PROGRAM EVALUATION IN THE GOVERNMENT OF CANADA: PLUS ÇA CHANGE ...

R.V. Segsworth
Laurentian University
Sudbury, Ontario

Abstract: This article describes the evolution of evaluation policy and practice from 1977, when the first policy on program evaluation was issued by Treasury Board, to the current state of evaluation in the Government of Canada. It discusses the changing nature of evaluation policy, resources, and conceptions of the function as well as a number of concerns about federal evaluation that have been expressed during this period. The article concludes by providing suggestions that, if implemented, might enhance the status and role of evaluation in the federal government.

Résumé: Cet article décrit l'évolution de la politique d'évaluation et de ses pratiques depuis 1977 — date de publication de la première politique d'évaluation du Conseil du Trésor — jusqu'à l'état actuel de l'évaluation au sein du gouvernement canadien. L'article montre la nature changeante de la politique de l'évaluation et ses ressources et conceptions fonctionnelles, ainsi qu'un certain nombre de préoccupations soulevées sur l'évaluation durant cette période. L'article se termine en fournissant quelques suggestions qui, si elles étaient mises en œuvre, pourraient améliorer le statut et le rôle de l'évaluation au niveau du gouvernement fédéral.

Twenty years ago, most of the articles in the first issue of this Journal dealt with the place of program evaluation in Canadian federal government departments and agencies. For example, Michael Rayner, then serving as the Comptroller General of Canada, the agency responsible for ensuring that government departments carry out program evaluations, reported on the status of federal-level evaluation efforts. Leonard Rutman also reported on the status of federal-level evaluation, commented on the sorry state of program evaluation in provincial governments, and raised questions about the nature and role of program evaluation in the federal government. Performance measurement systems and their place in the federal
approach to program evaluation were the subjects of an informal debate between Neil Maxwell, arguing for linking performance measurement to program evaluation work in government departments and agencies, and John Mayne, who argued that performance measurement and program evaluation serve different aims and constitute different functions.

Most of the topics addressed in these articles from the first issue remain current and still controversial, at least within the confines of federal government evaluators. Some of these recurring themes will be addressed in this article. The article is intended to provide a discussion of the historical path federal government evaluation has taken and what paths it seems likely to follow in future. It examines the changing evaluation policies, resources, concerns, and conceptions of the role of evaluation in the Government of Canada. It suggests alternative roles for, and understandings of, evaluation that, if adopted, could enhance the function considerably in the federal government. In the spirit of the aims of the original journal issue, observations and judgements are offered to stimulate debate.

A BRIEF HISTORY

The history of program evaluation in the Government of Canada reveals a pattern of change, expansion, and contraction. This pattern is reflected in alterations to evaluation policy, resources, the subjects of evaluation as well as internal concerns about the evaluation function.

Evaluation Policy

Although there is evidence (Muller-Clemm & Barnes, 1997) that evaluation research existed in the Government of Canada earlier, Treasury Board issued the first formal policy on program evaluation in 1977. This is the formal starting point of federal government program evaluation. Treasury Board circular 1977-47 stated that “departments and agencies of the federal government will periodically review their programs to evaluate their effectiveness in meeting their objectives and the efficiency with which they are being administered” (Treasury Board, 1977, p. 2). Deputy heads were to ensure that procedures were in place such that all programs were evaluated periodically, that evaluation reports were to be communicated to deputy heads and other appropriate levels of management, and that
evaluation reports were objective. The circular went on to note that the Treasury Board policy on performance measurement (Treasury Board, 1976) complemented program evaluation. Evaluations were expected to provide more comprehensive data and to more thoroughly examine programs than performance measurement systems would do. In addition, the program evaluation policy would ensure that all programs would be more thoroughly examined than would be possible under the performance measurement policy. The 1977 circular went on to suggest that each program was to be evaluated at least once every 3 to 5 years. In 1981 Treasury Board policy was spelled out in operational terms when the Office of the Comptroller General (OCG) published the Guide on the Program Evaluation Function (1981a) and Principles for the Evaluation of Programs by Federal Departments and Agencies (1981b). The Guide was issued as Treasury Board circular 1981-25 and became part of the formal policy on program evaluation together with Treasury Board circular 1977-47. The Guide established four basic program evaluation issues and seven basic evaluation questions. These were:

1. **Program Rationale** (does the program make sense)
   a) To what extent are the objectives and mandate of the program still relevant?
   b) Are the activities and outputs of the program consistent with its mandate and plausibly linked to the attainment of the objectives and the intended impacts and effects?

