The Canadian Journal of Program Evaluation Vol. 12 No. 1 Pages 47–70
ISSN 0834-1516

A HISTORICAL PERSPECTIVE ON FEDERAL PROGRAM EVALUATION IN CANADA

Werner J. Müller-Clemm
E.R.G. Evaluation Research Group Inc.
Victoria, British Columbia

Maria Paulette Barnes
Department of Psychology/Faculty of Law
University of Victoria

Abstract: Federal program evaluation has been a central force in defining the history of evaluation activities in Canada. This article reviews the key policy issues that underlie federal program evaluation by analyzing the historical context within which the federal evaluation function evolved. Because federal evaluation has been linked to a perceived need for accountability, and because evaluation has been expected to play an important role in the policy making, planning, and budgeting of the government (McQueen, 1992), a key issue in the history of program evaluation in Canada is outcome *promise* versus *performance*. The authors identify and discuss factors that shaped the practice of federal program evaluation, such as federal policy issues, political actors, organizational structures, and political and governmental changes, in order to generate a more complete understanding of the current state of program evaluation and its future within the federal government.

Résumé: Le programme d'évaluation féderale a été une force centrale en déterminant l'histoire des activités d'évaluation au Canada. Cet article traite des questions de politique principales qui sont au fond du programme d'évaluation fèderale, en faisant une analyse de l'histoire dans laquelle la fonction d'évaluation féderale s'est développée. Puisque l'évaluation féderale a été lié au besoin de l'imputabilité, et puisque l'évaluation a dû jouer un rôle important dans la politique, l'organisation et la finance du gouvernement, une question principale dans l'histoire du programme d'évaluation féderale au Canada, en ce qui concerne les résultats d'évaluation, est celle de *la promesse* contre *l'accomplissement*. On discute les influences qui ont formé la pratique du programme d'évaluation féderale, par exemple, les questions de la politique féderale, des intervenants politiques, de l'organisation et des changements dans la politique et dans

le gouvernement, dans le but de mieux comprendre l'état d'évaluation des programmes à notre époque et en comprendre aussi le futur dans le gouvernement féderale.

Federal program evaluation has been a central force in defining the history of evaluation activities in Canada. In fact, the history of program evaluation in Canada is largely the history of program evaluation *in the federal government* of Canada. This is due, in part, to the unified parliamentary structure of the Canadian political system under which the nation's social welfare is a political responsibility (in contrast to, for example, the United States, which rigidly separates its executive and legislative branches).

In this paper we offer an understanding of the key policy issues that underlie federal program evaluation in Canada. In the first section, we present an analysis of the historical context (up to 1994) within which the federal evaluation function evolved, for we believe that only by appreciating the history of federal program evaluation will evaluators be able to anticipate future developments in their field. In the second section, we discuss some of the trends in program evaluation. Our conclusions are based on information collected through literature search, review, and synthesis and by conducting personal interviews with experts in the area of federal evaluation.

Program evaluation is an enormous field: many factors may be identified as influencing its growth at the federal level. In the following brief historical account of the development of federal program evaluation, we have consciously decided to include some information and omit other aspects. In our analysis we focus on "internal" influences on federal program evaluation, which include federal policy issues, organizational structures, political and governmental changes, and the results of evaluation strategies and practices. We note, however, that "external" influences such as the program evaluation policies of other countries and the two World Wars contributed to the way in which federal program evaluation developed. Moreover, though we recognize the political and societal complexity of the federal evaluation context — for example, that there were many important contributors to the practice of evaluation, and that private-sector evaluation influences that of the public sector — we leave these topics to other commentators. In a decade-by-decade analysis of federal program evaluation, we identify key changes in program evaluation over its history in Canada. We conclude by looking at the future of federal program evaluation, and comment on how we see program evaluation evolving after 1994. Our examination begins with the issue of accountability and the development of early legislation and federal offices. Each of the following subsections will be prefaced by a tabular, chronological summary as a reference guide for the reader.

THE BEGINNING OF ORGANIZED PROGRAM EVALUATION IN CANADA

Year	Event	Description/Impact on Program Evaluation
1931	Consolidated Revenue & Audit Act (amendments)	Traditional emphasis on financial audit and compliance audit

According to McQueen (1992), the evolution of federal program evaluation was somewhat sporadic in the early stages, and it unfolded through a series of phases that hinged upon a growing need for evaluation as a means to accountability and as an important component of policy making, planning, and budgeting. As a starting point, we examine accountability as a central feature in defining program evaluation in Canada.

Accountability and the justification of spending of public funds has long been a national concern. This concern can be traced back to the establishment of federal offices and legislation. The Treasury Board was created in 1867, the Finance Act in 1869, the Civil Service Commission began in 1908, and the consequent Civil Service Act was enacted in 1918. All of these developments culminated in the Consolidated Revenue and Audit Act of 1931 (Government of Canada, 1990). Early federal "evaluations" took the form of financial audits, because "programs" such as unemployment insurance or Canada pensions were assessed strictly on their actuarial soundness (Corbeil, personal communication, August 1994). From its inception in the 1870s until the changes introduced in the 1970s, the Office of the Auditor General's (OAG) evaluation activities were based on financial and actuarial considerations, emphasizing a financial/compliance audit approach. We begin our look at the 1960s with an examination of this central influence on federal program evaluation.

