations produced by program managers.

Keywords: *evaluation use; organizational evaluation capacity; instrumental use; evaluation reports; Canadian federal government; document review; organizational decision-making.*

Introduction

In the Canadian federal government, the program evaluation function is responsible for the development and implementation of studies meant to assess the relevance and performance of policies and programs. This function is decentralized across departments and agencies. Each departmental evaluation unit is responsible for coordinating external evaluations or conducting them internally, depending on available resources and expertise. Each unit is accountable to the senior administrator of the organization, but must follow the government-wide Policy on Evaluation (TBS, 2009) that outlines specific requirements regarding program coverage, timing, and the main issues to be addressed in evaluation studies. According to the Policy, the evidence gathered through program evaluations is meant to be used by various stakeholders in a number of ways; for instance, evaluation data can be used to inform government-wide exercises focused on budgetary reallocations between programs and organizations; it can also be used to improve program design and delivery, or to show accountability to politicians and the public.

Even though the program evaluation function has been in existence in the federal government since the 1970s, there remain questions as to the actual use of program evaluation for these three purposes. Even though there is some agreement as to the usefulness of evaluations towards program improvement, the Auditor General of Canada (2013) has found that the internal capacity of individual departments and agencies to understand program evaluation and apply its findings towards program improvement is generally insufficient across government. Similar findings have been identified in empirical studies focusing on the use of performance measurement, which is closely linked to program evaluation (McDavid and Huse, 2012). Evidence that evaluation is used to inform government-wide exercises focused on budgetary reallocations, such as Strategic Reviews, is even more difficult to find (Dumaine, 2012). In other words, it seems as though program evaluation, despite its history and prominence, does not typically provide government decision-makers (e.g., primary evaluation users) with the required information to improve programs, or to reallocate funds between and within government departments and agencies. However, aside from audits of the federal evaluation function and self-reported departmental data, very little empirical research has been conducted on program evaluation use in the Canadian federal government; therefore, we don’t yet know to what extent evaluation is actually used in government, and how it is used. In addition, even though studies of evaluation utilization in other jurisdictions abound in the 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 the evaluation users themselves (e.g., Fleischer and Christie, 2009).

The ongoing study reported here seeks to fill this knowledge gap by studying the extent to which program evaluation results produced between 2010-2013 by two anonymous Canadian government organizations have been used for accountability, program improvement and budget reallocation. This study considered program evaluation use from the perspective of evaluation users in order to better understand their needs through analysis. This paper highlights the findings from the first component of the study, a review of published evaluation reports, in an attempt to assess the usability of the evaluation reports produced by the two participating organizations.

Conceptual Framework and Research Questions

The concept of “evaluation use”, defined generally as the utilization of program evaluation findings to make strategic organizational decisions or to improve programs, is a key focus of the program evaluation research literature, as demonstrated through the large number of theoretical and empirical studies published on this topic (Cousins and Leithwood; 1986, and Johnson and associates; 2009, have produced significant reviews of the literature on evaluation use and together, highlight 106 empirical studies on this topic published since 1986). These studies have enabled researchers and practitioners to explore the various ways in which evaluation can be used in organizations. Four essential types of evaluation use have been identified in the literature, the first three being associated with the use of findings: *instrumental use*, which refers to using evaluation findings as a basis for action and change; *conceptual use*, which occurs when an evaluation influences stakeholders’ understanding and attitudes about a program; and *symbolic use*, which occurs when individuals use evaluation information for political self-interest (Johnson, 1998). In addition to these three types of evaluation use, researchers have also identified *process use*, which is observed when behavioral and cognitive changes occur in evaluation participants by virtue of their proximity to it. Process use moves beyond the use of findings and involves learning to apply evaluative logic and thinking to a broader program and organizational context (Alkin and Taut, 2002; Amo and Cousins, 2007; Johnson, 1998; Leviton, 2003; Patton, 1997).

