

## Using Evaluation In The Federal Government

Michael H. Rayner  
Comptroller General Of Canada

### ABSTRACT

L'activité d'évaluation a connu une importante intensification au sein du gouvernement fédéral, mais qu'elle a atteint un plateau, ces dernières années. D'un autre côté, certains critiques estiment que l'évaluation n'a pas contribué suffisamment aux efforts qui sont déployés actuellement pour réduire la taille du gouvernement fédéral. Pourtant, l'évaluation présente tant de possibilités de contribution utile que l'auteur demeure très optimiste quant à l'expansion de l'activité, de l'utilité et de l'influence de la fonction. Les gestionnaires axiaux peuvent recourir à l'évaluation pour améliorer le fonctionnement des programmes; les ministères peuvent l'utiliser pour assurer la réforme des programmes; ministères et organismes peuvent s'en servir pour la répartition stratégique des ressources; et on peut l'utiliser pour l'élaboration des politiques, le contrôle de la responsabilité à tous les niveaux, ainsi que pour la recherche. Certes, ce ne sont pas toutes et chacune des études qui peuvent fournir une contribution de toutes ces manières; néanmoins, l'auteur est convaincu que bon nombre d'évaluations demeurent trop restreintes quant à leur portée, parce qu'on ne prévoit pas les nombreuses manières d'utiliser les résultats des études. À l'aide de cinq exemples d'évaluations fructueuses, l'auteur montre qu'on a fait du bon travail à bien des égards. L'on a déjà lancé les initiatives nécessaires pour susciter la demande d'évaluation dans plusieurs des contextes susmentionnés. Les évaluateurs fédéraux feraient bien de prévoir cette évolution et de produire les quantités voulues d'évaluations utiles.

I believe in the evaluation of government programs. I feel that evaluation is crucial at both the operational and strategic levels of government management and decision-making. It has the potential to complete the line of accountability from bureaucrat to politician to the public. But evaluation does not always deliver, and because of this, the function is not safe from the cold winds of cutbacks that are blowing continually through all our governments.

Since my appointment as Comptroller General, I have had no shortage of advice on how to "save" the evaluation function:

- Some say make it popular with line managers by doing free management consulting.
- Or get in good with the Deputy Ministers (the top public servants in each department) by doing the special jobs – be his or her special SWAT squad.
- Others say align the function more closely with the Policy area – that's where the action is.

\* (This article is based in part on a speech given to the joint conference of the CES, the Evaluation Research Society and the Evaluation Network in Toronto on October 17, 1985.)

- Or concentrate on serving the information needs of central agencies — that's where the power is.
- People also suggest that evaluation should become an instrument of accountability to Parliament.
- Some say be the one to deliver the quick and dirty analyses that are required in a crisis.
- Others suggest that evaluators should be more rigorous with higher standards and more professionalism.

It all sounds good to me. And I say "go to it", although I am aware of the conflicts among all these ends. Can you serve line management and central agency needs simultaneously? Would either side trust you if you tried? Can you be both strategic and tactical? Policy-oriented and operations-oriented? Quick and rigorous? Timely and timeless? This is probably impossible in any single study or report. But it is not impossible for the function as a whole. As the examples in this article illustrate, evaluation can be useful in many ways.

In Canada, at least at the federal level, we are at a seeming "plateau" in the development of the evaluation function. I will describe that situation more fully in a moment. In the provinces and municipalities, despite many pioneering efforts, evaluation is not catching on universally. In recent years in the United States, there has been a downgrading of the resources available for evaluation, if not of the respect for the function itself. Is the bloom off the rose? I hope not. It doesn't have to be. Evaluation can thrive if it is being used. There are so many ways to use evaluations that, by keeping open the channels to all potential users, and through plain hard work, the promise of evaluation for better programs and better government will be realized more fully.

Let me list some of these many uses for evaluation. I have seen evaluation used for operational improvements to programs, for fundamental redesign of programs, for departmental or agency-level resource allocation, for resource allocation by central agencies, for policy formulation and for accountability to the legislature and to the public. It can be useful as well in an academic context. I want to see more of all of these uses. I think, in fact, that evaluation can be made *so useful* that any talk of decline or demise will just fade away.

My examples and experiences are drawn from the Canadian federal context, and I am aware of the institutional differences with other governments and agencies. My contention however is that useful, successful evaluation ought not to depend on any particular set of administrative arrangements. To be sure, it can help to have the right people listening, acting, paying the bills, and so on. But our institutions in Ottawa are not perfectly set up for the convenience of evaluators. They probably never will be. This year could prove to be a time of real ferment and change in the authorities, responsibilities and accountabilities of our federal departments and agencies. I hope that a better climate for evaluation will come out of this. For my part, my office will be working as hard as it can to improve systems and attitudes.