THE 1960S: DECENTRALIZATION AND "PROGRAM" DEFINED

1960s — Confident "Social Indicators" Movement

Year	Event	Description/Impact on Program Evaluation
1960	OAG under Maxwell Henderson	Initiated radical changes to the function of the OAG
1960–63	Glassco Commission	Royal Commission on Government Organization
1962	OAG change of focus	From strictly financial management to a more comprehensive, operational management perspective
1966	TB's *Financial Management Guide* (directive MI-3-66)	Identified the need to monitor and evaluate the performance of government activities
1968	Planning, programming, budgeting (PPB)	A new financial management approach
1969 the	Decentralization	Move from central control by the comptroller of
		TB to decentralized departmental management
1969	Formal adoption of PPBS	A framework for evaluating results of government programs that explicitly called for program monitoring and assessment
1969	TB secretariat	Planning Branch formed

In 1960 Maxwell Henderson became the auditor general of Canada (Segsworth, 1990a). In the 13 years that spanned his leadership of the OAG, the office witnessed considerable change in evaluation policy and practices. By the mid-1960s, Henderson had introduced his own commentary and criticisms into the auditor general's annual reports, which expanded the customary purview of the auditor general's reporting practices from their traditional and rather impartial reporting of financial audit results, to include personal judgements of the auditor general.

Henderson's commentary on issues such as "non-productive expenditures," for example, popularized the status and role of the OAG by drawing considerable media attention to it (Segsworth, 1990a), as well as emphasizing evaluation's shortcomings in the area of performance. In reaction to this unprecedented behavior of an auditor general, the Liberal government under Pierre Trudeau's leadership introduced Bill C-190 in 1969. Had the bill not been dropped in response to embarrassing media pressure, it essentially would have stifled the auditor general's powers of commentary (Segsworth, 1990a). Instead, according to one observer, by the end of his term in 1973 Henderson "had extended the mandate of the OAG, developed considerable political support outside Parliament, created a belief among Canadians that something was seriously wrong with the system, and created a climate favourable to even more power for his office" (Segsworth, 1990a, p. 44).

Legitimation of a federal program evaluation capability came about through the reports of the Royal Commission on Government Organization in 1960–1963, commonly referred to as the Glassco Commission, which in turn spawned a series of amendments to the Financial Administration Act in 1967 (Office of the Comptroller General [OCG], 1981a). In opposition to the centralized expenditure control policy that had been in place since the implementation of the Consolidated Revenue and Audit Act in 1931, the Glassco Commission recommended management reforms: decentralizing authority to ministries, departments, and agencies in order to "let the managers manage" (Government of Canada, 1990). Although the commission did make recommendations on how this decentralization process should be structured and implemented, only some recommendations were subsequently implemented.

The Glassco Commission emphasized the importance of checks and balances and the need to maintain the accountability of managers to the government (Auditor General of Canada, 1975). The embodiment of this "counterbalance" to the shift in management responsibility, financial management, and control was to have been the Treasury Board. However, owing to the burden of additional administrative responsibilities, the Treasury Board found that the "somewhat technical and continuing financial control aspects appear[ed] to have had a somewhat lower priority" (Auditor General, 1975, para. 2.12). In fact, by 1975 few of the 28 departments and agencies examined were fulfilling the role envisioned for them by the Glassco Commission (Auditor General, 1975). Once again, federal expectations of what program evaluation could produce exceeded its performance.

By 1962 the OAG had moved away from adherence to a strict financial management model toward promotion of a more comprehensive, operational management perspective. As first suggested by the Glassco Commission in its report on the accountability *and* justification of the spending of public funds, the "ultimate test of [the federal government's] spending is not of its legality but of its effectiveness in the view of the public" (as cited in Jordan & Sutherland, 1979, p. 585). In 1966 the *Financial Management Guide,* accompanied by Directive MI-3-66 issued by the Treasury Board, noted the need to monitor and evaluate the performance of government activities using modern management techniques (Dobell & Zussman, 1981; Segsworth, 1990b). By 1969 the introduction of a new financial management approach called the Planning, Programming, Budgeting (PPB) system heralded the beginnings of social program evaluation policy in Canada (Segsworth, 1990b). This, along with the creation of a Planning Branch to the Treasury Board Secretariat, marked the move from central control by the Comptroller of the Treasury Board through pre-audit of all expenditures to decentralized departmental management as suggested by the Glassco Commission.

As decentralization, to be effective, required accountability measures, the PPB policy was conceptualized to provide an analytical vehicle for constructing a coherent and hierarchical program framework of all federal government activity for the primary purpose of resource allocation (Sutherland, 1990). The *Planning Programming and Budgeting Guide of the Government of Canada,* published in 1969, contained an explicit call for program monitoring and assessment.

The PPB system was superimposed upon the existing parliamentary appropriation process and was allowed to evolve and take control in an incremental fashion. In spite of high expectations, Sutherland (1990) argued that the anticipated unified accountability system did not materialize out of its "ambitious theoretical vision … [which] would have required a revolution in social knowledge, let alone in management capacity" (p. 143).

With the federal government's formal adoption of the PPB system in the 1960s, however, the word "program," previously used informally to identify micro- to macro-activities within the government, became conceptualized more clearly (Sutherland, 1990). The PPB *Guide* defined a program as "a collection of activities having the same objective or set of objectives" (Sutherland, 1990, p. 140). A formal definition of "program" helped to shape the questioning of government practices and the usefulness of "innovative" social programs (Jordan & Sutherland, 1979) that characterized the 1970s.