2. **Impacts and Effects** (what has happened as a result of the program)
   a) What impacts and effects, both intended and unintended, resulted from carrying out the program?
   b) In what manner and to what extent does the program complement, duplicate, overlap or work at cross purposes with other programs?

3. **Objectives Achievement** (has the program achieved what was expected)
   a) In what manner and to what extent were appropriate program objectives achieved as a result of the program?

4. **Alternatives** (are there better ways of achieving the results)
   a) Are there more cost-effective alternative programs which might achieve the objectives and intended impacts and effects?
   b) Are there more cost-effective ways of delivering the existing program? (OCG, 1981a, p. 7)
The Guide established that the deputy head was to be the client of the evaluations undertaken in his/her department, suggested that program evaluation units should be established in every department and agency, and proposed a model of the program evaluation process.

In August 1991, Treasury Board rescinded both circular 1977-47 and circular 1981-25 and issued a new policy on program evaluation. The new policy (OCG, 1991, p. 1) stated that “it is government policy to periodically evaluate the continued relevance, success and cost-effectiveness (i.e., program performance) of federal programs and use that information to reconfirm, improve or discontinue programs.” This new policy contained several important changes to the original version. The first was that not all programs were necessarily to be evaluated. Although the policy’s guidelines recommended a six-year evaluation cycle, it noted that “evaluations for which there is no priority need for the performance information produced or which would require excessive resources or time to carry out should not be undertaken” (p. 8). The second change was that “evaluation criteria — the means by which the performance of the program can be judged — be established for all programs” (OCG, 1991, Foreword). The third involved a greater role for other governmental actors, specifically that departmental evaluation planning was expected to respond to evaluation issues that reflected concerns of Treasury Board or other cabinet committees. Furthermore, Treasury Board Secretariat (TBS) and the Office of the Comptroller General staff would review evaluation criteria in departmental Treasury Board submissions. The policy also envisioned evaluations to be requested by Treasury Board. Finally, the four basic evaluation issues contained in the 1981 policy became three program performance issues:

1. Continued Relevance: Does the program continue to be consistent with departmental and government-wide priorities and to address realistically an actual need?
2. Program Success: In what manner and to what extent is the program meeting its objectives, within budget, and without significant unwanted outcomes?
3. Program Cost-Effectiveness: To what extent is the current program the most appropriate, efficient means for achieving the objectives, relative to alternative design and delivery approaches?

This 1991 policy was quite short-lived. In 1994, Treasury Board rescinded it and issued a new Review Policy, which defined evalua-
tion as one of four means by which departments were to “determine and review the performance of their policies, programs and operations in a timely, relevant and cost-effective fashion, emphasizing results, innovations, alternatives and lessons learned” (Treasury Board, 1994, p. 3). The three other review mechanisms were internal audit, performance monitoring, and other types of review. The policy indicated that there were six types of review and described four of them in some detail. Interestingly, the basic evaluation issues remained virtually identical to those described in the 1991 revisions.

In 2001 Treasury Board published a new evaluation policy stating:

It is government policy that departments and agencies embed evaluation into their management practices:

• to help design policies, programs and initiatives that clearly define expected results and that embody sound performance measurement, reporting and accountability provisions at the outset; and

• to help assess in a rigorous and objective manner the results of government policies, programs, and initiatives, including their impacts both intended and unintended and alternative ways of achieving expected results. (Treasury Board, 2001, p. 3)

The policy clearly describes evaluation as a management tool, while also continuing the mandate to include the periodic assessment of policy. The evaluation policy also suggests that departments should “establish ongoing performance monitoring and performance measurement practices” (Treasury Board, 2001, p. 2). The 2001 policy does not define the deputy head of departments and agencies as the client for evaluations, but does require the deputy head “to establish an appropriate evaluation capacity” (p. 3) within his/her organization. The deputy head is expected to provide Treasury Board Secretariat with access to evaluation plans and to warn TBS of evaluation findings that raise major concerns. TBS is required to provide direction for evaluation in the Government of Canada, and the policy calls for the creation of a Centre of Excellence for Evaluation in TBS to take on that responsibility. The 2001 evaluation policy also requires that an evaluation of the policy be undertaken within five years of its implementation date.
Evaluation Resources

To some extent, the importance of program evaluation in government is reflected in the expenditures on evaluation, staffing resources, and arrangements assigned to carrying out evaluation work. Establishing groups of dedicated evaluation staff in designated organizational units is a key indicator of the importance placed on program evaluation in government departments and central agencies. Although Treasury Board evaluation policy (circulars 77-47 [Treasury Board, 1977] and 81-25 [Office of the Comptroller General, 1981a]) both stressed the importance of departments establishing and staffing evaluation units, actual practice varied from department to department as well as ebbing and flowing over time.