Based on the four essential types of evaluation use derived from the literature, the specific research questions that guide this study are the following: 1) Have program evaluation findings produced between 2010 and 2013 been used in the context of budgetary reallocation exercises undertaken by the two participating organizations? 2) If so, how were the findings used? 3) Are program evaluation findings routinely applied to ongoing program design and delivery improvements within these organizations? 4) What organizational systems and structures exist within these organizations to support the use of program evaluation results?

Methodology

Participants

This organizational case study features two participating organizations[1]. These organizations were selected on the basis of prior research experiences with the research 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 Bourgeois et al.'s (2013) Organizational Evaluation Capacity Self-Assessment Instrument. In order to assess the usability of published evaluation reports, we first had to ensure that the participating organizations had produced a sufficient number of reports to enable an in-depth content analysis, and that these reports reflected commonly accepted evaluation practices in the Canadian federal government. We determined that the two participating organizations were appropriate for this study based on the application of the Organizational Self-Assessment Instrument, which touches on these specific aspects of capacity.

Both participating organizations are relatively small compared to other large government departments, and have a long-standing history of program evaluation due to the nature of their respective programs. Program managers and their staff are exposed regularly to evaluation and have developed considerable knowledge of evaluation practices and methods over the years. We felt, therefore, that these organizations would be in a position to demonstrate different types of evaluation uses, if and when use occurs.

Procedures

The research activities undertaken in this first phase of the study involved a review of all program evaluation reports published by the organizations from 2010 to 2013, with the aim of generating qualitative data pertaining to program evaluation results and their potential utilization. This four-year time frame was chosen to ensure that the reports feeding into broader reallocation exercises, such as the Strategic Review and the Deficit Reduction Action Plan (DRAP) are included in the analysis. In total, seventeen reports[2] were analyzed in depth to capture common themes, which were identified through the literature review and emerged from the reports themselves. The information contained within each report was first divided up according to the following sections/variables:

- *subject area* of each program evaluated
- evaluation *methods* used and their respective *limitations*
- *internal/external* evaluators responsible for the evaluation
- presence of an *evaluation steering committee*
- *program objectives*
- *type of program*
- *inclusion of stakeholders*
- time since the *last evaluation* of the program
- *evaluation approach* (i.e. participatory, utilization-focused, other)
- *type of evaluation* (summative or formative)
- existence of a program *logic model*
- evaluation issues and *questions*
- *key findings*
- *recommendations*
- *management response* to the recommendations (i.e. accepted or rejected)

[1] Both participating organizations requested anonymity in all reports and publications stemming from this study. Any information that would lead to identification of the organizations has been removed from this article for this reason.

[2] Although 17 reports were examined in total, they represented 23 evaluated programs.

- *types of evaluation use anticipated by management* (i.e. instrumental, conceptual, symbolic, process or imposed).

While the data were mostly captured through insertion of literal excerpts from each report in an Excel matrix, in some instances the recommendations and findings were paraphrased or summarized in order to focus the analysis on only the most critical data.

After the initial capture in Excel, we decided to transfer all of the raw data into NVivo in order to more easily code and categorize the report excerpts. The category variables outlined above were used to organize the data in NVivo. The use of different query techniques (text search, content search, word search, coding cluster analysis) followed by in-depth content analysis and review by a second reader, produced the preliminary findings outlined in the following section.

Findings

This section presents the findings of the report content analysis. Through the data coding and interpretation process, we sought to identify elements of evaluation "usability", that is, those aspects that enabled us to assess the evaluator's credibility and the report's quality, two of the main precursors for evaluation use.

Evaluation Methods and Approaches

In general, the evaluations reviewed the relevance and performance of programs, in line with Treasury Board policy requirements. All evaluations were conducted using multiple lines of evidence and employed qualitative and quantitative methods. Table 1 outlines the data collection methods that were used to produce evaluation results.