THE 1970S: CHANGES AND CHALLENGES

1970s — Questioning of the Government and the Usefulness of "Innovative" Social Programs

Year	Event	Description/Impact on Program Evaluation
1973	Office of the Auditor General (J.J. Macdonell)	Called in external auditors to conduct studies and restructured the OAG
1973	Operational Performance Measurement Systems (OPMS)	
1974	TB studies (initiated by Macdonell)	
1974	*Operational Performance Measurement* manual	"Means-end hierarchy" of performance levels (the "how-to" elaboration of the PPB system)
1975	"Cabinet Evaluation Studies Program"	
1976	Circular number 1976-25	Secretary of the Treasury Board. Subject: Measurement of the performance of government operations
1976	Lambert Commission	(The Royal Commission on Financial Management and Accountability)
1976	More TB studies	
1977	Bill C-20	The "three e's" (economy, efficiency, effectiveness) value-for-money philosophy
1977	Circular number 1977-47	Secretary of the Treasury Board. Subject: Evaluation of programs by departments and agencies
1978	Office of the Comptroller General created	
1978	IMPAC Survey and the Task Force on Program Evaluability	(Improvement in management practices and controls)
1978	The OCG	Rogers (the first CG): Big "P" and little "p"
1978	PEMS and OPFs	Implementation of the Policy and Expenditure Management System, a planning approach employing operational plan frameworks
1978	Speech from the Throne	Review by Parliament of evaluations by the government of major programs
1978-81	Program Evaluation Branch	(Within the OCG)
1979	Lambert Commission report	Need to improve program evaluation

In 1973 the OAG experienced a change in leadership, with J. J. Macdonell replacing Henderson as auditor general of Canada. Macdonell's tactics differed from those of his predecessor. Initially, Macdonell's reports remained nonconfrontational in that they did not challenge the federal government's spending activities or policies. However, in both 1974 and 1976 Macdonell called in external auditors to conduct evaluation studies; restructured his office by down-sizing and flattening the administrative hierarchy; and, in 1976, issued a critical annual report. Macdonell's report as auditor general expressed disappointment with the government's accountability, as is illustrated in the following passage:

> I am deeply concerned that Parliament — and indeed the Government — has lost, or is close to losing, effective control of the public purse … Based on the study of the systems of departments, agencies and Crown corporations audited by the Auditor General, financial management and control in the Government of Canada is grossly inadequate. (as cited in Segsworth, 1990a, p. 45)

This was "the political storm" that in 1976 led to the formation of the Royal Commission on Financial Management and Accountability, commonly known as the Lambert Commission, which introduced Bill C-20 into the House of Commons and resulted in Parliament's passing the 1977 Auditor General Act (Segsworth, 1990a). We return to the importance of the Lambert Commission later on in this article.

What followed Bill C-20 and the Auditor General Act of 1977 was a return to a more prescriptive form of central financial control driven by the federal government's preoccupation with financial management and accountability. By way of a traditional attest audit, the Auditor General Act confirmed that existing authority rested with the auditor general and ensured the auditor general's relative autonomy. Further, the Auditor General Act contributed to the federal evaluation enterprise the "three e's" — economy, efficiency, effectiveness — as a value-for-money philosophy advocated by Macdonell (Segsworth, 1990a).

Development of a framework for the evaluation of the results of ongoing government programs was initiated by the Treasury Board Secretariat in 1969. Thus departments in the 1970s were encouraged to devise strategies for conducting comprehensive evaluations of their programs' effectiveness and efficiency. Program monitoring was also recommended and was known then as part of a concept called operational performance measurement (OPM). The OPM system was the focus of Treasury Board policy in the 1970s (Treasury Board, 1976).

The OPM system was approved by Treasury Board ministers in 1973, after a period of testing and development. Departments were instructed to supply performance data to the Treasury Board prior to the 1977–1978 program forecast date. Meanwhile, the Treasury Board Planning Branch in 1974 issued their *Operation Performance Measurement Manual,* which featured a "means-end hierarchy" of performance levels (Jordan & Sutherland, 1979). The OPM was the "how-to" elaboration of the PPB system's *Guide* of the 1960s that presented a general evaluation framework, the OPM concept, and related techniques. After an implementation review by Treasury Board ministers in early 1976, departments were directed to accelerate completion of the performance measurement work because the Treasury Board required performance data for "program review and resource allocation purposes" (Treasury Board, 1976, p. 4).

At this point, the Treasury Board largely was concerned with the extent to which departments were meeting its policies and "the extent to which performance data supplied with resource requirements [were] representative of program performance" (Treasury Board, 1976, p. 4). Departments were mandated to improve and increase their evaluation function, for internal management purposes as well as for resource allocation, with the goal that *all* departments would be "using adequate and reliable means, wherever feasible, for performance measurement" by 1980 (Treasury Board, 1976, p. 5). Measures of efficiency were to be related to resource use and measures of effectiveness to program objectives. All of these measures were to facilitate the planning, controlling, and evaluating of programs and program performance, according to the Treasury Board. Program data were to be quantitative and statistical in nature. Finally, a process of internal and external departmental review was also implemented to monitor the progress and quality of departmental evaluations and the OPM system.

The first formalized program evaluation policy in Canada, Policy 1977–47 (Treasury Board, 1977), was put forward by the Treasury Board. In this policy, managers were reminded that they were required to conduct "periodic evaluations in carrying out their responsibilities for the management of their programs" (Treasury Board, 1977, p. 2). Evaluation reports were to be communicated to the appropriate levels of management and were to be "objective." Objectivity was established via three criteria: the development of appropriate terms of reference, the independence of evaluators from the evaluand, and the establishment of appropriate reporting levels. Two evaluation goals identified in the policy were (1) to review the effectiveness of programs in terms of the achievement of their objectives, and (2) to determine the efficiency with which programs were being administered. Based on Treasury Board documents, the scope of evaluation extended to changing the operation of programs, clarifying program objectives, reducing or eliminating programs, and identifying the relative priority of programs.