Designated Staffing Units

The 1983 audit of the evaluation function by the Auditor General of Canada (Office of the Auditor General, 1983) intensively examined 19 departments and agencies. The Auditor General found that by 1982–83, all 19 had established distinct evaluation units with a median staff size of five. Mayne (1994) reported that by the mid-1980s, most departments had established distinct evaluation units. In an informal survey I conducted in 1992, of the 19 departments that responded, only 8 retained distinct evaluation units but prior to 1988, 15 of them had separate evaluation structures. A second informal review undertaken in 1999 revealed that only 6 of the 20 departments that responded continued to maintain an autonomous program evaluation section, unit, branch, or division. A very simple examination of department structures that I undertook in May 2005 indicated that 11 of 21 departments had autonomous evaluation units while 10 did not, suggesting a significant increase in distinct evaluation units, but not to the levels achieved in the mid-1980s.

One explanation for these changes may have been the changing signals from the centre. For much of the 1980s, the Program Evaluation Branch of OCG provided guidance, support, and monitoring for evaluation in the Government of Canada. In the late 1980s, the Program Evaluation Branch was merged with Internal Audit and became the Internal Audit and Evaluation Branch of the OCG. By 1993, the OCG had been eliminated as a separate entity, and evaluation found itself as one of several functions housed in the Government Review and Quality Services Division of TBS. Only in 2001 did a separate evaluation entity again emerge, the Centre of Excellence for Evaluation. It was situated in TBS a central agency.
In this context, it is hardly surprising that departments, having first supported autonomous evaluation units, merged internal audit and evaluation units in the 1990s in response to a clear message from the TBS that the complementarity of audit and evaluation could be reflected in structural adjustment. But the pattern changed, again in response to leadership from the centre. Given the TBS lead in establishing the Centre of Excellence for Evaluation in 2001, more distinct evaluation units were created in major departments than was the case in the 1990s.

Evaluation Staffing

By evaluation resources we mean the ability of departments and agencies to undertake and/or competently commission evaluation studies. One indicator of capacity is expenditures on evaluation. A second indicator is evaluation staffing. In this case we include not only the number of staff but also the presence of evaluators in departments and agencies. A third indicator of capacity is the number of evaluation reports produced.

The 1983 Auditor General’s Report on Program Evaluation (Office of the Auditor General, 1983) found that 19 departments and agencies employed some 168 person-years (PYs) in 1982–83 and that the median size of the evaluation units was 5. Mayne’s 1994 discussion of the evaluation function in the 1980s indicates that some 60 government organizations, including all major and medium-sized departments, had evaluation staff in units ranging in size from 1 to 15 and that about 350 evaluators worked for the Government of Canada. Segsworth (1994) reports that there were 279 evaluation staff by 1989–90 and this number declined to 261 by 1991–92. The 1993 Report of the Auditor General on the evaluation function (Office of the Auditor General, 1993) notes that the number of in-house staff employed in the 19 departments and agencies audited in 1983 had been reduced. More recently, the Centre of Excellence for Evaluation states that 293 evaluation staff work within the federal government: “The 41 large departments and agencies employ 284 FTEs [full-time equivalents]; the 37 reporting small agencies employ 9 FTEs” (Centre of Excellence for Evaluation, 2005, p. 12).

Evaluation Studies Completed

The 1983 Auditor General’s report noted an increase in the number of evaluation studies completed from 31 in 1980–81 to 87 in 1982–83.
Mayne (1994) reported that about 80–100 evaluations were completed annually during the mid-1980s. Segsworth (1994) claimed that the number of evaluation studies reported to the Public Accounts Committee increased from 75 in 1984 to 100 in 1987 and then declined to 79 in 1989. The 1993 Report of the Auditor General (Office of the Auditor General, 1993) found that the number of completed evaluation studies had declined from 99 reports in 1987–88 to 80 in 1991–92. Segsworth (2002) noted that some 60 evaluation studies were completed in 1997–98. Hunt and Geiger (2005) reported that 130 evaluation studies were finished in 2004–05.