Where are we now in evaluation in the federal government? One re-

sponsibility of my office is to see to the establishment of the evaluation function. Program evaluation has been delegated to departments as something they must do for themselves. Provisions have been made to promote objectivity, credibility and utility inside and outside the department. We in the Office of the Comptroller General do not do evaluations ourselves except, of course, for our own programs. Instead we monitor the function in departments and advise on infrastructure, issues and methods. We also assume a quality assurance role for other users of evaluation in the central decision making bodies. Our central position allows us to see the patterns and trends in the function across government.

It is clear that the rapid growth phase is over, as almost all major departments have established an evaluation capacity. There are over 300 public servants working in the function and lots of consultants making a living at this work. Before my current job, I was a partner in one of those consulting firms so I know both perspectives. By and large, the federal program evaluation community is competent and I'm glad there are so many talented people interested in the field. The quality and technical sophistication of the studies are improving but remain uneven. The rate of production is holding steady at a bit under 100 studies a year. This is about half the rate we need if virtually all government programs are to be reviewed at least every five years. This slow rate reflects inevitable slippage on studies in the most conscientious departments, and feeble or no effort in the remainder. The required doubling of activity is going to be very hard to achieve because, as with most North American governments, we are going through a period of freezes and cutbacks.

In fact, one of the loudest criticisms of the evaluation function in Ottawa has been that it is not enough help in the drive to "down-size" government. This alleged lack of usefulness might make evaluation itself a worthy candidate for the chop. I will go along halfway with that argument. I too want to see more use of evaluation in decisions to privatize, to delegate, to reduce or otherwise trim government programs. After all, those decisions can only be improved by the availability of information on the actual results of the programs. In the past we have too infrequently asked the hard, basic questions about program relevance and continuation. We have focused instead on improvements and enhancements. And so I, as Comptroller General, have asked departments in their evaluations to include investigations which reflect that economizing philosophy. They should ask what can be done with *fewer* resources as well as what *more* resources would do to enhance effectiveness. It is more work, but evaluations that ask both kinds of questions would be useful in more contexts. We need more strategic thinking in the design, execution and use of evaluations.

Making more evaluations more useful to more people is the key to the survival and growth of the function. Earlier I gave a list of uses for evaluation and I promised some examples. I mentioned operational improvements to programs, fundamental redesign of programs, departmental resource allocation, resource allocation by central agencies, policy formulation, accountability to the legislature and public, and academic research. I will begin with an example at the operational level. Evaluation should and can

be useful to the line manager in improving the operation of his or her program.

The federal Department of National Revenue Customs and Excise completed an evaluation of the Carrier Declarations part of its operations in 1984. This area spends about \$22 million a year and employs more than 700 people. As goods arrive and leave Canada by rail, truck, ship or plane, carriers must declare each shipment for possible duties or taxes, for trade statistics and for control requirements. The area for this study also included customs processing of international mail.

The main issues were primarily operational: achievement of the operational objectives of carrier compliance and export declarations, client satisfaction and efficiencies concerning resource allocation.

The study found a lot that was right about the operation but I will focus on the problems. The evaluators noticed, for example, that different modes of transportation were getting different amounts of scrutiny and were subject to different controls and paperwork requirements. They noted responsiveness problems with respect to service needs in the postal declarations area, along with a significant potential for productivity improvements and staff reductions on the order of 100 person-years. On the other hand, the export declarations area seemed to need more resources. Finally, in both export and import carrier areas, some opportunities for improving enforcement verification and controls were noted.

This evaluation is resulting in the rationalization of procedures and paperwork for carrier declarations, some reallocation of staff, better client service systems in the postal declarations area, and improvements in verifications and controls. Nothing shocking or scandalous. Nothing that even surprised the program's management. What the study did was document some problems while confirming that many other aspects were operating satisfactorily. This has enabled program management to take action on the necessary changes *with confidence*.

Although this study had criticisms, it also had some good news and was extremely useful to line managers. To them the study amounted, in many ways, to a bit of free management consulting but with visibility to senior management. This is certainly a model worth emulating and in fact is fairly typical of a large number of operationally-oriented evaluations that have occurred in the federal Government. It is something we are doing reasonably well in many departments. And many departments would be content to consolidate their evaluation functions at this operational level.