Table 1.
Evaluation methods used

Method	Number of Reports	Observations
Document and literature review	17	Four reports explicitly mention that previous evaluation reports were included
Review of program files and administrative data	17	Mainly administrative data bases with information on program clients and other partners
Key informant interviews	17	Mainly individuals with knowledge of the programs such as organizational staff, partners, and some professional associations
Surveys	16	Mainly former program clients; in a few cases, other stakeholders also participated
Case studies	9	To complement other lines of evidence
Focus groups	5	Mainly with current or former program clients
Other methods	4	E.g., Benchmarking, economic impact analysis

The reports do not provide a clear rationale for the selection of the methods used. This information might have been included in the evaluation framework developed at the outset of each project, however. We also found that often, the term 'evaluation approach' is used to signify 'evaluation methods'. Although these are closely related (the choice of an evaluative approach is usually linked with the selection of commensurate methods), it may have been useful to clarify the terms used in order to avoid potential confusion with evaluation models or approaches which frame evaluation practice and establish its epistemological positioning (for instance: developmental evaluation, participatory evaluation, utilization-focused evaluation, etc.).

Methodological Challenges and Limitations

In order to assess the reports' credibility, we analyzed the methodological challenges and limitations outlined by the evaluators. All but one of the reports clearly identified the limitations of the methods used. The following limits were presented in the reports:

- The lack of availability, consistency and reliability of performance measurement data to feed into the evaluation (9);

- The small number of survey responses, especially from specific sub-groups of program clients (5);
- The low level of visibility of the programs or the limited awareness of the programs among its target population, which made it difficult to determine the specific impact of each program as opposed to the overall impact of the organization (4);
- The lack of comparable data (comparable programs, comparable groups) (4);
- Potential weighting issues related to the greater participation of certain types of survey respondents (4);
- Problems with identifying, reaching or including sub-groups of survey participants (3);
- Problems with quality of self-reported data (3); and,
- Analytical challenges (1)/results could not be generalized (2) particularly in evaluations in which case studies were conducted.

In most reports, these challenges and limitations were mitigated either by the triangulation of sources, the evaluators' experience, or by taking into account the limitations during the interpretation of the findings. Therefore, these limitations do not undermine the credibility of the evaluation findings and should not, therefore, impede evaluation utilization.

Internal/External Evaluation

Most of the evaluations reviewed as part of this analysis were at least partly conducted externally: more than half of the evaluations were entrusted to external consulting firms (11 external evaluations), and five others were hybrid evaluations conducted both by external and internal evaluators. In these cases, internal evaluations were responsible for literature, document and file reviews as well as administrative data reviews, while most of the interviews, surveys, and case studies were entrusted to external evaluators[3]. In only one case was an entire evaluation conducted exclusively by internal resources.

Evaluation Steering Committee

Most of the reports do not explicitly state the contribution of an evaluation steering committee to the evaluation process. This is not atypical, however, as this concerns the management of the evaluation project itself more than the evaluation process and results. This issue will be further developed through the second phase of the research.

Time Since Last Evaluation

The time elapsed since each program was previously evaluated was considered in the analysis, since it may yield useful clues about organizational priorities and potential decision-making. Out of the 23 programs reviewed (in the 17 reports analyzed), in six cases no information regarding previous evaluation activities was available. The time elapsed since the previous evaluation, at the time of the reported evaluation, was obtained for the remaining 17 programs and is shown in Table 2:

Table 2: *Time since last evaluation*

Time	Number of programs
Never evaluated	7 programs
More than 20 years prior	1 program
Between 16 and 20 years	0 programs
Between 11 and 15	2 programs
Between 6 and 10 years	5 programs
Between 1 and 5 years	2 programs

[3] One report did not mention how responsibilities were shared

Existence of a Program Logic Model

Most of the evaluation reports reviewed (13) contain logic models. The reports do not indicate, however, whether the logic models were developed prior to the evaluation by program staff or by evaluators as part of the evaluation process. This issue will be further explored as part of the key informant interviews to be conducted during the second phase of the research process.

Type of Evaluation

Many of the evaluations reviewed (7) were explicitly identified as summative in nature. The remainder of the reports did not specify the type of evaluation conducted; however, given the nature of the recommendations made in these reports, it was determined that all of these evaluations were summative rather than formative. This is consistent with the requirements of the *Policy on Evaluation* (2009).

