

PEER ASSESSMENT OF EVALUATION PRODUCTS: PRACTICE IN THE DEPARTMENT OF NATIONAL DEFENCE

Ira Lewis
Director General Program Evaluation
Department of National Defence

J. Edward Browell*
Director General Program Evaluation
Department of National Defence

Abstract: Although evaluation reports produced in the Department of National Defence have been subject to peer assessment since 1972, limited evidence to support this practice was found in the evaluation literature. Peer assessments are used to ensure that evaluation standards have been met, to correct gaps in reasoning and errors, and to solicit insights from reviewers.

Résumé: Quoique les rapports d'évaluation produits au sein du ministère de la Défense nationale ont été assujettis à un processus de révision par les pairs depuis 1972, il n'existe que peu d'exemples publiés de cette pratique dans la littérature portant sur l'évaluation. La révision par les pairs est employée afin de s'assurer que les normes d'évaluation sont respectées, corriger les lacunes dans la logique et les erreurs matérielles et obtenir des commentaires auprès des réviseurs.

Carrying out effective evaluations requires a mix of formal controls and informal management practices. Since 1972, evaluation studies in the Department of National Defence (DND) have been subject to peer assessment. This practice is not required by government policy, nor does extensive support for the practice exist in the evaluation literature. However, peer assessment has proven a valuable method of ensuring the high quality of evaluation studies,

* Mr. Browell retired from the public service of Canada in December 1995. Any opinions expressed in this article are those of the authors and not of the Department of National Defence or the Government of Canada.

and may be as important as the work of the evaluators who actually conduct evaluation studies.

The Director General Program Evaluation (DGPE), part of the Review Services Branch at National Defence Headquarters, has a staff of 12 evaluators, which is equally divided between public servants and military officers. DGPE is larger and contracts out fewer evaluations than most federal government evaluation organizations. Recent evaluation work has included reviews of a number of defence policy, operational, and support areas.

EVALUATION QUALITY ASSURANCE

The auditor general of Canada (1993) has pointed out that program evaluation in the Canadian government is faced with an “intractable issue,” described as “one that is not easily dealt with because it is complex and challenges the status quo” (p. 17). Evaluation groups within departments and agencies need to “ask the tough questions and assess significant expenditures,” but the auditor general found that program evaluations often provide incomplete coverage of the tough evaluation questions. Also, evaluation groups were found to seldom deal with all of the main evaluation issues: cost-effectiveness, program success, and program relevance.

This uneven quality of program evaluations reduces their credibility and usefulness. However, the auditor general also found some evaluations that were well done and that provided important information on program results; those findings “provide a demonstrated potential for what the evaluation function can achieve when it produces work of quality” (p. 231). The auditor general recommended that program evaluation managers “ensure that they have the required management processes in place to build quality into their evaluation products” (p. 256).

The evaluation policy for federal departments and agencies set out by the Treasury Board of Canada (1994) establishes specific requirements for evaluation, and is intended to strengthen evaluation practices. To achieve that objective, the skills and competencies of the evaluation staff are to be developed and enhanced to ensure that they have the ability consistently to conduct and manage highly professional, state-of-the-art evaluation work.

The Treasury Board of Canada monitors the implementation of the evaluation policy through performance measures and standards,

which provide departments and agencies with a workable basis for improving the quality of evaluation practice. The evaluation policy also provides quality assurance criteria for peer assessment.

EVALUATION PEER ASSESSMENT

Guba and Lincoln (1989, p. 237) have described the “peer debriefing” process (a synonym for “peer review”) as:

the process of engaging, with a disinterested peer, in extended and extensive discussions of one’s findings, conclusions, tentative analyses, and, occasionally, field stresses, the purpose of which is both “testing out” the findings with someone who has no contractual interest in the situation and also helping to make propositional that tacit and implicit information that the evaluator might possess. The disinterested peer poses searching questions in order to help the evaluator understand his or her own posture and values and their role in the inquiry; to facilitate testing working hypotheses outside the context; to provide an opportunity to search out and try next methodological steps in an emergent design; and as a means of reducing the psychological stress that normally comes with fieldwork—a means of catharsis within a confidential, professional relationship.