Expenditures on Evaluation

Table 1 presents information on federal evaluation expenditures for selected years. A review of the data presented in the table reveals that in constant dollars, expenditures on evaluation in 1991–92 were almost $7 million less than expenditures on evaluation in 1983–84. Furthermore, in the 1996 Annual Report (Office of the Auditor General, 1996), the Auditor General noted that expenditures on evaluation for 1994–95 had fallen from the 1991–92 values. The expenditure pattern shows evaluation to be in considerable decline from the early 1990s on.

Table 1
Federal Expenditure on Evaluation, Selected Years, 1983–84 to 2004–05

<table>
<thead>
<tr>
<th>Year</th>
<th>Current expenditure ($)</th>
<th>Constant $ ($2005)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1983–84</td>
<td>22,652,000</td>
<td>39,806,861</td>
</tr>
<tr>
<td>1989–90</td>
<td>32,904,000</td>
<td>44,966,420</td>
</tr>
<tr>
<td>1991–92</td>
<td>26,059,000</td>
<td>32,966,204</td>
</tr>
<tr>
<td>2004–05</td>
<td>54,787,092</td>
<td>54,787,092</td>
</tr>
</tbody>
</table>

At this stage a clear pattern emerges. From 1977 to the end of the 1980s, there is a distinct policy on evaluation; a central agency presence to guide, monitor, and support the function; and considerable growth in evaluation capacity and productivity in the Government of Canada. From 1991 to the end of the millennium, we see the elimination of a clear evaluation policy, the loss of a central agency evaluation unit, confusion regarding a relationship between internal audit and evaluation, and a decline in evaluation capacity and production. Beginning in 2001, a distinct policy is promulgated; the Centre of Excellence for Evaluation begins its work in TBS; and evaluation capacity and production increases. Despite the recent improvements, in 2005 a senior federal evaluation official told me that, in his opinion,
the evaluation function had not yet fully returned to the levels it had reached in the late 1980s.

There appear to be obvious similarities between the current status of evaluation and the situation that existed in the 1980s in terms of policy, central agency presence, and capacity. This does not mean that there are not substantial differences in what gets evaluated, what the major concerns about evaluation might be, and what the current understanding of the role of evaluation in the Government of Canada is.

What Gets Evaluated

In the 1983 audit of the evaluation function, the Auditor General noted that the 43 studies examined contained recommendations for using the study results in four overlapping ways. One involved changes in program size. Another recommended changes in program design, while a third type of recommendation was for changes in support operations. The fourth type of recommendation was for doing additional research on program effectiveness.

In a 1992 publication, Guy Leclerc, former Deputy Comptroller General, Program Evaluation, OCG, reported on the utilization of evaluation studies based upon a review of some 200 studies completed between 1982 and 1989, finding the following types of use:

- 45% operational improvements
- 16% program redesign
- 8% confirmation as is
- 3% abandon the program
- 26% improve understanding of program cost-effectiveness and monitoring
- 4% no impact

About 18% of these reports led to an adjustment of the resource levels for the programs concerned and approximately 25% responded to concerns raised by Cabinet or cabinet committees (Leclerc, 1992, p. 52). In another 1992 publication, McQueen (p. 40) reported that 17% of federal evaluation studies led to design or structural changes, 50% resulted in operational improvements, and 34% involved conceptual utilization.

In 2005, the Centre of Excellence for Evaluation noted “that evaluation work tends to focus on formative evaluation (e.g., management
practices) rather than summative assessments (e.g., what difference did the program make?” (p. 18). The 2003 Interim Evaluation of the Treasury Board Evaluation Policy (Centre of Excellence for Evaluation, 2003) noted a caution regarding the role of program managers in evaluation, that is, that they might tend to focus more on process or program delivery concerns. Recent interviews in Ottawa tend to confirm the conclusion that a significant majority of evaluations focus on operational/managerial concerns.

The evidence suggests that rather little has changed in terms of the issues addressed by evaluators over time. Despite a policy mandate to deal with results, most evaluation studies continue to emphasize operational concerns.