The Treasury Board's policy in 1977 marked a significant change in federal evaluation policy even though its authors explicitly stated that the policy reaffirmed the Treasury Board's commitment to follow a program-oriented approach to management and resource allocation. The Treasury Board policy differed from that contained in the PPB *Guide* published in 1969 in the frequency of evaluations mandated and in the evaluation techniques to be employed. Whereas the *Guide* promoted an "ideal," thermostat-like financial control system that constantly self-corrects and self-reports on itself according to objective, quantitative data and criteria, the Treasury Board's policy directed departments and agencies to establish three- to five-year evaluation strategies entirely based on the best judgement of program managers. "Best judgment," of course, was a very different focus from the "objectivity" described in the earlier 1969 *Guide*. The switch made by the Treasury Board in 1977 has been described as an "ad hoc analysis, only a judgemental kind of control … without providing any formulae for how the information should be weighted, evaluated and interpreted within the bureaucratic-political context" (Sutherland, 1990, p. 145). Although cost-benefit analysis continued to be mandated through Treasury Board policy (Treasury Board, 1976), the Treasury Board's performance measurement policy, and other allusions to cost-benefit analysis in the 1977 policy, were not enforced (Sutherland, 1990).

The Treasury Board's evaluation reforms nonetheless "represented not merely a difference of degree, but a major difference in kind" (Segsworth, 1990a, p. 41). The concomitant birth of the "comprehensive auditing" approach in 1977, which "purport[ed] to be comprehensive, cyclical, coordinated with other audits, and constructive" (Segsworth, 1990a, p. 46), added support to the change in policy.

In 1978 the Office of the Comptroller General was created. The period from 1978 to 1981 saw the creation of the Program Evaluation Branch (PEB) within the OCG. The PEB and the OCG set out to generate and "sell" two documents, the *Guide on the Program Evaluation Function* and *Principles for the Evaluation of Programs by Federal Departments and Agencies* (Segsworth, 1990b). The latter document attempted to explain and interpret Policy 1977-47 of the Treasury Board, to assist departments in establishing and maintaining an evaluation function, to act as a statement of Treasury Board and OCG program evaluation expectations, and to outline the responsibilities of the OCG concerning program evaluation (OCG, 1981b). The former document stated that "the review and evaluation of existing programs has always been a part of managing in government" (OCG, 1981a, p. 2), and described evaluation activities in the federal government as having evolved from merely concentrating on the resources used toward examination of *how* the resources were used. Its authors emphasized that evaluation activities should also investigate the purpose of programs, their impacts, and in what manner their impacts affect our society.

At its inception, the OCG had two main functions: to develop and supervise federal financial management policies, and to give substance to Policy 1977-47 of the Treasury Board and engage in meta-evaluation (Segsworth, 1990a). Based on an earlier Improvement in Management Practices and Controls (IMPAC) survey and the report of the Task Force on Program Evaluability, the first comptroller general of Canada made a distinction between "big-*P*" programs and "little-*p*" programs (Jordan & Sutherland, 1979). Big-*P* programs were major federal programs addressed by the annual *Estimates* (a regular reporting mechanism from the OAG to Parliament). Little-*p* programs were departmental programs such as the delivery programs for policy objectives, which were to be the focus of OCG evaluation activities (see Jordan & Sutherland, 1979). A Speech from the Throne in a 1978 House of Commons session under the heading "Openness and Social Responsibility" declared that "in the further promotion of open and efficient government, a proposal will be placed before you to provide for the review by Parliament of evaluations by the government of major programs" (as cited in Dobell & Zussman, 1981, p. 405).

The implementation of the policy and expenditure management system (PEMS) in 1978 as a planning approach akin to a type of portfolio management process in the private sector directed program managers to formulate an operational plan framework (OPF) for their programs. These OPFs were to document program objectives, elements, causal linkages, and measurement indices (Sutherland, 1990). PEMS featured the competitive allocation of funds to large programs or departments, supradepartmental policy planning to take place concomitant with budget allocations, and consistent use of language and categories at all levels of reporting (Sutherland, 1990).

In spite of the apparent efforts devoted to the establishment and improvement of a federal evaluation function, the Lambert Commission's report in 1979 emphasized the need to improve federal program evaluation and recommended a cyclical five-year review of government programs to ensure evaluation policy compliance. The mandate of the commission members was to develop a framework for federal accountability; to define the roles and responsibilities of departments, agencies, and individuals within this framework; and to indicate mechanisms for reporting to Parliament (Auditor General, 1979). The Lambert Commission issued its report in 1979. Soon after the formation of the commission, however, the government had introduced Bill C-20 (Segsworth, 1990a). The bill established the OAG as a separate entity with greater authority, autonomy, and responsibility to investigate and report on the financial affairs of the federal government of Canada. This was to be the first of a series of reforms that would change both the audit and the evaluation functions of the federal government in the years to come. Thus ended the 1970s, allowing attention to turn to fiscal restraint and concomitant demands for internal and external accountability in the federal government.

