

ENSURING QUALITY FOR EVALUATION: LESSONS FROM AUDITORS

John Mayne
Ottawa, Ontario

Abstract: This article addresses ways to enhance the quality of evaluations with weak designs through a variety of quality assurance practices. Many types of evaluations are restricted in the types of designs they can use. Evaluations of development programs with widely dispersed projects in different countries are often a case in point, where the design uses visits to a number of dispersed sites, interviews with staff and stakeholders, and reviews of documentation to draw conclusions. These interview-based evaluations are quite similar in methodological approach to many performance audits. National audit offices devote considerable resources to their quality assurance practices, and, for the most part, the quality of their performance audits is not questioned. It is argued that evaluations, and not only interview-based ones, could usefully adopt many of the quality assurance practices used by national audit offices to ensure the quality of their products.

Résumé: Cet article traite des façons de rehausser la qualité des évaluations mal conçues à l'aide des pratiques d'assurance qualité. Plusieurs types d'évaluations sont limitées quant aux types de conceptions qu'elles peuvent utiliser. Les évaluations de programmes de développement portant sur des projets dispersés dans plusieurs pays sont souvent des exemples de conceptions basées sur des visites à un certain nombre de sites, des entrevues avec le personnel et les intéressés, et des examens de la documentation pour tirer des conclusions. Ces évaluations basées sur des entrevues sont assez semblables aux pratiques de vérification du rendement. Les bureaux de vérification nationaux consacrent énormément de ressources à l'assurance de la qualité et, la plupart du temps, la qualité de leurs vérifications du rendement n'est pas remise en question. L'auteur de l'article soutient que les évaluations — pas seulement celles basées sur des entrevues — pourraient peut-être tirer profit de l'adoption de plusieurs des pratiques d'assurance de la qualité utilisées par les bureaux de vérification nationaux pour assurer la qualité de leurs produits.

Corresponding author: John Mayne, 654 Sherbourne Rd., Ottawa, ON, K2A 3H3;
<john.mayne@rogers.com>

What constitutes good quality in evaluation is fairly well established. In a number of jurisdictions, evaluation standards for good quality have been developed and widely distributed within the profession. In the US, evaluation standards were developed by the Joint Committee on Standards for Educational Evaluation in 1981 and a revised version issued in 1994 (Joint Committee, 1994) with less focus on education. These standards set out what constitutes good evaluation and good evaluation practice for members. These standards have served as the basis for evaluation standards developed in other countries, such as Switzerland (Widmer & Beywl, 2000) and some countries in Africa (Nairobi M&E Network et al., 2002). Widmer (2005) argues these standards have had a positive effect on the quality of evaluations in Switzerland.

Schwartz and Mayne (2005) have identified (see Table 1) three types of evaluation standards, those covering evaluation products, those for the evaluation process, and those focused on the usefulness of the evaluation.

This article does not discuss criteria for good quality per se, but rather asks how good quality can be arrived at. In particular, what

Table 1
Standards for Good Quality in Evaluation

Product quality criteria

- *Well-defined scope.* The objectives of the evaluation, the purposes to be served, and the range of coverage should be clearly set out.
- *Validated criteria.* The evaluative criteria to be used should be validated (drawn from some authoritative source or through agreement).
- *Accurate data.* The data collected should be valid and reliable.
- *Sound analysis.* The analysis of the data collected should be based on robust methodology.
- *Substantiated and impartial/objective findings/conclusions.* The findings and conclusions presented should be supported by the evidence gathered (data and analysis) and should be presented in an impartial (objective) manner

Process quality criteria

- *Efficiency* in production
- *Fair and objective* data gathering and production
- *Involvement of stakeholders* in the planning and conduct
- *Involvement of qualified investigators*
- *Disclosure to affected parties* of pertinent findings

Usefulness criteria

- *Relevance* to key decisions
- *Timeliness*
- *Transparency* in analysis
- *Clarity* of messages

Source: Schwartz and Mayne, 2005 (pp. 304–305).

quality assurance practices are there that might help ensure good quality, especially when, perhaps for good reason, the design of an evaluation is weak?