Internal Concerns About Evaluation

The major concerns about the evaluation function in the Government of Canada expressed publicly by federal evaluators reveal a similar pattern. One concern that remains throughout the period is the failure of evaluation to tackle interdepartmental (horizontal) issues. The Auditor General’s Report (1983) as well as Ulrich (1984) express the view that more interdepartmental evaluations are required. The OCG (1985) reiterated that concern a year later. A second issue involves considerations about the roles of the evaluation function in the Government of Canada. McQueen (1992), for example, argues that a major role is to demonstrate the relevance and cost-effectiveness of programs. Leclerc (1992) adds the need to demonstrate performance.

The Draft Evaluation Function in the Government of Canada (Centre of Excellence for Evaluation, 2004b) argues that evaluation should contribute to important management practices, should support executive committees and central agencies, and should assist in improving management practices and results. In their May 2005 presentation, Evaluation in the Government of Canada: The Roles of TBS and Departments, Hunt and Geiger make the case that evaluation should provide better advice on program performance, should assist in identifying savings and assessing value-for-money, and should lead to program improvements. They note that evaluation has a significant role to play in justifying future allocations under the Treasury Board Transfer Payment Policy.

Another concern about quality has been raised by the Auditor General
in every audit of the evaluation function. In particular, the Office of the Auditor General (OAG) has raised the issue of methodology. In its 1993 report, for example, the OAG “identified only 2 programs that had been evaluated consistent with standards of quality” (p. 13). Rutman (1986) also emphasized this point in the first issue of CJPE and noted “an anti-rigour orientation in which insufficient emphasis is placed on such important research principles as validity, reliability, attribution of cause and effect, and generalizability” (p. 25). Currently, the Centre of Excellence for Evaluation is spearheading an effort to certify evaluators in an attempt to ensure that the quality of evaluation studies and reports is improved.

There is a significant concern about evaluation capacity in current discussions about evaluation. This issue is discussed in some length in the *Interim Evaluation of the Treasury Board Evaluation Policy* (Centre of Excellence for Evaluation, 2003) and in the 2005 Centre of Excellence for Evaluation report on *The Health of the Evaluation Function in the Government of Canada* (Centre of Excellence for Evaluation, 2005). The argument is that the Government of Canada does not possess sufficient capacity to allow evaluators to meet all of the demands placed upon them. This precise point was raised in the very first issue of the *Canadian Journal of Program Evaluation* in 1986, when the Comptroller General, M. Rayner, noted that 100 evaluation studies per year “is about half the rate we need if virtually all government programs are to be reviewed at least every five years” (1986, p. 3).

The Concept of Evaluation

Contemporary evaluation literature reveals a rather different approach to understanding evaluation than the conceptions that existed in the 1960s and 1970s, when the term became part of the lexicon and practice of federal public servants (Dobell, 2003; Muller-Clemm & Barnes, 1997; Segsworth, 1993). It is possible that current understandings of evaluation in the federal government are different now from what they were in the 1980s.

McQueen (1992) defined federal evaluation “as an integral part of management decision-making and as a means of supporting management accountability” (p. 28). Leclerc (1992) claimed that “program evaluation is a management function” (p. 52). The *2001 Evaluation Policy* (Treasury Board, 2001, p. 2) describes evaluation as a management tool with two main purposes:
• to help managers design or improve the design of policies, programs, and initiatives; and

• to provide, where appropriate, periodic assessments of policy or program effectiveness, of impacts both intended and unintended, and of alternative ways of achieving expected results.

The 2004 Evaluation Guidebook for Small Agencies (Center of Excellence for Evaluation, 2004a) indicates that evaluation has two main purposes — improvement and accountability.

This limited evidence indicates that the understanding of what evaluation is and the reasons for its existence in the Government of Canada remain largely unchanged. Thus evaluation policy and practice in the federal government are not without their critics, including several with first-hand experience in the government’s evaluation function.

The 1986 Task Force on Program Review concluded that “government program evaluations are generally useless and inadequate” (Canada, 1986, p. 23). Dobell and Zussman (1981, p. 406) described the situation as one in which “a solid decade — almost two — has gone into changing the words and the forms.” Douglas Hartle’s critique was harsher. He claimed (1990, p. 23A:30) that “a largely vacuous form of program evaluation has been institutionalized.” He went on to state that since evaluation reports are “too innocuous to even make acceptable self-serving propaganda” (p. 23A:28–29), it was not surprising that nobody uses them. Segsworth (1994, p. 258) reported that “program evaluation did not contribute in any meaningful way to the downsizing decision making process of the Government of Canada.” Donald Savoie provided the following explanation of why evaluation did not play a greater role in government efforts to deal with Canada’s fiscal crisis.