THE 1980S: PUBLIC PROFILE AND PUBLIC SERVICE

1980s — Restraint and Demand for Accountability

Year	Event	Description/Impact on Program Evaluation
1980	Re: TB 77-47	Date by which *all* departments would be using adequate and reliable means, wherever feasible, for performance measurement
1981	Re: TB 77-47	Only 12 departments out of 58 agencies and departments have adequate evaluation functions
1981–83	Publication of resource materials by OCG	*Program Evaluation: An Introduction; Guide on the Program Evaluation Function; Evaluation Framework Studies*

Year	Event	Description/Impact on Program Evaluation
1982	**TB Policy 1982-8**	**Decision to include performance information in the reporting process to Parliament via the annual *Estimates* documents**
1984	**Changes to the Financial Administration Act**	**Evaluation reforms**
1984–86	**Neilson Task Force**	**(The Ministerial Task Force on Program Review)**
1985	**Federal Court decision**	**The auditor general receives greater access to cabinet documents**
1987	**Prime minister refuses to further increase OAG access to government documents**	
1988 **system**	**The IMAA handbook**	**Third attempt at creating a basic information**
1989	**The IMAA "working standards"**	**Stronger emphasis/responsibility for senior line managers to evaluate their programs**
1989	**Supreme Court ruling**	**Auditor general has no recourse to the courts to seek documents when access has been refused by Parliament**
1989	***The Manager's Deskbook***	**New definition of "program"**

By 1981, 12 departments out of 58 agencies and departments that were covered by the Treasury Board's Policy 1977-47 had adequate evaluation functions, and 6 more were working on creating evaluation capabilities (Dobell & Zussman, 1981). Additionally, the formation of the Canadian Evaluation Society in 1981 increased the profile of program evaluation as a profession and provided a much-needed support network of resources.

In 1982, the Treasury Board issued Policy 1982-8, which reflected a governmental decision in 1981 to include performance information in the reporting process to Parliament through the annual *Estimates* documents (Sutherland, 1990). The goal of the Treasury Board's policy in 1982 was to "convey information on results of programs so that parliamentarians [could] carry out their duties to review expenditures more effectively" (Sutherland, 1990, p. 147). The problem, however, was that the judgemental control demanded by Policy 1977-47 of the Treasury Board had replaced any regular "hard-core" performance measurement information. In a 1983 discussion paper entitled "Evaluation Framework Studies," the OCG stated:

> An evaluation framework study is the process through which a (new) program's purpose, background and description are documented and, if need be, clarified. At the same time, it is determined what information must be collected in order to be able to conduct a *useful* evaluation of the program. (OCG, 1983, p. 5, emphasis added)

Significantly, in 1984 evaluation reforms as amendments to the Financial Administration Act were seen as the start of the second phase of program evaluation in the federal government (McQueen, 1992). These reforms extended the value-for-money policy espoused earlier in the OAG. The auditor general's access to information also was expanded in 1985 after a Federal Court decision gave the auditor general greater access to cabinet documents. However, in 1987 Prime Minister Turner refused to further increase OAG access to government documents.

The third attempt at creating a basic information system was the production in 1988 of a document entitled *Increased Authority and Accountability for Ministers and Departments* (IMAA) (Sutherland, 1990). The IMAA handbook (Treasury Board, 1988) and the consequent *Working Standards* released in 1989 (Office of the Comptroller General, 1989) supported the judgemental scheduling of program evaluations but shifted the decision-making from those in charge of the program to a mandated negotiation between the program/department and the Treasury Board in the form of a "memorandum of understanding." Thus a re-emphasizing of the link between evaluation results, the budget, central agencies, and Parliament was endorsed (Sutherland, 1990). Finally, the IMAA system supported stronger responsibility for senior line managers to evaluate their programs (Segsworth, 1990b).

It is clear that in the 1980s the federal government continued to feel the need to improve not only its public profile but also its organizational structure, financial management, and control mechanisms. To this end, Prime Minister Brian Mulroney announced an initiative for the renewal of the Public Service of Canada. This policy was optimistically named "Public Service 2000" (PS2000), and it was to complement and run in parallel with the Treasury Board's IMAA system, given that the Public Service Commission is a separate administrative institution that acts as an agent of Parliament. Although PS2000 reflected a significant change in federal philosophy with regard to organizational structure, policy, and accountability or evaluation, the new system was superimposed upon the existing framework. This approach puts into question the utility of changing the federal structure in this manner.

A main theme in the history of reform in the public service is shifts in the balance between centralization and decentralization of control (Government of Canada, 1990). The purpose of PS2000 was to encourage public servants to be professional, highly qualified, and nonpartisan and to foster a client-focused attitude. Governmental red tape was to be significantly simplified or eliminated, central administrative controls were to be reduced, and innovative methods for improving program efficiency and delivery were to be developed (Government of Canada, 1990). As McQueen (1992) stated, “federal evaluation was well placed to support managers and departments in building on these themes” (p. 39).

One of Brian Mulroney’s first acts as prime minister in 1984 was to initiate a task force on program review. The Ministerial Task Force on Program Review, commonly known as the Nielsen Task Force, was announced one day after the new government was sworn in (Ministerial Task Force on Program Review, 1986). Its mandate was to identify waste, duplication, and red tape in federal departments and then overhaul programs in an ambitious attempt to make them more efficient, effective, accessible, and “decentralized.”

The Nielsen Task Force chose 989 federal programs and services for review that amounted to annual federal expenditures of more than $92 billion, and were given one year in which to complete their work (Ministerial Task Force on Program Review, 1986). With the use of study teams, the task force collected data from consultations or discussions with private-sector groups, federal departments and agencies, and provincial and territorial governments.