INTERVIEW-BASED EVALUATIONS

Evaluation designs are many and varied, each with their strengths and weaknesses. Evaluators are encouraged to adopt as robust a design as possible in the circumstances. Many evaluations are able to use quite sophisticated and ingenious designs. Others, however, often for defensible reasons, use less sophisticated and simple designs. In particular, we see evaluations that entail:

- visiting several sites where the program is operating,
- interviewing program staff, recipients, and stakeholders,
- reviewing existing program documentation,
- perhaps conducting a survey, and
- based on all this information, writing an evaluation report.

Such interview-based evaluations are not necessarily the result of inadequate development of a more robust methodology. Often they are all that can practically be done for a reasonable investment of evaluation resources. Evaluation of many development aid programs is a case in point. Evaluation here faces a number of serious methodological challenges:

- The program is implemented through a large number of projects.
- These projects are usually geographically dispersed in a number of countries around the world.
- Many individual projects may be relatively small.
- The outcome sought very often takes many years to achieve.

The evaluation of such programs is usually undertaken, based on terms of reference (TOR), by a small evaluation team who visit a number of the projects for a short period of time, interview those associated with the program (staff and recipients), and review pertinent administrative documents. Sometimes a survey questionnaire is sent out to a wider audience in order to supplement the information gathered. The team then pulls together what was found and drafts an evaluation report, responding to the TOR as best they can. These evaluations are often summative in nature, but could be formative.

In my experience, and in discussions with others working in evaluation of development programs of donor organizations, this interview-based methodological approach to these evaluations is quite common. One sees similar approaches in other evaluation settings. While all agree that better evaluation designs would enhance the robustness of the evaluation findings, the interview-based evaluation may be all that is doable, given that:

- Widely dispersed projects mean only a few can be examined, since visiting is costly and time-consuming.
- Comparisons with other projects or situations is not usually feasible, or too expensive. It is often quite difficult to get the same data and information across a range of diverse and geographically dispersed projects or programs in a number of countries around the world.
- Examination of a specific project can really only involve interviews, collection of available administrative data, and inventorying project outputs.

While one might be sympathetic to the problems faced in interview-based evaluations, these evaluations clearly have serious weaknesses to their credibility:

- self-interest testimonies from recipients of program and project benefits, and from program staff,
- limited or no counterfactual evidence,
- evaluator bias, since analysis and interpretation is based on the evaluator assessing interviews and documentation,
- spending money (although usually producing some benefit to someone) that, on its own, doesn't tell much about how well the program is doing.

Some of these weaknesses can be reduced through clear upfront evaluation objectives and, in my view, through the use of theory-based evaluation approaches. These approaches are not considered here. This article considers ways of strengthening the interview-based approach to evaluation by building in some robust quality assurance practices borrowed from performance auditing. In so doing, possible useful quality assurance practices are suggested for evaluations generally.

The next section looks at some of the similarities between performance audit and evaluation, noting the similarities in methodological

approach between performance audits and interview-based evaluations. It then briefly looks at several examples of quality assurance practices in evaluation and discusses independence and objectivity in audit and evaluation. The following section discusses how the quality assurance practices in performance audit could be adopted for use in interview-based evaluations, and more generally. I conclude with some overview comments on quality assurance for evaluation.

QUALITY ASSURANCE IN PERFORMANCE AUDIT AND EVALUATION

Performance Auditing

Auditing compares what is with what ought to be, as determined by some form of criteria. In financial auditing, the criteria (generally accepted accounting principles) have become widely accepted. Performance audit concerns itself with some aspect of performance of or within an organization, either some element of the actual performance of programs, or the systems and procedures used to manage programs. For performance audits, the criteria are determined, set out, and usually agreed upon with management at the outset of the audit. The performance audit assesses the extent to which the criteria have been met.

Performance audit and evaluation share many characteristics, as numerous authors have observed (Leeuw, 1992; Wisler, 1996). Both are trying to examine, measure, and report on some aspect of a public sector policy, program, or project and rely on trained professionals to undertake the work. Both are evidence-based, seeking to bring together the best evidence available in the circumstances at hand. Performance auditing as carried out by national audit offices — the Supreme Audit Institutes (SAIs) — have well-established quality standards of good audit practice, both at the international level (International Organization of Supreme Audit Institutions, 1992) and within each SAI. Of interest in this article, however, are their approaches to quality assurance: what practices are adopted to ensure that good quality audits are produced.

