Evaluation and Program Planning Vol. 10, pp. 119-123, 1987
Printed in the USA.

0149-7189/87 $3.00 + .00

USING MULTIPLE LINES OF EVIDENCE AND INDEPENDENT TEAMS TO ENHANCE THE CREDIBILITY OF EVALUATIONS

ALAN WINBERG

Evaluation Directorate, Indian and Northern Affairs, Canada

ABSTRACT

The utilization of evaluation findings is an ongoing issue among evaluators. This article presents an approach to conducting evaluations that can be used to enhance the credibility of findings and to facilitate the production of adequate documentation in support of evaluation findings. The concepts of multiple lines of evidence and the use of independent teams are discussed. Use of this approach is expected to improve the utilization of evaluation findings.

Two program evaluations that were undertaken at Canada's federal department of Consumer and Corporate Affairs are presented. To undertake the data collection and analysis, these studies were broken into separate "evaluation modules," each one undertaken by an independent team. The teams reported to an in-house study director, responsible for weighing the evidence and findings of each of the teams and writing the final evaluation report.

INTRODUCTION

Canada's federal Department of Consumer and Corporate Affairs has completed internal program evaluation studies of several regulatory programs using the approach outlined in this article, including:

- Evaluation of Canada's Metric Conversion Program (Canada, Department of Consumer and Corporate Affairs, 1983); and
- Evaluation of Canada's "Energuide" Appliance Labelling Program (Canada, Department of Consumer and Corporate Affairs, 1985).

The approach used for the conduct of these studies involved using multiple lines of evidence and independent study teams. Each team assessed specific evaluation issues and presented a separate report. Each of these "pieces" of the evaluation used a different analytic technique and was called an evaluation "module." In these cases, this approach had a number of positive impacts in terms of facilitating the understanding of the analyses by users (and the appreciation of users' needs by evaluators) and thereby, improving the utilization of evaluation findings. The disadvantages of this approach, in terms of increased project management demands, are also considered.

Three key issues are addressed. First, this paper suggests that multiple lines of evidence are often more compelling than only one line of evidence and presents some potentially effective ways to design and manage the collection, analysis, and presentation of these multiple lines. Second, this paper discusses the pros and cons of parcelling out specific tasks within an evaluation to different evaluation teams. This is seen to be an important topic, very relevant to emerging government practices. In both the United States and Canada, current trends in this area indicate that more and more functions which were previously performed by in-house staff are going to be undertaken by private consulting firms, working under contract. As this trend continues, one would expect that many evaluation studies will be managed by an in-house study director, but actually conducted by teams of private evaluation consultants, or by mixed teams of in-house analysts and private consultants. Third, the paper presents a non-traditional

Requests for reprints should be sent to Alan Winberg, Director, Evaluation Directorate, Indian and Northern Affairs, Ottawa, Ontario, K1A, 0H4, Canada

method for interacting with program staff and decision makers through conducting separate discussions on the separate evaluation "modules." This is a potential way to obtain more involvement and response from program staff, both as an evaluation progresses and also at its conclusion.

EVALUATION MODULES

In order to complete the studies over the relatively short timeframe available before decisions had to be taken, internal departmental evaluators were supplemented by external consultants working under contract. For both projects, approximately four-fifths of the time spent on collecting data and about one third of the time used for data analysis was supplied by external consultants. To undertake the data collection and analysis for these evaluations, studies were broken into separate "evaluation modules," each one undertaken by an independent team. These teams were made up of in-house, departmental evaluators and outside consultants, working under contract. The teams reported to an in-house study director, responsible for weighing the various pieces of evidence and the findings of each of the teams, and then writing the final evaluation report.

The findings and recommendations produced in these studies had a very high level of acceptance by management and staff of the programs examined and were implemented soon after completion of the studies. In some cases, actions were initiated in response to evaluation data, as presented by the team undertaking a specific evaluation module, even before the final evaluation report was prepared.

There are a number of reasons why the use of multiple lines of evidence and independent teams, each undertaking a specific evaluation module, contributed to improving the utilization of evaluation results in these two cases. It is widely accepted that a major factor affecting utilization of an evaluation study is the quality of the lines of communication between the people undertaking the evaluation work and those responsible for decision making and implementation of decisions on completion of a study (Patton, 1978; Seigel & Tuckel, 1985; Weiss, 1972). As will be shown, the independent team approach can serve to improve the quality of this communication. The numerous meetings held in studies using this approach also facilitate evaluators acquiring a better understanding of the decision makers and their key concerns (Hill, 1980).