As it is, there are two central agencies singing from “different hymn books” about program effectiveness and efficiency. One can make recommendations on resource allocations to ministers and the other is responsible for evaluating program effectiveness. The one with clout makes only limited use of program evaluation results, not simply because they are largely “useless and inadequate,” but also because the responsibility for program evaluation lies with the Office of the Comptroller General.
and not with the Treasury Board program branch. The program branch also has little opportunity to orient the program evaluation function to the requirements of the expenditure budget. (1990, p. 116)

Muller-Clemm and Barnes (1997, p. 67) “identified the promise versus performance problem” as one of the key issues in federal program evaluation. In essence, their argument was that although the evaluation function had the potential to contribute significantly to policy decisions, expenditure decisions, and operational improvements, it had not lived up to its potential.

DISCUSSION

The evidence above suggests that little has changed insofar as the evaluation function in the Government of Canada is concerned. Recent discussions with officials in the Office of the Auditor General (OAG) support this conclusion. The OAG had tentatively scheduled an audit of the evaluation function report for 2003. Based on initial inquiries, the OAG decided to postpone the audit because it felt that there would be nothing new to report.

The evaluation function, then, is likely to continue largely unchanged. Given the OAG’s mandate to report instances where satisfactory procedures have not been established to measure and report the effectiveness of programs, where such procedures could appropriately and reasonably be implemented, and given the OAG’s support for evaluation as evidenced in its Report for 1996, the function is likely to survive. There may even be grounds for expansion of the function. The Centre of Excellence for Evaluation has a rather ambitious list, including more emphasis on demonstrating value-for-money and supporting the expenditure review process. In the latter case, the Centre of Excellence for Evaluation has encouraged departments to include the review issues — public interest, role of government, federalism, partnership, value for money, efficiency, and affordability — in their evaluations of programs. A number of recent developments indicate a serious effort in this direction. David Maloney, Senior Assistant Secretary, Expenditure Management Sector, TBS, noted in September 2005 that the new home of the Centre of Excellence for Evaluation is the Expenditure Management Sector. Following the recommendation of Peter Aucoin (2005, p. 22), the evaluation function will support expenditure management decision-making in this new “Budget Office.” In addition, the
Centre is encouraging a risk-based approach to departmental evaluation plans to ensure coverage of important priority programs and altering the focus of evaluation to better serve the needs of deputy heads, ministers, and central agencies.

Two other developments are of interest and potential importance to federal evaluation initiatives. The first involves the Results Management and Accountability Frameworks (RMAFs). The second relates to the Mayne-Maxwell debate found in the first issue of the *Canadian Journal of Program Evaluation* and subsequent thinking on the subject.

RMAFs are required for funding submissions of grant and contribution programs to assist in planning, monitoring, evaluating, and reporting on policy/program results. They involve program logic models and the development of performance measures to monitor the programs. They also include an evaluation plan. Treasury Board has adopted a requirement that for 2005–06, each department must develop program activity architecture that “reflects how a department allocates and manages the resources under its control to achieve intended results and how programs are linked to the department’s strategic outcomes” (Centre of Excellence for Evaluation, 2005, p. 5). This requires performance measures and evaluation. Particularly interesting in this regard is that departmental evaluation units have been involved in the production of some 495 RMAFs between 2001 and 2003–04.

The Mayne-Maxwell debate was, in part, a discussion of the definition of evaluation in the Government of Canada. Crudely put, Maxwell (1986) argued that performance measurement was evaluation and Mayne (1986) disagreed. In terms of the Government of Canada and its policies, Mayne was correct. Treasury Board issued two distinct policies — one on performance measurement in 1976 and one on program evaluation in 1977. It followed then that, at least in the Canadian federal government, these were two distinct functions. More recently, Mayne (2001) has suggested the use of “contribution analysis” as a form of evaluation that is committed heavily to the intelligent application of performance measures.