Corbeil and McQueen (1991) suggest one way of improving the quality of an evaluation is through peer reviews conducted by colleagues, steering or advisory committees, or stakeholders. The authors see peer review as a means of ensuring that there are no oversights or biases in the evaluation. An earlier paper by Cook and Gruder (1978) dealing with evaluation of evaluation, also known as “metaevaluation,” suggested that metaevaluation consultants used peer review as an evaluation-monitoring technique.

The U.S. General Accounting Office (GAO) (1993) requires evaluation teams to submit reports to a “referencer,” who helps ensure that the product will be of the highest possible quality, will accurately and objectively communicate results, and complies with GAO reporting policies and standards. Referencing (i.e., peer reviewing) is one of the GAO’s principal quality control processes.

The referencer is an independent evaluator not involved in the same evaluation study but who has at least three years’ experience.

Referencers are selected on the basis of independence, objectivity, experience, analytical ability, and knowledge of the rules of evidence and of GAO's reporting policies and standards. The referencer checks each statement or fact for consistency within the report and with supporting work papers, independently checks all figures, and ensures that the findings, conclusions, and recommendations are adequately supported by credible evidence. Finally, the referencer looks for evidence in working papers that contradicts or calls into question facts or statements in the report.

Peer assessment should not be confused with quality assurance reviews of evaluation reports prepared by contractors. The peer reviewer is not a member of the evaluation team, but he or she is employed by the same evaluation organization, and so the arm's-length relationship required in a client/contractor relationship is neither present nor required.

Carrying out peer assessments to ensure that evaluation studies are of a high, professional quality is in part appropriate because evaluation work is similar to academic or scientific research in that its major inputs are human intellect, initiative, and creativity.

PEER ASSESSMENT IN DND

Peer assessment, an integral part of the DND program evaluation integrated process and methodology, has been described by Clark (1994). Two or three members of the Program Evaluation Division are chosen as peer reviewers, usually at the outset of the evaluation study. Occasionally, when technical expertise in subject matter is required and not available within the Program Evaluation Division, evaluators may request an individual from outside the evaluation organization, but within DND, to carry out a peer assessment on specified technical material. Such individuals, however, cannot come from the ranks of those that will be responding to the final evaluation report. The process, though required by internal instructions, is an informal one, and the authors of the evaluation study are not obliged to adopt changes suggested by the reviewers.

Sometimes it is useful to provide peer reviewers with parts of the draft report in stages as they are prepared. Comments from peer reviewers at each stage can help with the drafting of later parts.

Reviewers are often asked to sit in on rehearsals of oral briefings, which play an important role in the presentation of evaluation find-

ings. The peer reviewers contribute their own background, experience, and knowledge of DND as well as the knowledge they have acquired in reviewing the draft report, and can advise on briefing presentation methodology in addition to playing the role of the program managers who will be receiving the presentation and final report. As pointed out by Caron (1993), a sound knowledge of the organization's culture is critical to the success of both the evaluation report and accompanying oral presentations.

Figure 1 provides the peer assessment criteria checklist used within the Program Evaluation Division. Another source of peer assessment criteria is provided by the U.S. General Accounting Office (1993). These criteria are used for the general guidance of both reviewers and the authors of the evaluation study, and are not intended to be explicitly answered in every case.

CONCLUSION

Peer assessment provides a useful means of assessing compliance with the evaluation policy set out by the Treasury Board of Canada, confirming that Program Evaluation Division evaluation methodology has been properly followed, and identifying factual errors and gaps in logic that can be corrected before a report is sent to stakeholders for comment. Information that may create legal, security, or public-access-to-information concerns can be identified early on. Additionally, this technique provides a fresh look at a report by someone familiar with the evaluation process but who has not been involved in the preparation of the report's findings. Reviewers contribute their own background and experience and learn more about evaluation methodologies and the subject of the evaluation, which may be of assistance in later evaluation studies. Accordingly, participating in a peer assessment as a reviewer provides a useful form of professional development for evaluation staff.

The relatively large size (twelve evaluators) of DND's Program Evaluation Division makes it feasible to involve as peer reviewers two or three evaluators who have not been members of a specific evaluation team. Smaller evaluation groups, particularly those that rely extensively on contractors, may not have the same option. Those organizations might consider asking qualified colleagues from outside the evaluation group to act as peer reviewers, which would have the added benefits of providing a professional development experience for the reviewer and making the role of the evaluation group

better known in the organization as a whole. Peer assessments offer an important quality enhancement means of ensuring that excellence is achieved in the evaluation process.