Evaluation Issues and Questions

All of the evaluations reviewed were based on the five core evaluation issues required by the *Policy on Evaluation* (TBS, 2009), in addition to program design and delivery issues. First, all of the reports (17) addressed the issue pertaining to program relevance. As can be expected, the evaluation questions dealing with relevance generally focused on: the alignment of the program objectives with federal government and agency priorities and policies; the niches that the programs occupy in relation to similar programs at the federal and other levels of government; the existence of a necessary role for the federal government to provide such programs; and finally, a demonstrable need for the program to exist or to continue to exist.

Secondly, the issues and questions regarding program performance were also addressed in all reports, although these tended to be more tailored to the program under study. Such evaluation questions included both the program's efficiency and economy and/or the program's effectiveness[4]. Questions on program efficiency and economy focused on whether there was use of efficient and effective means, tools or models to deliver the program. They also examined how efficiency and effectiveness could be improved, for example through the use of performance management tools and systems. Questions on effectiveness focused mainly on the program's success or impact, and examined progress towards expected program outcomes.

In addition to the five core issues required by the *Policy on Evaluation (2009)*, most of the reports (12) addressed program design and delivery in a separate set of questions (distinct from those focusing on efficiency and economy). More specifically, design and delivery questions focused on whether the program was implemented as planned (i.e., program fidelity), whether it was designed appropriately to achieve its objectives, and whether the program meets the needs of the organization. Moreover, specific questions were raised regarding the mechanisms at work in each program (4).

Evaluation Results/Findings

The preliminary results show that the findings were well aligned to the evaluation issues and questions.

Relevance. All evaluations clearly stated that the programs continue to be relevant. Most reports, for example, confirmed the alignment of the programs evaluated with government and agency priorities (14), and confirmed the continued need for the programs. The federal government is seen as playing a crucial role in delivering these programs. The evaluations provide clear examples that illustrate how the programs continue to be relevant, through the results achieved thus far. A few evaluations, however, found that some programs were not designed to meet all of the needs of their respective target populations (such as sub-groups of program clients), which may limit the reach of the organization and its progress towards stated outcomes.

Design and Delivery. Several reports tackled the issues of program design, delivery and/ or implementation (12). The findings pertaining to these issues can be grouped into common themes such as flexibility of program design and administrative processes, pre-established relationships between program staff and clients, information collection systems, and the division of responsibilities between program staff. Specific recommendations on all of these points were made in the reports.

---

[4] Variations found in evaluation terminology relating to program performance: cost-effectiveness, cost-efficiency, cost-analysis, effectiveness, efficiency and economy.

Program impacts. In terms of program performance, success, impact, and effectiveness, most evaluations found the programs effective in achieving planned outcomes (16). Although specific outcomes cannot be listed here for reasons of confidentiality, overall, very few concerns were raised by evaluators on the progress made toward program objectives.

Recommendations

All of the reports reviewed include recommendations to program managers and organizational decision makers, as per the *Policy on Evaluation (2009)*. Although all of the reports conclude that the programs evaluated should be maintained (17), some modifications or adjustments could be made in all cases to ensure greater efficiency and effectiveness. Here again, the recommendations were specific and tailored for each program but some common themes can be identified throughout the reports. Many of the recommendations focus on program design, such as increasing program budgets to improve reach (11), and improving the application process (3). Other recommendations focus instead on program management, such as: reviewing program logic models (6); improving performance measurement systems or tools (3); finding alternatives to delivery in order to reduce costs (4). Finally, some recommendations encouraged sharing of knowledge sharing and best practices within (between) program partners. (4)

Management Responses to the Recommendations

The management responses submitted by program directors following each evaluation show that management agreed with most of the 54 recommendations presented in the 17 reports reviewed. In five cases, they partially agreed with recommendations, mainly because these were in areas in which managers have little or no control (for instance, some recommendations focused on program partners, something over which the organizations have no jurisdiction). Managers also only partially agreed with other recommendations due to the additional human and financial resources required to implement them.