The task force identified several broad, interrelated patterns as “generic” issues. One prevalent issue was the inadequacy of program evaluations. This, when linked with the “universal subsidy phenomenon” (i.e., many government subsidies supported “activity” rather than “results”) and governmental ignorance of the “fiscal totality,” which included an incomplete knowledge of direct expenditures, program expenditures, and the hidden costs of regulation, understandably gave rise to the observation that government spending was out of control.

The Task Force report stated that “many study teams found that routine government program evaluations were generally useless and inadequate” (p. 23). Because routine evaluations were conducted by departmental officials for their deputy minister, the evaluation product was, by definition, self-serving. Additionally, as public documents, evaluations are subject to disclosure under the Access to Information Act, a fact that further restricts the quality and usefulness of their content. Program evaluations were found to concentrate on program impact and delivery without questioning the program’s basic rationale. These criticisms led to much debate in the evaluation community about how to increase the usefulness, relevance, and credibility of evaluations (McQueen, 1992).

A Supreme Court ruling in 1989 dictated “that the Auditor General has no recourse to the courts to seek documents when access has been refused by Parliament or cabinet,” and had the effect of keeping cabinet solidarity intact “based upon political bargaining and compromise” (Segsworth, 1990a, p. 50). The auditor general’s response was to go to the media, whose ensuing reports, incidentally, were explicitly based on incomplete information. For the purpose of examining the development of federal program evaluation, the auditor general’s instigation of media involvement severely compromised the OAG’s tradition of neutrality and objectivity.

The *Manager’s Deskbook,* published in 1989, provided a new definition of “program,” as being “any of the various means a department uses to achieve its objectives including, for example, expenditure programs and tax expenditures and regulations” (as cited in Sutherland, 1990, p. 156). This policy extended the purview of the OAG beyond its original mandate to evaluate federal programs to include “off-budget” entities and policy evaluation.

UP TO 1994: A NEW AGE FOR FEDERAL EVALUATION?

Year	Event	Description/Impact on Program Evaluation
1993	OCG folded into TBS	Retreat to audit and statistical security

The federal program evaluation function has had a tenuous reputation. Although sporadic review of policies and programs can be traced back to Confederation, culled from the reports of Parliamentary committees, central agencies, line departments, Cabinet, royal commissions, and task forces, the results of such "evaluative" efforts are of questionable quality (Segsworth, 1990b).

A Canadian program evaluation capability, as a recognizable entity, first came into being in the 1960s through the recognition of its importance in the reports of the Glassco commission from 1960 to 1963. It continued with the introduction of the planning, programming, and budgeting system in 1969, and was followed by the official federal evaluation policy put forth by the Treasury Board in 1977. Unlike in other countries, including the United States, where social experimentation played a large role in the development of social programs and program evaluation, in Canada political and ideological debate fueled its evolution. Consequently, the field of Canadian program evaluation began intuitively rather than systematically, and under considerable government leadership and control.

Criticisms of the federal program evaluation function in Canada have emerged over the years, not the least of which have originated from the Office of the Auditor General. The authors of one article summed up the history of program evaluation in Canada from the early 1960s to the early 1980s in the following way: "A solid decade — almost two — has gone into changing the words and the forms" (Dobell & Zussman, 1981, p. 406).

Criticisms like these are not without merit. The purpose of federal program evaluation has been to review and monitor program performance and improve accountability as a management tool. Even after the Treasury Board in the late 1960s targeted funding specifically at establishing an evaluative function, few studies were ever carried out on program evaluation practices (Mayne & Hudson, 1992). According to Dobell and Zussman (1981), there are three possible reasons for Canada's program evaluation "pathology": lack of agreed theory and purpose, general (i.e., group/organizational) resistance to evaluation activities, and attempts to establish comprehensive evaluation systems that failed to consider the true nature and information needs of the user. Further, commentators have noted problems related to the importance of policy management versus policy content, which are in turn related to the varying levels of expertise at higher levels of bureaucratic hierarchies and the replacement of complex evaluation problems with simplistic program effectiveness indicators (Segsworth, 1990a).

A stronger conclusion, drawn by Sutherland (1990), is that the current federal program evaluation enterprise, as based upon the PPB system, has failed: rather than being an objective field program evaluation has become a political process, and is therefore value-laden or "a negotiated scheduling of wary 'victims' through elite struggle and accommodation at many levels of bureaucratic activity" (Sutherland, 1990, p. 164).

As support for this view, an OCG treatise on the new face of program evaluation entitled *Into the 90s: Government Program Evaluation Perspectives* (1991a) claimed that federal evaluation placed such high demands on evaluators and their products that there seems to be little realistic regard for the theoretical and methodological problems and issues that have historically challenged the profession. The OCG continued to question the usefulness of the evaluation function. In a 1992 discussion paper it recognized that although "federal evaluation clearly pays for itself … it is not a major player in resource reallocation" (1992, p. 6), and that "previous Treasury Board policy called for a 5-year cycle for evaluation coverage … *Recent federal practice, then, is closer to a 12-year cycle*" (p. 3, emphasis added). Moreover, the Standing Senate Committee on National Finance in 1991 reported that the OCG must be strengthened significantly. The increased decentralization of authority through the IMAA and Public Service 2000 initiatives evoked concern that the central control of the evaluative function would be further undermined. The committee recommended that "the personnel responsible for program evaluation in departments and agencies be transferred to and become employees of the Office of the Comptroller General" (p. 21).