REFERENCES

- Auditor General of Canada. (1993). *Report to the House of Commons* (chaps. 8–10). Ottawa: Supply & Services.
- Caron, D.J. (1993). Knowledge required to perform the duties of an evaluator. *Canadian Journal of Program Evaluation*, 8(1), 59–78.
- Clark, A.A. (1994, May). Effective program evaluations in times of resource restraints: A five-part integrated process for conducting timely and relevant program evaluations. Paper presented at the 14th annual conference of the Canadian Evaluation Society, Quebec City.
- Cook, T.D., & Gruder, C.L. (1978). Metaevaluation research. *Evaluation Quarterly*, 2, 5–51.
- Corbeil, R.C., & McQueen, C. (1991). Improving the quality of evaluations. In A. Love (Ed.), *Evaluation methods sourcebook* (pp. 204–206). Toronto: Canadian Evaluation Society.
- Guba, E.G., & Lincoln, Y.S. (1989). *Fourth generation evaluation*. Beverly Hills, CA: Sage.
- Treasury Board of Canada. (1994). *Treasury Board manual: Review, internal audit and evaluation*. Ottawa: Canada Communication Group—Publishing.
- U.S. General Accounting Office. (1993). *Communications manual* (chaps. 12 and 13, pp. 1–8, App. 1). Washington, DC: Author.

Figure 1 **Peer Review Criteria Checklist**

- 1. Have all the evaluation issues and questions in the Evaluation Directive been answered?**
- 2. Have generally accepted evaluation methodologies been properly followed?**
- 3. Is there evidence that the broader environment in which the program exists has been taken into account?**
- 4. Does the Program Profile provide a comprehensive statement on the program, including its mandate, objectives, description, resources, performance indicators, activities, outputs, expected impacts and effects and the program structure?**

Figure 1 (continued)

-
5. Has the evaluation produced timely, pertinent and credible findings and conclusions that managers and other stakeholders can use with confidence, based on practical, cost-effective and objective data collection and analysis?
 6. Is the report written in a style that communicates with its audience? Is the order of the material logical? Has the standard outline for evaluation reports been used?
 7. Has data gathering and analysis in investigations specific to the evaluation and from supporting ongoing measurement produced documented evidence on performance? Is the evidence sufficient in relation to the decision-making context? Are evaluation findings relevant to the issues addressed and do they follow from the evidence?
 8. Has an independent technical expert examined and approved the manner in which highly technical data have been developed and used?
 9. Is every calculation of a key total, percentage, or similar figure accurate?
 10. Have the evaluation information and data been collected, analyzed and presented in an objective manner such that if others conducted the evaluation they would reach similar findings?
 11. Are the conclusions on the evaluation issues consistent with and clearly distinguished from the evaluation findings?
 12. Do the conclusions flow directly from each of the evaluation issues specified in the directive, the information and data collected, and the analysis? Are they convincing? Is there any evidence in the report that either contradicts or calls into question facts or statements in the report?
 13. Are the conclusions the answers to the issues questions posed in the evaluation directive?
 14. Have any conditional conclusions been made based on incomplete evidence on a program, i.e. conclusions that are absolute in nature but are conditional upon certain senior management assumptions, points of view, predictions or conditions? Are these conditional conclusions valid?
 15. Are findings and conclusions presented in a clear and balanced manner along with an indication of their degree of reliability?
 16. If the recommendations were adopted, would significant change result? In other words, if the recommendations had been in place, would the deficiencies found have been avoided?
 17. Are the recommendations likely to be adopted and implemented?
 18. Do the findings and recommendations reflect an accurate perception (or understanding) of the culture of the organization or activity being evaluated?
 19. Does the report comply with government policies related to information collection, use, preservation and dissemination?
 20. Does the text appear to breach any personal confidence, i.e. is it possible to deduce who said something?
 21. Does the report contain appropriate technical annexes including a bibliography, glossary and adequate reference material?
 22. Does the evaluation report meet the Standards for Individual Evaluations reported in the Treasury Board Standards for Evaluation in Federal Departments and Agencies?
-