Another example of an evaluation is a little further down the spectrum from operational changes to more strategic program reconsideration. The Department of Energy, Mines and Resources administered the Canadian Home Insulation Program, perhaps best known by its acronym "CHIP". This program provided a taxable grant to homeowners to defray part of the costs of upgrading insulation and air-tightness. It has provided about \$1.5 billion in grants over almost ten years.

There were many interesting findings from the evaluation. It was found, for example, that the program was effective in getting homeowners to take action but that most of the work was being done on attic insulation. This is

a good place to start, especially in older homes, but there is also a large potential saving in such areas as weather stripping, caulking and wall insulation. Insulation contractors liked the relatively easy attic jobs and this is what they promoted. Even after seven years of operation there were still problems with the awareness and image of this program. In its first six years, about 44,000 person-years of employment had been generated. A cost-benefit ratio was conservatively estimated for CHIP at about 1.26 although a wide range of values could be generated by varying incrementality assumptions.

These findings helped the department make a number of decisions to alter the program fundamentally. In response to a recommendation that policies be developed to prevent fraud, the incentive was reduced substantially, requiring the homeowner to pay more of the costs. Standards were raised to make sure that the effort would result in significant savings. A new marketing approach was installed and substantial improvements were made to program management, administration and delivery. The reduced incentive encompassed a greater range of eligibility and slanted the program to those houses where the savings would be the greatest. It was expected that reduced participation in the program would save tens of millions of dollars in the first year alone. Thus the cost-effectiveness of CHIP was substantially increased. More recently, the program has been scheduled to phase out altogether and is now in its last year of operation.

This evaluation gave the department, and later the central agencies, reliable knowledge of what the program had accomplished and what it could expect to do in the future at what possible costs. They could make fundamental changes, not merely marginal shifts, in the program's design and philosophy with confidence. Later they could know reliably what benefits and costs they were passing up by phasing out CHIP.

In a third example, evaluation can be useful not only in improving programs but also as input to a decision as to whether or not the program should continue at all. The Department of Consumer and Corporate Affairs has been administering a program called ENERGUIDE. It provided for a standard label to be placed by manufacturers on each of six major home appliances such as refrigerators and dryers. The label gave the estimated monthly energy consumption under a laboratory simulation of typical use. A listing of the ratings for all makes and models sold in Canada also was published.

The program cost the government only about five hundred thousand dollars to run chiefly for testing manufacturer claims, but industry said it cost them millions. Conceived during the energy crisis of the last decade, the program was originally scheduled to end in 1986 but plans had been prepared for its renewal and extension. Evaluation was needed to provide balanced and objective information with respect to those plans and other options for the program.

The evaluation found that the program had been successful in that appliances are now generally much more energy-efficient. However, this success also tended to argue for the end of the program since there was little room left for economically sensible increases in efficiency in modern designs for

major appliances. Furthermore, consumers apparently did not understand or use the information. Neither did retail salespeople. The manufacturers were not supportive, chiefly because of the costs of testing. Even consumer advocacy groups had criticisms of the current implementation; they had their own ideas on how the program should be delivered.

The evaluation was able to conclude that neither continuation of the program nor any of several possible modifications would provide good value-for-money. These conclusions allowed the Department to decide on the termination of ENERGUIDE with confidence that it was doing the right thing. Consumer and Corporate Affairs no longer funds testing or publishes the handbook. At this stage, however, manufacturers are still required to affix the ENERGUIDE stickers. These regulations are under the jurisdiction of another department.

In the three above examples, the departments or operators of the programs used evaluations *operationally* to improve their programs and *strategically* in the allocation of resources. Another example illustrates how the central planning and budgeting agencies can use evaluation to help in their strategic decisions. Some years ago, the Treasury Board commissioned a special Task Force on Service to the Public which was attached to the Department of Supply and Services. This group was outside the regular administration of the Department as a deliberate strategy to make it innovative. Many of the Task Force staff were seconded from other areas and many contract personnel were used to run the various experimental services that were devised. At the time of the evaluation, expenditures were running at several million dollars annually with pressures to grow quickly.

Previous surveys had revealed a lot of general public dissatisfaction about contacting the government. So the Task Force tried walk-in or store front service bureaux for those who don't like to use the phone. There were mobile versions of the service bureaux to serve rural areas. Telidon or interactive video display units were installed in shopping malls and public libraries. These allowed access to up to 50,000 pages of information about the government on everything from the weather to job listings to tourist services. An effort was made to consolidate and improve government telephone listings. Telephone referral services were set up, often in cooperation with provincial authorities, so that the caller got a knowledgeable operator who could give the right number for the caller's problem right away. An index of all federal programs and services was compiled for the use of both the public and those front-line public servants who have to handle the public's questions. The Task Force also set up an advisory service to work with staff in all departments to improve their relations, service and image with the public. Funding was on a year-to-year basis with frequent demands for evaluation of the various projects.

