

Some Thoughts On Federal Level Evaluation*

Leonard Rutman

Price Waterhouse, Ottawa

ABSTRACT

Cet article présente quelques commentaires sur la pratique d'évaluation de programme du gouvernement fédéral. Il y soulève trois observations majeures. Premièrement, les évaluations manquent de précision parce que: le sous-ministre n'est pas en pratique le premier client et, plusieurs groupes influencent l'orientation d'une étude; le besoin d'une évaluation n'est pas clairement spécifié; et plusieurs questions sont adressées dans une même étude. Deuxièmement, on insiste seulement sur les études périodiques et l'exclusion d'un contrôle du rendement soulève des problèmes. Troisièmement, il y a une tendance à un manque de rigueur, ce qui crée des études ne possédant point la rigueur méthodologique dans les situations où elle est requise.

In this paper I present some thoughts about program evaluation in Canada. My interest in this subject is longstanding, based on participation in a number of formal reviews going back almost 10 years. In 1976, for example, Dick de Jong and I examined program evaluation activities in several federal departments. I had further opportunities to review federal evaluation activities as a result of work with the Office of the Auditor General in conducting government-wide audits of program evaluation for the 1978 and 1983 Annual Reports. In 1982, I worked with the Canadian Comprehensive Auditing Foundation in designing and implementing a study of provincial and federal evaluation activities. More recently, my views have been shaped by experiences at providing consultation services on program evaluation to federal and provincial departments and agencies. These experiences have given me an opportunity to observe the historical evolution of the program evaluation function and provide the context for the observations offered here. It should be noted, however, that these experiences do not lessen the personal and necessarily subjective nature of the views presented in this paper.

Background

Major impetus for the development of the program evaluation function at the federal level arose from the expanded mandate given to the Office of the Auditor General in 1977, as implemented through the Study on Procedures in Cost Effectiveness (SPICE). The value-for-money mandate appears in Section 2(e) of the Act which directs the Auditor General to assess whether satisfactory procedures were in place to measure and report on program effectiveness, where this was reasonable and appropriate. This meant that there was an implicit requirement for departments to conduct program ef-

* This article is based in part on a speech given to the joint conference of the Canadian Evaluation Society, Evaluation Research Society and Evaluation Network in Toronto, October, 1985.

fectiveness evaluations. The SPICE findings were included in the Office of the Auditor General's Annual Report in 1978, where he concluded:

A review of 23 programs in 18 departments has disclosed few successful attempts to evaluate the effectiveness of programs. The scope and quality of effectiveness evaluation will have to be increased significantly before management, the Government and Parliament, each with its respective interests, can be reasonably informed on the achievements of public programs (83).

The Federal Government's program evaluation policy (Treasury Board Policy 1977-47) was issued after SPICE had already completed audits in many departments. In the policy, the implicit requirement in the Auditor General's Act for departments to evaluate their programs became explicit Government policy. The policy stated:

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 of Canada, 1977).

The manner in which the program evaluation function evolved and the rate of developments can be credited to the Office of the Comptroller General. The Office established a Program Evaluation Branch which provided: guidelines for planning and conducting program evaluation, advice and assistance to departments through liaison officers, and training courses to program evaluation managers and practitioners. The Office of the Comptroller General also influenced government policies to provide program evaluation in departmental policy and expenditure plans which are reviewed by Cabinet Committees.

Five years after SPICE, another government-wide audit was conducted on the program evaluation function. The findings in the 1983 Annual Report of the Auditor General noted considerable progress in the *level* of evaluation activities:

- all 19 audited departments had established corporate evaluation units.
- the 19 audited departments used 168 person-years of staff time and spent just over \$3 million on consultants.
- there were 211 completed initiatives (113).

In contrast to these developments, the Canadian Comprehensive Auditing Foundation's study on program evaluation which I directed found that 77 of 220 provincial departments had evaluation units and their combined expenditures on evaluation studies were one-third less than the amount spent by the Federal Government (1983:pp. 11-12). The Auditor General's 1983 Report expressed some concern about the *quality* of the evaluation work completed by Departments.

A substantial number of the evaluation studies had significant weaknesses in the methods to carry out the planned work, particularly with respect to measuring program effectiveness (126).

Nielsen Task Force

Debate on the usefulness of program evaluation has recently been fueled by the Ministerial Task Force on Program Review (commonly referred to as the Nielsen Task Force). It was established on September 18, 1984, just one day after the new government was sworn in. The Task Force reviewed 989 federal programs and services, reflecting more than \$92 billion in annual federal expenditures, including tax expenditures. Its report was made public on March 11, 1986 (Task Force on Program Review; 23-24).