Two observations can be made at this point. First, the methodological approach of many if not most performance audits is that followed in what have been called interview-based evaluations: interviews, administrative file review, and surveys. Second, for the most part, performance audits by SAIs are seen by most as quite reliable and credible; they are seen as of good quality (Lonsdale &

Mayne, 2005). There is occasional criticism of the role of and methodologies used by SAIs (Bowerman, 1996; Power, 1997; Roberts & Pollitt, 1994; Schwartz, 1999; Sutherland, 2001), but there are few instances of performance audit findings being seriously challenged. Performance audits appear to be able to handle the inherent weaknesses in their methodological approaches fairly well. We argue that this is in large part due to their quality assurance practices.

Schwartz and Mayne (2005) conclude that there are significant differences between the two disciplines of performance audit and evaluation when it comes to quality assurance. In particular, it is pointed out that performance audits as carried out by national audit offices (SAIs) have in place quite formidable quality assurance practices. SAIs put considerable value on their credibility and want to ensure

Table 2
Quality Assurance Practices of SAIs

Quality assurance during audits

- Use of standard phases for each audit (typically a survey/preliminary study phase, an examination phase, and a reporting phase)
- Clear quality standards, especially requirements for
 - independence
 - objectives for the audit, criteria against which assessments will be made, and the need to conclude against the objectives/criteria
 - adequate substantiated evidence on which findings and conclusions are based
- Use of an oversight/advisory committee who advise at key stages in the audit, who often include, in addition to internal specialists and senior management, expert subject-matter external advisors, stakeholders, and sometimes the auditee
- Use of a quality reviewer who signs off at various stages of the audit, ensuring amongst other things that there is adequate substantiation for the findings in the report
- Draft audit reports sent to the auditees for review and agreement on the facts uncovered in the audit
- Standardization of audit report structure and availability of editorial assistance for final drafting of report

Quality assurance after audits are completed

- Post-audit quality practice reviews (meta-evaluation) undertaken on a sample of completed audits by an independent unit with that responsibility, to determine if the audit met office quality standards
- Quality practice reviews leading to lessons learned
- Self-evaluation by the audit team, leading to lessons learned being reported and discussed in the office
- In a few cases, review of completed audits by an outside independent academic body
- Surveys of auditees to obtain their views on the extent to which the audit followed proper processes

Note: Not included in the above are the extensive audit guidance and training provided on the audit standards to be followed.

that reports they produce are as accurate, timely, and relevant as possible in their findings, conclusions, and recommendations. Table 2 — developed from Lonsdale and Mayne (2005) — lists the kinds of quality assurance practices in place in SAIs.

Current Quality Assurance in Evaluation

The article does not review or summarize current quality assurance practices in evaluation. We can observe that the literature is not replete with discussion on the topic (see Corbeil & McQueen, 1991, for one discussion). Datta (2000) discusses strategies to use to achieve a non-partisan evaluation in a partisan world. Interestingly, a number of the strategies discussed make reference to practice in the General Accounting Office, the US SAI.

From an organizational perspective, two examples of quality assurance practices in evaluation can be pointed to: the MEANS approach used by the Directorates-General in the European Union (EU) and the independent evaluations carried out by the World Bank. Both cases focus on summative evaluations carried out to assess the effectiveness of the programs and projects examined.

Toulemonde, Summa-Pollitt, and Usher (2005) describe the EU's approach, whereby a panel of experts reporting to a steering committee of EU officials formally assess, at both a draft report stage and a final report stage, the quality of evaluations carried out by external evaluators for the EU, using the MEANS criteria grid (European Commission, 1999). This covers such elements as relevance, evaluation design, reliability of data, and robust findings. The evaluators are aware of the process, and work to produce an evaluation report, which has a good assessment (with the final assessment usually published along with the report). The process is structured and quite formal. Toulemonde et al. review the EU experience with this system and conclude that it has helped improve the quality of evaluations, but that the evaluations still fall far short of quality expectations.