On the basis of past developments, it is possible to speculate about what the future holds for federal program evaluation in Canada. For example, one could interpret the recent "integration" in 1993 of the evaluation function under the Treasury Board Secretariat (which, incidentally, renamed it "*audit* and evaluation") as a retreat to the traditional audit philosophy and its concomitant statistical and actuarial security. In 1993 the auditor general reported that:

> The title and responsibilities of the Comptroller General have been added to those of the Secretary of the Treasury Board … audit and evaluation have been assigned to the Deputy Secretary (Administrative Policy Branch) … [and] the former Comptroller General has been named Chief Informatics Officer. (Auditor General, 1993, para. 1.76)

We can surmise that part of the motivation for this recent integration was the fact that the federal program evaluation function ended up on the OAG "intractable issues" list contained in the 1993 annual report, which stated that the government must ensure "that program evaluation asks the tough questions and assesses significant expenditures" (Auditor General, 1993, para. 1.13). This is an obvious re-emphasis of the fact that federal program evaluation has not adequately fulfilled its mandate. There appears to be a return to the "hard facts" of a financial audit philosophy and a marginalization of "soft" program evaluation products.

The integration of program evaluation also represents the culmination of a history of gaps between "promise" and "performance." The task of comptrollership or stewardship of the federal purse has continued to suffer from serious shortcomings insofar as departments and agencies continue to lack, or fail to utilize, the performance and cost data necessary for adequate understanding and operation of their programs (Auditor General, 1993). Consequently, a major concern is that as the evaluation function is subsumed by the "review" function of the Treasury Board, it may be weakened by the general perception that such an integration is a really a reduction in the importance of evaluation. This, in turn, could lead to a downgrading of the evaluation function within departments and agencies, at a time when it should be strengthened.

The Treasury Board *Manual on Review, Audit and Evaluation* was published in 1994 for the purpose of "strengthening the value-added of review at all levels of management" (Treasury Board, 1994, p. 1). The evaluation function is subsumed under review and audit is nevertheless expected to deliver

> timely, relevant, objective, practical and cost-effective evaluation products … on the continued relevance of government policies and programs; the impacts they are producing; and on opportunities for using alternative and more cost-effective policy instruments or program delivery mechanisms to achieve objectives. (pp. 14–16)

The only uncertainty that remains is whether or not this order is too tall for evaluators to fill.

CONCLUSION

More recently, in a treatise on the new face of program evaluation entitled *Into the 90s: Government Program Evaluation Perspectives* (OCG, 1991a), the OCG called for timely, relevant, credible, and objective findings and made heavy demands on federal evaluators and their products with little realistic regard for the theoretical and methodological problems and issues that have historically challenged the profession. Although methodological issues such as measurement problems and attribution problems are addressed in an OCG publication called *Program Evaluation Methods* (OCG, 1991b), the problems evaluators face remain deeply rooted in the application of quantitative social scientific research methods to evaluation questions that are embedded in a sociopolitical context (Guba & Lincoln, 1989). Rigorous empirical methods are inherently time consuming and encourage the collection of data that are abstract, and create problems of relevance and utility. For example, many evaluators have assumed that their evaluands are homogenous (Müller-Clemm & Barnes, 1994). Such bias is in direct conflict with the inherent heterogeneity of Canada, for in reality, a variety of persons participate in programs. Moreover, some charge that evaluators have tended to over-recognize stakeholders with power (e.g., persons in upper hierarchies) and neglect less powerful stakeholders in the evaluation process (e.g., Guba & Lincoln, 1989).

In this article we have described the evolution of the program evaluation function in Canada. In compiling a history of program evaluation in Canada, we provided a decade-by-decade examination of the many factors that shaped the field. Attention to accountability and fiscal responsibility occurred in the form of legislation prior to the end of the last century. Other concepts familiar to contemporary program evaluation emerged sporadically in the 1960s. With the advent of social programs, changes in the federal bureaucratic structure, revisions to the budgeting system, and the increased popularity of the "social indicators" movement, program evaluation activities became more widespread. The late 1960s and early 1970s heralded the establishment of structured, financial management systems and a general questioning of the usefulness and relevance of social programs. By the late 1970s, with its infrastructure firmly in place, the evaluation function in the federal government was rapidly evolving and clearly mandated. In a series of policy documents beginning in the late 1970s, the Treasury Board (1977, 1991, 1994) stated that evaluation products should provide results that may be used to change the operation of programs, clarify program objectives, reduce or eliminate programs, or identify the relative priority of programs. A series of evaluation reforms in the 1980s were initially driven by a push for better and more evaluation products. By the late 1980s, reforms were driven by a need for quality assurance and evaluation standards.

Finally, the role in Canadian program evaluation played by various federal offices cannot be understated. The Office of the Auditor General was created over 100 years ago, and the 1970s marked the beginning of the Office of the Comptroller General. Over the following decade, the comptroller general was responsible for initiating considerable changes to evaluation policy and practices; for instance, newer management systems for accountability were imposed upon the existing organizational structure. Royal commissions and task forces, though ad hoc in nature, also made very significant contributions to the evolution of evaluation practices in Canada. As one example, we showed that the Nielsen Task Force was a relatively recent historical turning point for federal evaluation. Its recommendations in 1986 called the existing federal evaluation function into question and ultimately led to a shift in emphasis, from evaluation to a return to audit. We also considered the impact of government policy on the evolution of federal program evaluation, especially of policy makers' vacillation between centralized and decentralized financial control strategies.