METRIC CONVERSION PROGRAM

Canada's metric conversion program was a program to facilitate the use of metric measures in major economic sectors to the point that there was sufficient momentum to ensure the widespread adoption in Canada of the metric system of measurement. For the study of Canada's metric conversion program, eight study teams, working independently on the various evaluation modules, reported to the study director.

The evaluation modules for the metric evaluation included:

- a review of program files and documents
- an examination of the experiences of other countries that had converted to the metric system
- a survey of program staff
- a survey of industry
- case studies of various economic sectors where conversion had occurred and others where little or no conversion had occurred (comparisons between the various sectors were drawn in order to assess environmental factors affecting the differing levels of conversion)
- a marketing/communications module to assess the communications, public relations and media relations activities in support of the metric conversion program
- an analysis of legislation which could inhibit or prohibit conversion
- an organizational analysis to assess the feasibility of implementing the various options and to prepare an implementation plan for the recommended option

Based on the results of the consultations undertaken in the various study modules, the factors which affect conversion to the metric system in various economic sectors and the experiences of other countries which had converted to metric, the study found that there existed significant momentum for conversion and that the foundation had been laid for successful completion of the conversion to metric in Canada. Conversion was found to be a gradual process, expected to require 30 to 50 years to complete. The study forecast that the Metric Commission's workload would decline in a gradual and steady manner over a twenty-month period to a much reduced level. The study further found that federal government activities which were still required to support the conversion program after this twenty-month period could be undertaken by other federal departments and agencies.

The recommendations of the study included setting new priority areas for the program, and closing down the government co-ordinating body—Canada's Metric Commission—after a two-year period. A work schedule and a list of tasks to be completed over the two-year period by the Commission were also recommended.

These recommendations have now been implemented by the government of Canada.

The multiple lines of evidence and multiple-team approach used in this study made a significant contribution to its utilization by management. The quality of the reports prepared by each of the independent teams, the intensity of the work, and the fact that much of the evidence pointed in similar directions (although based on different data sources by different analysts) were crucial factors. Largely as a result of this approach, the evidence was well understood and considered, by the federal government of Canada, to be sufficiently valid and reliable for decision-making related to the program.

ENERGUIDE

The evaluation of the Energuide appliance labelling program used a similar approach. The Energuide program, similar in many respects to the U.S. Energyguide program, was designed to assist consumers in their decision-making and to accelerate the improvements in energy-savings features incorporated in six types of household appliances offered for sale in Canada.

The program made mandatory the labelling of these appliances according to their consumption of electricity (measured in kilowatt-hours). There was also an Energuide directory which was a compilation of the energy consumption ratings of appliances. The label and the directory were meant to be used together in the appliance purchase decision.

The various evaluation modules used to conduct the evaluation of Energuide included:

- a file review
- a literature review
- a survey of manufacturers
- surveys and other analyses of consumer perceptions
- a survey of retailers
- expert opinion
- modelling, simulation, and benefit/cost analysis to examine program design and logic and the key variables that influence the size of potential benefits from the program

The evaluation of Energuide found that the program may have had some impact on the energy efficiency of household appliances in the past. The energy efficiency of appliances had improved since 1977 when the program was introduced and there was some evaluation evidence to suggest that Energuide made some contribution to accelerating the pace of these improvements. However, retention of the program in its then current form was not seen as providing good value for money. The evidence indicated that all practical energy-savings improvements had been made and it was unlikely that the Energuide program would produce any significant gains in the energy-efficiency of appliances.

As for providing purchase information, the survey module indicated that few consumers used the program and that none of the few consumers who did use the program attempted to calculate life-cycle costs. In addition, retailers in general did not make use of the Energuide information in the promotion or sale of appliances. Ways to improve the marketing aspects of the program were considered in the study.

The evaluation identified and assessed a number of ways of overcoming these weaknesses in the program. Six options for the future were assessed in the study. These options ranged from a "status quo" option to options which included reducing the scope of the program as well as an option to convert the program to a voluntary one. The option to terminate the program and revoke the regulations was recommended. Suggestions were made that further analysis to consider a voluntary program or a program to monitor appliance efficiency in the absence of the mandatory Energuide program be undertaken by Canada's federal department of Energy, Mines, and Resources.