Types of Uses Identified in Management Response

As can be expected, the reports do not explicitly state the ways in which the evaluations were to be used. Based on the production of recommendations and their management responses, we can, however, infer that instrumental use has occurred when and where the recommendations were implemented as planned. In addition, 14 reports clearly state that the evaluations intend to provide solid evidence in support of program renewal or facilitate decision-making, two clear instrumental uses of evaluation. Other types of use will be explored as part of the key informant interviews. These include conceptual, process and symbolic uses of evaluation (Johnson, 1998).

Discussion and Conclusion

This preliminary content analysis of the evaluation reports published by the two participating organizations over the period 2010 – 2013 highlight common themes and trends. 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. 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 relevant and appropriate, and their implementation will lead directly to instrumental use. What is less apparent from this analysis, of course, is where utilization goes from there. Now that the evaluation process and reports have been found credible and "usable", whether they were used and how they were used remains to be studied. The second phase of this study, therefore, will focus on key informant interviews, to be conducted with the evaluation project managers and the program managers. The interviews will continue to explore instrumental use, but also broaden the scope of the study to also include conceptual, process, and symbolic uses. It is hoped that these results will inform us as to the actual use of evaluation, both toward program improvement, but also toward greater organizational objectives.

References

Alkin, M. C., and Taut, S.M. (2002). Unbundling evaluation use. *Studies in Educational Evaluation*, 29, 1-12.

Amo, C., and Cousins, J.B. (2007). Going through the process: An examination of the operationalization of process use in empirical research on evaluation. In J.B. Cousins (Ed.), Process use in theory, research and practice. *New Directions for Evaluation* (p. 5-26). San Francisco: Jossey-Bass.

Bourgeois, I., Toews, E., Whynot, J., and Lamarche, M. K., (2013). Measuring organizational evaluation capacity in the Canadian federal government. *Canadian Journal of Program Evaluation*, *28* (2), 1-19.

Cousins, J.B., & Leithwood, K.A. (1993). Enhancing knowledge utilization as a strategy for school improvement. *Knowledge: Creation, Diffusion, Utilization*, 14, 305-333.

Dumaine, F., (2012). When one must go: The Canadian experience with strategic review and judging program value. In G. Julnes (Ed.), Promoting valuation in the public interest: Informing policies for judging value in evaluation. *New Directions for Evaluation*, 133, 65–75.

Fleischer, D., and Christie, C.A. (2009). Evaluation use: A survey of U.S. American Evaluation Association members. *American Journal of Evaluation*, 30, 158-175.

Johnson, K., Greenseid, L.O., Toal, S.A., King, J.A., Lawrenz, F., and Bolkov, B. (2009). Research on evaluation use: A review of the empirical literature from 1986 to 2005. *American Journal of Evaluation*, 30 (3), 377-410.

Johnson, R. B. (1998). Toward a theoretical model of evaluation utilization. *Evaluation and Program Planning*, 21, 93-110.

Leviton, L.C. (2003). Evaluation use: Advances, challenges and applications. *American Journal of Evaluation*, 24 (4), 525-535.

McDavid, J. C., and Huse, I. (2012). Legislator uses of public performance reports: Findings from a five year, study. *American Journal of Evaluation*, 33 (1), 7-25.

Government of Canada, Treasury Board Secretariat (2009). Policy on Evaluation, retrieved on May 3 2015, from: http://www.tbs-sct.gc.ca/pol/doc-fra.aspx?section=text&id=15024

Office of the Auditor General (2013). Spring Report of the Auditor General of Canada, Chapter 1 –Status Report on Evaluating the Effectiveness of Programs. Consulted online on April 26th, 2015 at: http://www.oag-bvg.gc.ca/internet/English/parl_oag_201304_01_e_38186.html#hd5m

Patton, M.Q. (1997). *Utilization-focused Evaluation: The New Century Text.* Thousand Oaks, CA: Sage Publications.