The Task Force used to claim, with some justification, that the experiments were always being changed and had just started up, making evaluation inconvenient or inappropriate. Finally, the Cabinet and its supporting central agencies decreed that the Task Force would be evaluated and then wound down with the successful parts to be incorporated into the regular Public Service in appropriate departments. The conflict here was that the

Task Force viewed itself as a continuing permanent program which was just getting started, while the central agencies thought of it as a one-shot experiment. The Task Force therefore submitted a counter proposal that it be made a permanent feature of the federal bureaucracy and given a mandate to continue its good works.

The Task Force hired several consulting firms, including my former firm, to examine the various projects. Unfortunately, no overall examination of the need for a single co-ordinating body was undertaken. And no investigation of its effectiveness as a whole was attempted. Instead, the services were looked at individually.

It was found that the Task Force was providing service which ranged from fair to excellent. Sometimes they were serving reasonable numbers of people with somewhat encouraging results. Often, however, the service was of only fair quality or was reaching only a few of the Canadians who were supposed to benefit from these services. A lot had been learned about the means and costs of improving service to the public. These were good results for experiments but were not compelling reasons to continue to fund the experiments.

For example, the telephone referral services were doing a good job for their callers but were not extensively used since they had relatively few staff, and the project managers were reluctant to publicize the service for fear they would be swamped. Staffing up to handle many more calls would have been beyond the means of the Task Force. Similarly, the walk-in and mobile service bureaux were appreciated by their clients but the Task Force could not afford to put a bureau in every community. The Index of Services was much admired but used only by a few. The Telidon/Cantel interactive video display kiosks were fun to experiment with but did not make an impact commensurate with the projected cost of installing a true national network.

One theme running through all the evaluations was that the services were being used disproportionately by those who were able, or ought to be able, to communicate with their government. The less educated, the less mobile, the less articulate were not being reached as much as might have been hoped.

So what to do with the Task Force on Service to the Public? Its staff wanted a shot at fixing up the deficiencies the evaluations had noted. But clearly the learning experience gains from the experiments were largely complete. It was now time to decide whether to spend the money for full-scale national implementations.

Despite good arguments about momentum and demonstration effects, Cabinet decided to give the Department of Supply and Services just one-third of the requested funds and only for one year's wind-up. The administration of the consolidated telephone listings was transferred to the Government Telecommunications Agency. The responsibility for developing Telidon/Cantel terminal demonstrations was integrated with other Telidon technology promotion under the Department of Communications. The service bureaux were closed. The telephone referral services continue in a low-key way with their provincial counterparts. Finally, the annual Index of Services continues to be published by Supply and Services.

This evaluation would have been improved if it had had a global dimension, looking at trade-offs among the services and the need for a central coordinating body. However, it provided solid evidence on performance and costs which was used in deciding the future of the individual initiatives. Otherwise, the decision would have rested solely on the anecdotes and promises of the program managers. Savings in the first year alone were \$6 million over what the Task Force proposed.

My final example is of an evaluation which has proved very useful in contributing to policy formulation. Canada Mortgage and Housing Corporation administers the Non-profit and Cooperative Housing program. The program subsidizes the mortgage payments of non-profit and cooperative housing developments in order that no more than the lower end of the local market rent is charged, and that some units be provided at a rent geared to the tenant's income. In 1982, subsidies amounted to \$177 million. The program applies to new or rehabilitated structures and to everything from single family housing to hostels, old-age homes and group homes. It was designed with the idea of mixing rent-to-income tenants with those who can afford at least the low end of market rents. The program was a deliberate departure from building tenements for the poorest people alone.

The evaluation showed the needs and demands for affordable housing are certainly still there and that the amount of accommodation encouraged by this program was only chipping away at the problem. The accommodation was neither over-crowded nor under-used in most cases. Disadvantaged groups were being served. The costs of construction were not always as low as they might have been but private sector resources were being mobilized and the supply of rental housing augmented. The program was not as cost-effective as other social or market housing initiatives, but because of the high mortgage rates of the early eighties, it provided the only examples of affordable housing construction in some years, and in some communities.