Upon completion of the report, as was the case with the Metric study, the findings were reviewed by the government and actions were initiated to modify the program in line with the evaluation findings. All federal funding of the program was terminated. In addition, policy development work was initiated by Canada's federal department of Energy, Mines and Resources to assess further the voluntary program option that was identified in the evaluation, and that could operate in the absence of Energuide. It is worth noting that this work was contracted to one of the consultants who had worked on one of the evaluation modules in order to build upon the knowledge that had been gained during the course of the evaluation.

MANAGING THE STUDIES

From the point of view of completing evaluations in a timely manner, the use of several independent teams has a number of advantages. This approach:

- permitted a number of pieces of work to be undertaken simultaneously;
- allowed the selection of each team to be totally tai-

lored to the type of issues being focused upon (i.e., communications experts looked at communications issues and produced a specialized report; engineers looked at technical, engineering issues; economists undertook modelling and policy analysis);
- permitted cross-assessment of the results of each team by the other "competing" teams to point out strengths, weaknesses, and conflicting evidence, thus providing valuable input to the study director;
- reduced the dependence of the study director compared with the situation of only one study team to complete all data collection and analysis; and
- provided a ready-made pool of knowledgeable outside consultants to whom follow-up tasks could be contracted more efficiently.

These benefits are very similar to those reported by the National Institute of Education (NIE) team which undertook a major evaluation of Title I compensatory education programs in 1976 and 1977 (Hill, 1980, p. 67).

But there are also potential drawbacks to using independent teams. For example, extra costs may be incurred due to duplication of read-in time. A larger number of analysts must acquire an understanding of the basic details of the program, the specific issues and the evaluation plan (however, these extra costs in the cases examined were relatively small compared with the total amount of work undertaken in each module).

A second type of disadvantage relates to the fact that while this approach simplified management of each of the smaller modules, management of the overall study became somewhat complex and required careful coordination and direction of the various teams by the study director. Upon completion, each of the various pieces of the study had to fit together. When external consultants were involved, much more effort was devoted to writing and coordinating multiple requests for proposals (RFP's) as opposed to a single, large RFP. In addition, the study director had to fully understand the details of each module and the problems being encountered by each of the teams undertaking the various modules. Using this approach, therefore, required that the study director be both an expert analyst and an efficient project manager at the same time.

It is important to caution that the multiple team approach may not be appropriate in certain cases. For example, a single, large study can probably address certain types of questions which are more complex (such as those that require complex sampling or survey methods) than several relatively small studies. In addition, in some instances, the multiple team approach may lead to difficulties in interpreting the data due to non-overlapping samples and other difficulties in research methodologies. In such cases, additional modules may have to be undertaken to resolve conflicts in the evidence and this could result in increased costs for completing the total study.

Other potential drawbacks to using this approach relate to the limited scope of each of the smaller evaluation modules. For example, if not carefully managed, there is the danger of limiting analysts to only their own field of expertise and thus missing possible cross-fertilization of having separate focused discussions on separate issues and thus possibly overlooking some relationships between issues. Again, these potential drawbacks emphasize the vital role of the study director in ensuring that they are kept to a minimum in the evaluation.

The study director is a key person both during and after the evaluation. An important element in utilization of the evaluation is the understanding of the program, its stakeholders and its users acquired during the study by the in-house study director. This knowledge remains within the organization long after the formal evaluation study has been completed. The existence of a well-briefed, respected, permanent in-house evaluator, who understands the needs of the client and who will stand behind the findings, is thought to be a major factor affecting utilization. Moreover, it may take some time for evaluation findings to work their way through an organization to the implementation state. Here, the continued expertise and understanding that is maintained in-house in the program evaluation unit has been a major factor in utilizing the evaluation findings.

MULTIPLE LINES OF EVIDENCE

Both of these evaluation studies were carried out over a timeframe of about one year, and involved data collection over one relatively compressed time frame. Both studies were based on a relatively limited number of interview modules, focus groups, expert opinion, surveys of various impacted groups, legal opinion and analysis, cost-benefit analysis, economic modelling and simulation, literature reviews, and program file reviews.