The RMAF and PAA requirements and Mayne’s more recent discussions of performance measurement and evaluation both recognize a much closer relationship between the two functions. Indeed, many evaluators such as Wholey (1983) and Hatry (Hatry, Winnie, & Fisk, 1981) see performance measurement as one form of evaluation. The Evaluation Research Society (1982) listed program monitoring as one...
of the types of evaluation amongst others. Segsworth (2002) develops this argument in a discussion of two perspectives on the future of the federal evaluation function. In their text, *Foundations of Program Evaluation* (1991), Shadish, Cook, and Leviton note that “the vast majority of evaluation practice is about the third step in this metatheoretical logic — measuring performance” (p. 460). It may be that this view is gaining support within the Government of Canada and that, in line with Neil Maxwell's original position, the two functions might be merged at some future date. Such a step would strengthen the two, generate greater coherence in results measurement and reporting responsibilities, and assist this revised evaluation function to respond better to the OAG's interest in accountability. A more robust evaluation function of this sort should enhance what McQueen (1992, p. 45) describes as evaluation serving “the more modest purposes of decision making and accountability within government.” It would respond to Rutman’s (1986, p. 25) argument that “the program evaluation functions in federal departments should turn their attention to the available opportunities for developing performance measurement systems that also serve the interests of program evaluation.”

It is not at all clear that such a step, one that would be considered a major change by federal officials, would be sufficient to fulfill the promise of evaluation. For that, we might look to two alternatives that are not necessarily mutually exclusive.

The first involves a serious rethinking of the role of evaluation and an extension of the current conception of evaluation in the federal government to include it as part of the policy function of government. One knowledgeable official in the Government of Ontario assessed the future for evaluation there in related terms. He suggested that if it were to become a policy function, this would imply that more substantive programs would be subjected to review (the Ontario Government term for evaluation at the time) and that reviews would focus on effectiveness and rational issues. If, on the other hand, the function were assigned to Finance and Administration, smaller programs would be reviewed, and the focus of such reviews would be primarily operational efficiency. (Segsworth, 1992, p. 76)

In many ways, such an adjustment logically reflects the understandings of evaluation that have emerged since its days as a key component of Planning, Programming Budgeting Systems. It would reflect
the fact that many of the ERS (1982) evaluation functions can serve policy development efforts.

In other words, evaluation could and should serve policy purposes. Ray Rist provides a useful discussion of the perceived differences between policy analysis (prospective) and program evaluation (retrospective).

A troubling aspect of this distinction ... is that these definitions essentially address two sides of the same coin. Good data on what has already happened can have profound influences on thinking about likely future effects. If history is prologue, then it is to be assumed that policy makers would look at what has occurred (and been learned) in similar circumstances before making the present decision. Likewise, once a decision has been made and the impacts of that decision begin to come into view, the likely effects are now transformed into actual effects. Consequently, what appears at first glance as a clear dichotomy — retrospective and prospective — becomes with closer scrutiny, two stages of an interactive process: decisions are made, information is gathered about the effects of those decisions, further decisions are made with data available on the results of previous decisions, etc., etc., etc. (1990, pp. 4–5)

Rist goes on to argue “that program evaluation now can encompass the various stages of life cycles of a program or policy — from conception, through execution, through impact” (1990, p. 5).

A similar perspective emerged in a conversation in May 2005 with a senior federal government policy analyst. The analyst indicated that in his department, evaluations of federal-provincial agreements had been undertaken, in part because of the TB policy on transfer payments. He stated that the information received through the evaluation process would feed into policy discussions with provincial officials and with his minister. He claimed that the evaluation findings would result in changes to the agreements when they are reviewed (renegotiated). This might result in more evaluations that are “focused on the life-and-death questions included under program rationale” (Rutman, 1986, p. 23). A May 2005 report by G. Breem and Associates found that the evaluation function linked poorly to policy and decision-making. Both Lee McCormack (2005) of the TBS Results-based Management
Directorate and Terry Hunt of the Centre of Excellence for Evaluation (2005) agree that a re-orientation of the function to better serve the needs of deputy heads and central agencies is warranted.

A much broader conception of evaluation’s role begins, for the purposes of this article, in Mayne’s (2003) “Results-Based Governance: Collaborating for Outcomes.” He argues that there are three key characteristics of contemporary developed societies: complexity, collaboration, and citizen participation. He acknowledges that governance has many meanings and defines it as “the manner in which power is exercised.” He restricts his discussion to a micro-level focused on power and responsibility in a specific programming area. His analysis leads to the conclusion that new approaches to governance are required, and he argues that results-based governance is an appropriate response. In that context he claims that measuring and reporting properly on outcomes has a number of benefits (pp. 163–164):

1. sorting out roles and responsibilities;
2. coordinating governments and citizens;
3. encouraging local and citizen involvement;
4. encouraging learning;
5. enhancing accountability.