The most celebrated finding, however, was that while many tenants were poor by any definition, a good proportion of those living in these units and paying no more than the lower end of the market rental rate were people of moderate and even comfortable incomes. Remember that this was, in effect, what the program was designed to do — to mix incomes and avoid low-income ghettos. The evaluation put forward the figures that while literally hundreds of thousands of Canadians still needed affordable housing, another group of better-off people were in subsidized housing although they could afford to live elsewhere. The clash of the two objectives of helping the poorest and achieving an income mix was documented and quantified.

What this evaluation did was to outline the effects and contradictions of the program and ensure that the continuing debate over housing policy has a reliable basis for discussion. When the report came out, people in the cooperative housing movement saw it as an attack on their way of life (and their subsidies). My office received critical letters and pleas for intervention but we were satisfied there were few methodological problems. I think by now that all sides realize the evaluation was relatively neutral and the



debate ought not to range over statistics or definitions of what is moderate and what is low income. Instead, this evaluation settled those issues for this particular program. The Non-profit and Cooperative Housing evaluation was one of a series in which CMHC examined all its Social Housing programs. In December of 1985, the government announced an overhaul of all these programs. Two basic thrusts were better focussing of benefits to the poorest households and new funding mechanisms included rent supplements and indexed mortgages. Evaluation certainly cannot claim entire credit for these changes but I am confident an objective base was provided for analysis and policy development.

The final use of evaluation mentioned earlier was research. Evaluation can be used to increase basic knowledge about society and the economy and especially about the effects of government intervention. We certainly need such information. I do not think there is much support for a purely technocratic society, but we would all like our politicians to have relevant and reliable information on which to act.

Most of the evaluations with which I am familiar would have trouble passing tests of complete scientific rigour. Government decision makers may need to act sooner than all the protocols of good research permit. It is better that they have some evidence that confirms or challenges what they intend to do, rather than none at all.

People have advised me, however, that increased rigour is the prescription for saving evaluation. I support increased rigour whenever appropriate but not at the loss of timeliness. I am convinced that the consultant's 80/20 rule often applies. In a government decision-making context, 80 per cent of the value of an evaluation often can be had for 20 per cent of the effort by focussing on key issues. In the federal government, we recommend a two phase approach where an assessment is done before the study in order to develop a wide range of possible issues and methods. Some issues will be of lower priority. Sometimes there will be other information sources that cover certain issues adequately. Let evaluation bring its particular strengths, including acceptable rigour, to the remaining critical issues.

True research is probably going to have to be left to the academics. Synthesizing the results of several less rigorous studies on the same area or related ones can be useful. Some research can grow out of contracts to academics for evaluation but there probably will be a line beyond which the evaluation budget will not stretch and research funding will have to take over.

In closing, I would like to return to an across government perspective of how we are using, or not using, evaluation at the federal level.

My staff has been following up on evaluation studies across the federal government to see how they have, in fact, been used. In about 150 cases that have been tracked through completion to implementation, the dominant results are operational improvements or simply better documentation of the program. There are far fewer "fundamental" uses such as cancellation, expansion, program redesign or governmental policy change. This could indicate that the majority of programs are perfect or in need of only minor adjustment. That is not a very credible explanation. It could mean

that the rest of the bureaucracy ignored all the challenges and wonderful advice that these intense and wide-ranging evaluation studies produced. I do not believe that either.

The federal government has undergone a series of program review task forces under Deputy Prime Minister Eric Nielsen. The Nielsen teams said that evaluation was often the only information they could get on program effectiveness, but that it was rarely adequate and never slanted toward economizing. No, the explanation is most likely that evaluators *and their clients* played it safe — for any number of understandable reasons — and produced studies that were constructive but inoffensive. We got some value — but not full value from these efforts.

Another evident problem for us is the continuing weakness of the accountability regime in the Canadian federal government. Some promising steps are being taken now. However, it will do no good to put in place the institutional arrangements for an accounting by public servants to the politicians, or by politicians to the public for the results of government programs if they have nothing to say. Evaluation must anticipate these developments and be ready with the evidence on program performance.

On balance, I do not feel that the batting average for evaluation has been unacceptably low. As the above examples have shown, evaluation is being used both operationally and strategically. Evaluation is affecting expenditures far in excess of the few tens of millions of dollars that the federal government spends on the function.

Still, programs remain where evaluation is feeble or non-existent. And evaluation is not accepted everywhere as a routine part of public sector management — as normal and necessary as planning or audit. Even though evaluation may bring good news and enlarge opportunities, it can be seen as a threat. I am proud that evaluation has made the progress and contributions that it has so far, despite this perception. But I am not satisfied. Evaluation can and will be used in many ways and at many levels to improve programs, influence resource allocation, guide policy development and enhance accountability.