We have identified the *promise* versus *performance* problem as one of the key issues in federal program evaluation. Our analysis has not uncovered a clear solution, yet we remain hopeful. To quote one author with regard to the current decade and beyond:

> We have identified clearly all the ingredients of administrative reform … and have experimented — sometimes successfully, sometimes less so — with different combinations of these ingredients. We are thus well positioned to find the remedies to the deficiencies we perceive. We have set in place certain durable institutions and processes, and we have set in place certain durable, if as yet unrealized, concepts. Above all, the public service seems to be in the course of developing a renewed will for change and return to the fundamental values of public service. (Johnson, 1992, p. 113)

We emphasize the importance of historical analysis in the development of governmental policies, programs, and in the area of program evaluation, and we encourage other observers to examine contemporary issues that continue to affect the future of program evaluation in the Canadian federal government.

ACKNOWLEDGEMENT

Many of the materials that we analyzed were gathered with the help of archivists, librarians, and government workers. In particular, we are grateful to Ron Corbeil. We also thank Laurie Jackson and Jim McDavid, who read and commented on earlier versions of this paper, and Ruth Müller-Clemm, who provided English-French translation.

REFERENCES

Auditor General of Canada. (1975). *Annual report of the Auditor General of Canada to the House of Commons*. Ottawa: Author.

Auditor General of Canada. (1979). *Annual report of the Auditor General of Canada to the House of Commons*. Ottawa: Author.

Auditor General of Canada. (1993). *Annual report of the Auditor General of Canada to the House of Commons*. Ottawa: Author.

Dobell, R., & Zussman, D. (1981). An evaluation system for government: If politics is theatre, then evaluation is (mostly) art. *Canadian Public Administration*, *24*(3), 404–427.

Government of Canada. (1990). *Public service 2000: The renewal of the public service of Canada*. Ottawa: Supply & Services.

Guba, E.G., & Lincoln, Y.S. (1989). *Fourth generation evaluation*. Newbury Park, CA: Sage.

Johnson, A.W. (1992). *Reflections on administrative reform in the government of Canada* (Protected draft for discussion purposes only). Ottawa: Office of the Auditor General of Canada.

Jordan, J.M., & Sutherland, S.L. (1979). Assessing the results of public expenditure: Program evaluation in the Canadian federal government. *Canadian Public Administration*, *22*(4), 581–609.

Mayne, J., & Hudson, J. (1992). Program evaluation: An overview. In J. Hudson, J. Mayne, & R. Thomlison (Eds.), *Action-oriented evaluation in organizations: Canadian practices* (pp. 1–20). Toronto: Wall & Emerson.

McQueen, C. (1992). Program evaluation in the Canadian federal government. In J. Hudson, J. Mayne, & R. Thomlison (Eds.), *Action-oriented evaluation in organizations: Canadian practices* (pp. 28–47). Toronto: Wall & Emerson.

Ministerial Task Force on Program Review (Nielsen Task Force). (1986). *Introduction to the process of program review*. Ottawa: Supply & Services.

Müller-Clemm, W.J., & Barnes, M.P. (1994, May). *Value pluralism and multiculturalism in fourth generation evaluation*. Paper presented at the annual convention of the Canadian Evaluation Society, Quebec City.

Office of the Comptroller General. (1981a). *Guide on the program evaluation function*. Ottawa: Treasury Board of Canada.

Office of the Comptroller General. (1981b). *Principles for the evaluation of programs by federal departments and agencies*. Ottawa: Treasury Board of Canada.

Office of the Comptroller General. (1981c). *Program evaluation: An introduction*. Ottawa: Treasury Board of Canada.

Office of the Comptroller General. (1983). *Evaluation framework studies* (Discussion Paper No. 83-002). Ottawa: Treasury Board of Canada.

Office of the Comptroller General. (1989). *Working standards for the evaluation of programs in federal departments and agencies*. Ottawa: Treasury Board of Canada.

Office of the Comptroller General. (1991a). *Into the 90s: Government program evaluation perspectives*. Ottawa: Treasury Board of Canada.

Office of the Comptroller General. (1991b). *Program evaluation methods*. Ottawa: Treasury Board of Canada.

Office of the Comptroller General. (1992). *Federal program evaluation: A 1992 status report* (Draft for Discussion). Ottawa: Treasury Board of Canada.

Segsworth, R.V. (1990a). Auditing and evaluation in the government of Canada: Some reflections. *Canadian Journal of Program Evaluation*, *5*(1), 41–56.

Segsworth, R.V. (1990b). Policy and program evaluation in the government of Canada. In R.C. Rist (Ed.), *Program evaluation and the management of government* (pp. 21–36). New Brunswick, NJ: Transaction.

Standing Senate Committee on National Finance. (1991). *The program evaluation system in the Government of Canada*. Ottawa: Supply & Services.

Sutherland, S.L. (1990). The evolution of program budget ideas in Canada: Does Parliament benefit from estimates reform? *Canadian Public Administration*, *33*(2), 133–164.

Treasury Board. (1976). Circular number 1976-25, Subject: Measurement of the performance of government operations. Ottawa: Author.

Treasury Board. (1977). Circular number 1977-47, Subject: Evaluation of programs by departments and agencies. Ottawa: Author.

Treasury Board. (1988). *IMAA handbook*. Ottawa: Author.

Treasury Board of Canada. (1991). *Program evaluation policy*. Ottawa: Author.

Treasury Board of Canada. (1994). *Manual for review, internal audit, and evaluation*. Ottawa: Author.