Each of these modules had its strengths and weaknesses as well as its specific focus. In both cases, it was not expected that any single data source or analytic method could provide a complete perspective on the program. The relatively soft nature of the data base, the short time frame for the studies and the reliance to a large extent on individuals' perceptions as a major source of information suggested a design incorporating multiple lines of evidence. No one module was expected to resolve all the evaluation issues with an acceptable level of certainty. In these evaluations, the different evaluation modules differed in methods of data collec-

tion and analysis, and in some assumptions. However, the combination of evaluation modules, taken together, allowed for validation and cross-checking of findings. Advisory committees which were created for these studies provided an additional mechanism to review the work of each module. These committees were made up of program management and other interested parties.

In the cases here presented, the use of multiple lines of evidence, combined with the independent team approach proved to be a very powerful persuader. It is possible, of course, that simply using multiple lines, even without the independent teams, would still have been successful in convincing skeptical audiences of the validity of the findings. However, there is a strong intuitive feeling by all those associated with the two cases here presented that the use of independent teams did add significantly to the level of credibility and utilization of the study findings.

CONCLUSIONS

The ultimate purpose of most evaluation work that goes on is to enable better decisions and ensure that what is decided is practical and can be implemented as desired. However, it is well known that some evaluation studies, while they have been carried out in a meticulous fashion, end up gathering dust on someone's bookshelf. Sometimes the problem is that the analysis was not well understood by potential users or well communicated to them. Evaluators may be largely to blame in such circumstances. Equally important, sometimes evaluators may not have an adequate understanding of the decision makers and the decision process (Datta & Perloff, 1979; Hill, 1980; Patton, 1978).

Using multiple lines of evidence and independent teams can help in many cases. As the subject matter of each team is significantly more straightforward than for the study as a whole, and often addresses fewer evaluation issues, this approach permits focused discussions with the program specialists. In the cases where it has been used, compared with other cases where one team has carried out all aspects of an evaluation, there has been a higher level of participation by program managers and staff. The extensive knowledge and experience of program staff thus can feed into the evaluation work. In our experience, program staff have tended to be more interested dealing with the selection of specific approaches for specific issues, as well as with the review of draft results and the discussion of the strengths and weaknesses of one specific study module. Work is going on at a finer level of detail. Furthermore, a given module can have significant problems without detracting from the overall validity of the full study (as compensations can be made to overcome these weaknesses in other modules).

A multiple team approach permits more frequent meetings to review tangible inputs to the evaluation, that is, the reports being presented by the study teams on each of the evaluation modules. Simply put, there are more special occasions to meet and discuss what is taking place, what has been found, and what is to occur in the next steps. This helps understanding of the evaluation by users; it also reduces the likelihood that errors of interpretation or fact by evaluators will go unnoticed for a long period of time. Each reporting occasion also becomes an occasion for mid-stream correction to ensure the work does not go off track. These occasions can also serve to improve evaluators' understanding of the decision makers and their key concerns.

To summarize, this approach produces few if any surprises. The findings are presented in a cumulative manner, long before the writing of the final report. For the cases here presented, once the final report was being prepared, the major findings and how they were arrived at were well known to decision makers. Without question, in the two cases here cited, the use of multiple lines of evidence and independent teams, reporting to an in-house study director, responsible for bringing together the various pieces of analysis, drawing conclusions and making recommendations, contributed significantly to producing evaluations which were immediately acted upon.

REFERENCES

CANADA. DEPARTMENT OF CONSUMER AND CORPORATE AFFAIRS (1983). *Metric evaluation*. Ottawa: Department of Consumer and Corporate Affairs.

CANADA. DEPARTMENT OF CONSUMER AND CORPORATE AFFAIRS (1985). *Energuide evaluation*. Ottawa: Department of Consumer and Corporate Affairs.

DATTA, L., & PERLOFF, R. (1979). *Improving evaluations*. Beverly Hills, CA: Sage Publications.

HILL, P.T. (1980). Evaluating education programs for federal policymakers: Lessons from the NIE compensatory education study. In John Pincus (Ed.), *Education evaluation in the public policy setting*. Santa Monica, CA: Rand Corporation.

PATTON, M.Q. (1978). *Utilization-focused evaluation*. Beverly Hills, CA: Sage.

SEIGEL, K., & TUCKEL, P. (1985). The utilization of evaluation research. *Evaluation Review, 9*, 307-328.

WEISS, C. (1972). *Evaluation research: Methods of assessing program effectiveness*. Englewood Cliffs, NJ: Prentice-Hall.