Program evaluation was one of the six generic issues covered under the highlights of the program review. The critical nature of the Task Force's comments are reflected in the heading for this section — "Inadequate Program Evaluation". The initial statement was: "Many study teams found that routine government program evaluations were generally useless and inadequate for the work of the program review" (23). The basis for this conclusion was that evaluations conducted for the department's deputy minister are, by definition, self-serving. As a result, evaluations tend to avoid examining the fundamental rationale for programs. The Task Force's view was that proper zero-based program evaluations are not a function of government (24).

Issues related to the Nielsen Task Force's conclusions are covered in the following sections of this article. However, considering the seriousness of these concerns, some direct responses to the Task Force warrant mention in this article which reviews federal level evaluation.

The observation that program evaluations to-date have not addressed the life-and death questions of programs appears valid. The clients for evaluation studies preferred to have other types of issues examined. Moreover, there is some justification for the view that the current structure within departments for carrying out fundamental reviews of program rationale may be inappropriate. However, this does not mean that program evaluations were "generally useless". The evaluations examined by the Task Force were not conducted for its use and it would therefore be unreasonable to expect these studies to meet its needs. The usefulness of evaluations should be judged by the "real" clients for the evaluations. In this regard, there are other criteria besides those used by the Task Force to assess the usefulness of evaluations, including:

- making programs more responsive to client needs;
- increasing efficiency of program delivery;
- shifting resources among programs;
- influencing policy decisions; and
- serving accountability requirements for senior management.

The above uses of program evaluation should not be considered trivial. Even in relation to cost savings, in practice the savings from studies that serve such purposes may exceed those from studies of program rationale which in theory could result in the greatest savings. The explanation lies in the practical difficulties of terminating or seriously reducing program budgets.

The Nielsen Task Force's conclusions is unduly negative about program evaluation on the basis that the issue of program rationale has not been sufficiently addressed. The risk of the sweeping generalization is that it directs attention to defending the overall program evaluation function rather than addressing some legitimate shortcomings. The remainder of this article aims to shed light on some issues and indicate how their resolution would contribute to more useful program evaluation.

This paper enters the current debate on the role and value of program evaluation by addressing important issues under three broad concerns:

1. lack of clear focus for program evaluations;
2. insufficient emphasis place on performance measurement systems compared to periodic evaluations; and
3. an anti-rigour orientation toward evaluation.

Focus For Program Evaluation

A clear and specific focus for an evaluation is essential to concentrate limited resources and produce detailed analysis on the normally complex questions addressed. Otherwise, there is a risk of producing a superficial evaluation that provides little insight or direction to decision makers.

Key elements of the Federal Government's policy on program evaluation which account for the focus of program evaluation efforts to date are:

- Deputy Minister as client;
- Multiple purposes; and
- Categories of evaluation issues.

Deputy Minister as Client.

In the Federal Government, the deputy head is considered as the client for program evaluation. This is understandable because the deputy is ultimately accountable for the management of programs and has the authority to ensure that evaluations are conducted and used. Making the deputy minister the client is likely the reason for the tremendous progress in establishing evaluation functions in departments and agencies.

In practice, however, it is rare for a deputy head to act in the role of primary client in respect to: defining the purpose(s) of the study, identifying and prioritizing evaluation issues, approving terms of reference, reviewing findings, and addressing recommendations. This is unrealistic and probably unnecessary in departments having numerous programs undergoing evaluation each year. Important decisions pertaining to an evaluation are made by assistant deputy ministers, program managers, evaluation officers/directors, Steering Committees, and others. The scope and focus of the evaluation is determined through the interplay of various stakeholders and the broad range of issues selected for an evaluation reflects the attempt to accommodate the various stakeholders. Consequently, issues are included that are of little interest to the deputy and there is the exclusion or insufficient emphasis paid to issues that would be of interest.

It is obvious that there can be many clients for a particular evaluation study. The failure to address the information needs of a particular client is that the evaluation covers a wide range of issues of interest to many stakeholders. Such overly ambitious studies often do not adequately meet the information needs of anybody. The identification of a primary client (the deputy head or others designated by him or her) for an evaluation would provide the basis for planning a focussed study.

Purpose(s) of Program Evaluation.

