

In Defense of Program Evaluation

John Mayne

President

Canadian Evaluation Society

To the Editor:

On March 12, 1986 and for the next few days, newspapers in Canada were full of stories on the *Final Report* of the Ministerial Task Force on Program Review (Task Force on Program Review, *Final Report*, 1986). Many of those articles had disparaging things to say about federal program evaluation: references to the general poor quality of evaluations and "useless" and "self-serving" evaluations abound. The articles by Christopher Waddell and Hugh Windsor on March 12, 1986 and by Bruce Little and Jeffrey Simpson on March 13, all in the *Globe and Mail* are examples.

Getting some attention in the media as the result of the findings of a particular evaluation study was one thing, but this seemed not to be the kind of attention our fledging profession needed. What was going on?

The newspapers were reporting the comments made on federal program evaluation in the Summary volume of the 21 volume *Final Report* of the Task Force, entitled *An Introduction to the Process of Program Review*. In March 1986, the Ministerial Task Force, more commonly referred to as the Nielsen Task Force after the then Deputy Prime Minister Erik Nielsen who was its Chairman, made available the findings from its review of 989 federal government programs which had taken one and a half years to complete. As part of its summary, the Task Force identified seven generic problems it found in carrying out the review, one of which was the reportedly sorry state of evaluation in the federal government:

"Many study teams found that routine government evaluations were generally useless and inadequate for the work of the program review." and further

"By definition... they tend to be self-serving, even though the scheduling and methodology of these evaluations are subject to the approval of the Comptroller General." (Task Force on Program Review, *An Introduction*, 1986:23)

For those who have been working in evaluation in the federal government this was disappointing if not startling news. There were in 1985/86 some 350 federal evaluators who had completed about 350 evaluations since 1981 (Treasury Board of Canada, 1986). They were aided and abetted by almost 20 officers in the Office of the Comptroller General who had developed the federal policy on evaluation and assisted departments in getting on with evaluation. The Comptroller General in the first issue of this journal had written about the use being made of federal evaluations (Rayner, 1986). Could all of this effort be in vain? Is it possible that these federal evaluators were all so incompetent that their products were "useless", "inadequate" and "self-serving"?

The answer, it seems apparent to me, is no and the evidence for that judgement is available in the Task Force reports themselves.

What was this Nielsen Task Force and what were they trying to do? Immediately into its new mandate, Prime Minister Brian Mulroney announced that the federal government intended to review the bulk of government programs to make sure that they were consistent with the new government's priorities and to rationalize the existing base of programs by identifying duplication, waste and inefficiencies. A Ministerial Task Force on Program Review was established, chaired by Erik Nielsen and included Michael Wilson, Minister of Finance, Robert de Cotret, President of the Treasury Board, and John Crosbie, Minister of Justice. Further, a Private Sector Advisory Committee chaired by Phillip Aspinall of the Canadian Institute of Chartered Accountants was established to assist the Ministers in their review.

One of the innovative features of Task Force — and one that is now acknowledged by all to have been very worthwhile — was to include the private sector in the study teams that were established to assemble the review material for consideration by the Ministerial and Private Sector Committees. Groups of government programs were identified such as Agriculture, Services and Subsidies to Business, Natural Resources, Procurement, Veterans Programs, Regulatory Programs and many more. The complete list can be found in the Summary volume. There is also a volume devoted to each program area and each contains the names of the study group members and their affiliation.

For each program area, a study team of between 6 and 20 persons was formed consisting of public servants and members from the private sector in equal numbers. Some study teams were headed by public servants and some by persons from the private sector. The Task Force had few resources of its own. The team leaders seconded public servants (usually not from the areas being reviewed) and acquired the private sector members who worked without remuneration. The program areas were reviewed in sequence over the one and a half years, with typically 3 to 5 teams operating at any one time and each study team taking 3 to 4 months to complete its report. The summary volume contains a more complete description of how the Task Force was organized and operated.

The teams reported to the Private Sector Advisory Committee and the Committee of Ministers as the reviews were completed. As a result, the government was able to report on the results of four of the study teams at the time of the May 1985 budget (Government of Canada, 1985). The complete set of reports were released in March of 1986.