He concludes (p. 175) that governments and governance cannot work in a complex, collaborative, citizen-focused society without effective measurement, accountability, and audit; hence results-based governance is essential.

In a brilliant report for the Ontario Panel on the Role of Government, Dobell (2003) extends this type of argument much further in a macro-level application. He would, I suspect, argue that Mayne is only partly correct. Dobell argues that there are two sets or types of issues with which government must deal. The first are rather routine matters, for which comprehensive auditing, performance measurement, traditional evaluation, and other positivist approaches of the sort espoused by Mayne should apply. For those issues, such techniques are appropriate and helpful.

Dobell’s discussion of “messy” issues is particularly important. Messy issues are inherently political and the criteria are inevitably ethical. He points out that our understanding of evaluation, and from evaluation, has grown tremendously. He cites, for example, utilization-focused evaluation, responsive evaluation, constructivist evaluation,
realistic evaluation, empowerment evaluation, and deliberative democratic evaluation. He argues that “most evaluation that matters must therefore be approached through the realm of interactive deliberation, or democratic evaluation” (2003, p. 50).

His argument addresses the limitations to unfettered deliberative democratic processes created by a federal, representative government. It recognizes the difficulties posed by attempting such efforts on a national, or even a provincial, scale. At the same time, Dobell’s report suggests that the prospects for e-democracy might provide the means by which such processes could be put it place. It provides rather practical, albeit rather dramatic, proposals for the role of evaluation in the Government of Ontario. Dobell suggests (p. 7) that evaluation should be focused on

- effective knowledge management systems emphasizing tacit knowledge, not information and communications technology, in which MPPs might serve constituents as knowledge brokers;
- assurance of accessibility (that is, truly effective access) for citizens to information in government hands, again facilitated through strong public service support for the efforts of MPPs in serving constituents directly, and perhaps through legislative task forces or roundtables;
- open analytical and procedural support by the public service for interactive deliberative processes of shared decision-making in communities, particularly through imaging, visioning, and simulation capacities, and participatory integrated assessment processes;
- provision of fora — safe places for ongoing participatory involvement of citizens from the full range of interest groups or communities in shared governance.

What Dobell (2003) does in his report is provide an application of two conceptions of evaluation to two kinds of problems. He demonstrates how the evaluation literature can lead to a different relationship between government and the governed and how the evaluation literature provides ideas and insights on how a new governance model might emerge to involve citizens and their elected representatives in fora to tackle and resolve very thorny policy and implementation issues. At the same time, he recognizes that older, positivist notions of evaluation support societal and governmental disposition of simple policy and management issues.
CONCLUSION

This review of developments related to evaluation in the Government of Canada from 1977 to 2005 suggest that very little in the way of fundamental change to the function has taken place. There is, of course, ebb and flow of resources, capability, and evaluation output, most of which is focused on understanding and improving managerial processes. There is some prospect that evaluation might feed more into expenditure review processes and have more influence on expenditure decisions than was the case in the past. There is a possibility (rather remote in my estimation) that performance measurement would be recognized as an element of evaluation and that it could be included in evaluation policy and practice in the future.

Although more extensive roles for, and conceptions of, evaluation exist in the literature that would result in evaluation being perceived as, and included in, the policy development function of government in at least a few ways, we are, at best, inching forward in such directions. I do not expect to see such dramatic developments being applied in the Government of Canada in the near future. In that regard, it is likely that evaluation will not fulfill its promise in the foreseeable future. On the other hand, it might be wise to remember what Mayne (2003, p. 175) had to say a few years ago: “It is hard to recognize trends when you are in the middle of the stream.”

ACKNOWLEDGEMENTS

The author acknowledges, with gratitude, the helpful comments and suggestions of his colleague, Dr. Mary Powell, and the anonymous reviewers of the Canadian Journal of Program Evaluation.

REFERENCES


**Bob Segsworth** teaches program evaluation in the Department of Political Science at Laurentian University, Sudbury, Ontario. He served as editor of the *Canadian Journal of Program Evaluation* from 1994 to 2002.