The stated purpose for federal level evaluation is to provide appropriate information on the results of programs in order to: make informed decisions on the management of programs, guide resourcing decisions, ensure accountability, and provide quality advice to Ministers. The purpose of the study helps determine the evaluation issues to examine and information requirements (i.e., type, amount and level of detail). For example, information on a program's outcomes might satisfy accountability requirements but be inadequate for program management purposes because it is essential to have information on how program delivery affects the outcomes for making management decisions.

The purpose for an evaluation is often not specified or alternatively, multiple purposes are stated for a particular study. The risk is that no particular purpose is adequately served. Specification of the primary client's purpose for the evaluation would facilitate the selection of particular issues and identification of information requirements for a focussed study.

Evaluation Issues.

The Federal Government's approach to program evaluation included four classes of evaluation issues under which specific evaluation questions are grouped.

1. Program rationale — Does the program make sense?
2. Impacts and effects — What has happened as a result of the program?
3. Objectives achievement — Has the program achieved what was expected?
4. Alternatives — Are there better ways of achieving the results?

Program evaluations are rarely focussed on the life-and-death questions included under program rationale. Presumably the continued existence of the program has not been considered a major concern among departmental bureaucrats who select the evaluation issues. This raises an important issue: To what extent is it appropriate for the Deputy Minister as client to be responsible for such evaluations? The crux of the issue is the presumed independence of the client for such studies.

Terms of reference for evaluation studies often require an examination of several questions for each of the four classes of evaluation issues. To a large extent this reflects the involvement of multiple stakeholders who can shape the evaluation and the lack of clear purpose for doing it. It also reflects an overzealous interpretation of government policy encouraging evaluations

to be concerned with all of the basic program evaluation issues. The specification of the purpose for an evaluation would provide the basis for determining whether each category should be covered and the emphasis each should receive. A broad range of questions can be considered for these categories in the planning stage of the evaluation study. However, a choice should then be made to focus on a more limited number of important questions, taking into account the available resources, time constraints, and the methodological requirements.

In summary, more focussed studies could be achieved by: (1) planning the study for a primary client while attempting to accommodate the interests of other stakeholders; (2) determining the primary client's main purpose for the evaluation study; and (3) selecting a restricted number of evaluation questions for study, leaving others for management attention, policy review, or future studies.

Insufficient Emphasis On Performance Measurement Systems

The Federal Government's approach to program evaluation emphasizes the conduct of periodic or "one-shot" studies to be conducted cyclically for all programs. In this respect, program evaluation is differentiated from ongoing performance measurement systems which was the subject of an earlier Treasury Board Circular (1976-25). These distinctions are elaborated in a publication titled *Internal Audit and Program Evaluation in the Government of Canada* (Office of the Comptroller General, 1979: 1-12). Numerous experts in the field do not share the Government's distinction between program evaluation and performance measurement systems, arguing that information for evaluating programs can be collected routinely and/or periodically (Glaser, 1973; Hatry et al, 1979; Rossi and Freeman, 1982; and Rutman, 1983).

Emphasis on periodic studies to the exclusion of performance monitoring can give rise to several problems:

1. There is an absence of baseline and time-series data for making causal inferences.
2. Information for decision making is likely needed more frequently than the cycle for evaluating the program (e.g., every five years).
3. Data collection for periodic studies cannot be easily integrated with program delivery. This makes it intrusive, difficult to implement and costly.
4. The data is collected from clients who are at different stages in their involvement with the program (e.g., recent recipients of grants are interviewed a few months after the award while older recipients may be interviewed three or four years after the award). The timing differences may influence the validity of recipients' responses.

Several policies at the federal level provide the basis for developing performance measurement systems. The Treasury Board Circular (1976-26) has already been mentioned. The preparation of Operational Planning Frameworks includes the specification of indicators for programs. Part III of the Revised Estimates provides for the routine reporting of program results.

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.

Anti-Rigour Orientation

Program evaluation in the United States had its roots in the social experimentation era which began in the mid 1960s. The "ideal" methodology consisted of experimental designs (or close approximations) and quantitative measurement. The leadership in the field came mainly from the academic community and many of the major consulting practices such as the Stanford Research Institute were linked to universities or involved large numbers of academics.

Program evaluation in Canada does not have the same academic roots. The developments came largely from within the Federal Government. The federal approach emphasized relevance, practicability and timeliness. The Guide on the Program Evaluation Function states:

" . . . program evaluation, as part of this decision making and management process, should not be seen as an exercise in scientific research aimed at producing definitive "scientific" conclusions about programs and results. Rather it should be seen as input to the complex, interactive process that is government decision making, with the aim of producing objective but not necessarily conclusive evidence on the results of the programs. (Treasury Board of Canada, 1981: 4).