This whole endeavor was perhaps the largest and most concentrated review of government programs ever mounted. Its history has still to be written but it is clear that it was, in effect, a large evaluation of government programs concentrating on their continued relevance and having clear political leadership. Coming as it did in 1984 and 1985, three and four years after the federal evaluation system had been in place one would have expected that it would be able to build on this base of evaluations. And as we shall see, it very much did so.

The mandate of the study team was to produce an inventory of government programs with special concern for identifying duplication, waste and inefficiencies. It is our hope that study team proposals to the Task Force will produce a system that, first, is simpler to understand and more accessible to the Canadian public and, second, contributes more effectively to the achievement of national economic and social objectives. (Task Force on Program Review, *An Introduction*, 1986:i)

In practice this meant that the study teams challenged the continued need for each program. Questions asked were: Why is this program needed? Who is actually benefiting from it? Why does the government have to provide the good or service? Is the program duplicating some other government activity such as a tax expenditure or regulation? The study teams frequently did not find answers to these questions nor discussions of these issues in the evaluations that they reviewed. Hence their criticism.

Behind these expectations would appear to lie a model of evaluation that differs from that being used in the federal government. The Task Force, as might many others, seemed to expect that evaluation studies should address a set of predetermined questions, such as continued rationale and effectiveness. Furthermore, they expected that the evidence provided in evaluations would aim for, even if not always attain, a consistent high quality of credibility in the eyes of readers of the reports. This all sounds quite reasonable and indeed is what many might assume evaluation to be all about. And it is a legitimate model of evaluation. But it is not the one that had been pursued in the federal government.

Evaluation in the federal government is designed to meet the information needs of a client at a specific point in time. That is, evaluations should indeed serve some well-identified client. This approach is based on the ample historical evidence that evaluations where the evaluator decides what questions to address or where the evaluation tries to meet the information needs of all potential clients are either not feasible or are not used, even if producing 'good' research. Rather, to be used, evaluations must focus on the issues facing the program as defined by the study's client at the time of the evaluation. In the federal government, the primary client of evaluations is the deputy head. This is the person who can decide significant questions about the program and who is expected to synthesize the numerous claims and viewpoints about the program. This is the person who can best ensure that the Minister's interests are reflected in the evaluation. This is also the person who can best decide on the level of credibility (and hence the level of effort required) that is needed in order that the findings can be acted upon. Yes, by definition, the evaluations are meant to serve the deputy, but not to be "self-serving".

Furthermore, in keeping with this approach, and unlike reported in the Summary volume, the evaluations were *not* subject to approval by the Comptroller General nor by his Office. Rather, that Office provided advice and assistance to departments.

Thus the evaluations reviewed by the study teams should have been addressing issues of concern to the deputy at the time of the evaluation. Ob-

viously, these may not have been the issues of interest to the Task Force, and indeed the Task Force report says that they found the evaluations inadequate *for the purposes of their review*, a caveat dropped in the media articles. The Task Force had a very high level focus based on the perspectives of the new government and took place in 1984 and 1985. The evaluations took place earlier – in many cases several years earlier when the evaluation functions in departments were just getting started – and were not motivated by a political desire to seriously review government programming from the perspectives of a new government in power. At the time the evaluations were done, the issues addressed were seen, presumably, as legitimate as those of the Task Force. Thus, while the expectations of the Task Force were understandable, they were destined not to be met.

Having said this is not to say that federal evaluations are, by definition, good quality evaluations, i.e., they always address the right issues, use the right methods and fairly report their findings. Many of the studies reviewed by the Task Force did avoid hard questions about the program and others had less than credible information in them. Overall the quality, perhaps not surprisingly, was mixed. The Auditor General had reported as such in his 1983 *Annual Report* (Auditor General of Canada, 1983: Chapter 3). Some of the poor quality is due to the newness of the evaluation function, some due to poor quality work done by evaluators, some due to inadequate oversight by the OCG, and some due to lack of demand by central agencies, Cabinet Committees and line Ministers to address the hard questions. All are real problems and all need to be improved. But were the evaluations as bad as the Summary volume seemed to indicate?

A review of the other 20 volumes of the Task Force suggests otherwise. Almost all of the reports have some reference to the evaluations reviewed. In many instances the conclusions of an evaluation are quoted . . . presumably an indication of agreement. In many instances the recommendations of the Task Force reports are those of evaluations. In many instances, the reports call for further evaluations before decisions are to be taken. And yes, there are numerous references to the lack of an evaluation (and hence no useful information available on the program?). All in all there are well over 100 references to evaluations. Not bad for inadequate and self-serving evaluations!