An emphasis on relevance is crucial for program evaluation to be useful. The thrust of the Federal Government's approach to focus on evaluation issues identified by senior management helps ensure the relevance of an evaluation study.

I have argued elsewhere about that the lack of useful evaluation studies is at least partly due to the inappropriate use of "rigorous" methods to measure impacts and effects under circumstances when the program was: ill-defined or not properly implemented, the objectives were vague, and the expected impacts were beyond the reach of a program (Rutman, 1980: 42-58). I support Grave's position that:

In many respects, this shift in the evolutionary direction of program evaluation has been beneficial. The correction for the excesses of the initial positivistic model of program evaluation research was both necessary and timely (1984; p. 42).

I sense 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. This orientation is often expressed by clients' terms of reference for evaluation studies as well as evaluators' proposed approaches for carrying out such assignments. In this regard, I share Graves' view that:

. . . many of the present proponents of relevant program evaluation have lost sight of the initial *telos* of research. Quite simply, information

may be timely, economical, lucid and germane to decision making, but it is really of no value if it is wrong (1984: 42).

The importance of rigour is evidenced in the counter-intuitive findings that numerous well-conducted evaluations have produced (often showing the ineffectiveness of programs held sacred by professionals) and the differences in results (usually more negative and disappointing) from studies which measure behavioural changes of clients rather than perceptions or viewpoints of stakeholders.

The usefulness of program evaluation comes not only from the independence and objectivity of evaluators but the assumed validity of the findings and conclusions. This is not meant to imply that rigorous evaluations require experimental or quasi-experimental designs, only quantitative data, and complex statistical analysis. Rather, a broad range of "quantitative" and "qualitative" approaches can be used to produce factual information for decision-makers.

Decision makers assume that the information produced by an evaluation study is reliable and valid. We need to strive to achieve the highest degree of *practical* rigour, not search for excuses to avoid it or inappropriately engage in excessive rigour.

Conclusion

This review of the evaluation function at the federal level attempted to identify achievements as well as issues which I believe should be addressed. It's important to recognize that the program evaluation function is still in its infancy. There have been remarkable accomplishments in developing an infrastructure for useful program evaluations. There is also some evidence that studies are being used by decision makers (Soper et al, 1984: 65). The viewpoints in this article aim to build upon these achievements and contribute toward a more effective role for program evaluation.

References

- Canadian Comprehensive Auditing Foundation. Value-For-Money Management: A Preliminary Study of Program Evaluation in Canada, 1983.
- Glaser, Daniel. *Routinizing Evaluation*. Rockville, Md.: Center for Studies of Crime and Delinquency, NIMH, 1973.
- Graves, Frank L., "Towards Practical Rigour: Methodological and Strategic Considerations for Program Evaluation", *Optimum*, Volume 15-4, 1984.
- Hatry, Harry P. Richard E. Winnie, and Donald M. Fisk. *Practical Program Evaluation for State and Local Governments*. Washington: The Urban Institute, 1981.
- Office of the Auditor General. *100th Annual Report of the Office of the Auditor General*. Ottawa: Supply and Services Canada, 1978.
- Office of the Auditor General. *Report of the Auditor General to the House of Commons*. Ottawa: Supply and Services Canada, 1983.

- Office of the Comptroller General. "Internal Audit and Program Evaluation in the Government of Canada," 1979.
- Rutman, Leonard and Dick de Jong. *Federal Level Evaluation*. Ottawa: Carleton University Press, 1976.
- Rutman, Leonard. *Planning Useful Evaluations: Evaluability Assessment*. Beverly Hills: Sage Publications, 1980.
- Treasury Board Policy (1977-47) "Evaluation of Programs by Departments and Agencies", 1977.
- Treasury Board of Canada. *Guide on the Program Evaluation Function*. Ottawa: Supply and Services Canada, 1981.
- Rossi, Peter H., Howard E. Freeman, and Sonia Wright. *Evaluation: A Systematic Approach*. Beverly Hills: Sage Publications, 1979.
- Soper, Nancy, Kathleen Petersen, Neil Maxwell and Rona Shaffron, "Facilitating Use: The Canadian Experience", *Optimum*, Volume 15-4, 1984.
- Task Force on Program Review. *Introduction to the Process of Program Review*. Ottawa: Supply and Services Canada, 1986.