Two of the study team reports make specific mention of the utility of the evaluations they reviewed. The *Services and Subsidies to Business* report states that:

First, program evaluations vary greatly in quality. Some are solid guides to understanding and reform. Some are little more than self-serving advertisements for more resources for pet departmental programs. Even where methodologically sound at the micro level, however, they seldom (a) examine alternative ways of accomplishing the same ends or (b) present the benefits of a program net of the good that would have accrued had the resources been allowed to percolate around in the private economy. The theology that has made evaluation a servant of the deputy minister rather than the system as a whole, an internal art not reviewed by outsiders, and an adjunct of

the audit rather than policy development function in most departments, needs re-examination. (1986: 18)

Despite these criticisms, a reading of their report indicates that this study team did make extensive use of evaluations they found. Evaluations are referred to, quoted from and conclusions and recommendations from evaluations are used. Nevertheless, they clearly expected more. Perhaps unfortunately for evaluation, this study team was the first to finish their work and these summary views on evaluation became accepted as the general conclusion of the Task Force and may indeed have biased the overall judgement on evaluations.

A somewhat different view is found in the *Job Creation, Training and Employment Services* report:

In carrying out its work over the past two months, the Study Team has drawn on a number of information sources, including internal departmental evaluations which have been available for most of the programs on our list. In our view, these evaluations could be improved in several ways: first, we have observed that they are generally not focussed directly on expenditure reduction; second, they often do not satisfactorily take into account the perceptions of the clientele concerning the programs; and third, the evaluations are sometimes out-of-date. However, notwithstanding these deficiencies, we wish to stress that, in our view, we have found the internal departmental program evaluations to be very useful and this report would have suffered considerably without them. (1986: 9)

The severe criticism in the Summary volume may indeed reflect the collective view of the study teams. However, the evidence as provided in the Task Force reports would seem to suggest otherwise; namely, that while some were of poor quality and many did not address the key issues of concern to the Task Force, many of the evaluations were used and useful to the Task Force. And this from evaluations that were done for a different client who had different concerns at an earlier point in time! How much more can one reasonably expect?

What can one conclude? It is fairly easy to show that the reports in the press were somewhat inaccurate in their description of what was said and of what could be found in the reports. Evaluations were used and useful to the Task Force. It is also clear, however, that the federal evaluations did not serve the Task Force as well as they could have and, indeed, should have. Given this experience, if another Nielsen type Task Force is mounted, the evaluations now being conducted probably will prove to be significantly more useful. But it is also clear, it seems to me, that those evaluations, no matter how good they become, will not, cannot and should not replace the reconsideration of programs by Ministers of the base of programs they feel their government should be providing to the public. Evaluation, the analytical exercise, should never be a substitute for evaluation, the management function of reconsidering programs as to their continued usefulness.

A more generic issue emerges out of this commentary, one that should be of interest to all evaluators. Should evaluations cover a given set of issues? Should evaluations only use state of the art methods and always be as definitive as possible? Should a reader be able to pick up an evaluation and expect to find a certain type of information? You know what your going to get out of a financial statement and sort of what you are going to get out of an audit . . . operational problems identified. Should there be a parallel with an evaluation report? Can evaluations both serve management and provide a basis for public accountability?

References

- Auditor General of Canada, 1983 Annual Report, Ottawa, 1983.
- Government of Canada, *New Management Initiatives, Initial Results from the Ministerial Task Force on Program Review*, May 1985. Tabled by the Hon. Erik Nielsen, Deputy Prime Minister.
- Rayner, Micheal H. "Using Evaluation in the Federal Government", *Canadian Journal of Program Evaluation*, Vol 1, No. 1, pp. 1-10.
- Task Force on Program Review, *An Introduction to the Process of Program Review*, Supply and Services Canada, Ottawa, March 1986.
- Task Force on Program Review, *Final Report*, 20 volumes, Supply and Services Canada, Ottawa, March 1986.
- Task Force on Program Review, *Job Creation, Training and Employment Services*, Supply and Services Canada, March 1986.
- Task Force on Program Review, *Services and Subsidies to Business*, Supply and Services Canada, March 1986.
- Treasury Board of Canada, *Comptroller General, 1986-87 Estimates, Part III*, Supply and Services Canada, 1986.