Grasso (2005) describes the World Bank's fairly extensive efforts at assuring the quality of their evaluative information, particularly at the project level. During implementation and at completion through Implementation Completion Reports, projects are self-rated by project managers using five evaluation criteria covering development outcomes, sustainability, institutional development impact, and

Bank and borrower performance. To assist in assuring the quality of this information, training in monitoring and evaluation is provided, and a separate Quality Assurance Group using a panel of experts rates a random sample of ongoing projects using the same criteria. In addition, the independent Operations Evaluation Department who report to the Bank's board (not its management), does its own desk review of all Implementation Completion Reports, carries out ratings based on field assessments of 25% of completed projects 2–5 years later, and impact evaluations of a small number of projects 5–10 years after completion. The World Bank quality assurance system is quite elaborate, well structured, and formal. Grasso concludes that this extensive effort in quality assurance has begun to improve the quality of project-level evaluation information in the Bank, but that “monitoring and evaluation remains a weakness in the bank's project management” (Grasso, p. 108).

There are other examples of quality assurance practices for evaluation in other organizations. The ones I have seen are all home-grown, developed to suit the circumstances of the organization in question, adopt a very formal assessment approach, and are often done after the fact, as in the case of the International Development Research Centre (2002). At IDRC, a corporate evaluation group applies quality assessment criteria to all evaluations conducted in the previous 12 months and communicates the findings to their senior management, their board, and the programs that conducted the evaluations. All these approaches differ somewhat from the well-established and proven approaches used by SAIs.

A major difference between the SAI approaches and those described above in the EU and World Bank is that the former are designed to ensure, with some degree of certainty, that *each* audit product meets quality standards. SAI quality assurance procedures are expected to *ensure* good quality, not to simply improve quality. SAIs believe they cannot afford to produce public audit reports that are below quality standards. Even one poor-quality audit, it is believed, would greatly affect the credibility of the SAI, and considerable effort is made to ensure this does not happen.

Independence and Objectivity

SAIs are uniquely positioned institutions whose main *raison d'être* is to produce quality independent external assessments *about others*. Public sector organizations, on the other hand, are not in busi-

ness to produce evaluations per se, but rather to achieve some result in the outside world. Evaluation is a tool to help them do that, through finding out what works well and what does not, making improvements in their operations, and being able to report externally on how well things are going. These different settings mean that external performance audits and evaluations done by organizations do differ. Hence not all of the quality assurance approaches used by SAIs are applicable for evaluation, and some that are applicable need to be adapted to make them relevant to the evaluation setting. But, in our view, there is a lot that could be usefully adopted.

Before discussing specific quality assurance practices, we need to address one possible major difference between evaluation and performance audit as carried out by SAIs: as just mentioned, SAIs do not audit themselves while organizations do indeed evaluate themselves. Thus, SAIs are quite comfortable with negative findings and conclusions about the entity they are auditing.

Organizations carrying out evaluations are less comfortable with such findings. Many organizations, recognizing the drawbacks of self-evaluations, deliberately set up their evaluations so as to enhance the independence and objectivity of the evaluation process as much as is practical. This is done by such things as having evaluations managed by a central group, independent from program managers, and using outside independent evaluators to carry out the evaluation.

This latter case is often called an “external evaluation,” in that it is carried out by an evaluator external to and, to a degree, independent from the organization.

In my view, despite the above-mentioned difference between SAIs and organizations conducting evaluations, the motives and expectations of the two are quite similar. Because performance audits and evaluations are made public, both have a desire to produce a credible product they can stand behind. If an organization does not want (or cannot tolerate) bad news, then of course they should probably not undertake evaluation, as it is unlikely that all is completely well in most organizations.

The more usual case, where an organization does indeed want to know what is working and what is not working (although perhaps not with the same vigour as an SAI!), is addressed here. In this case, attention to quality assurance ought to be essential to avoid wasted evaluation

resources and incorrect findings and conclusions, and to produce an evaluation that is credible to those outside the organization.

A practical difference in the two settings is the frequent use of outside evaluation experts to carry out evaluations, to enhance the independence and objectivity of the evaluation and/or to purchase skills and experience the organization does not have itself. SAIs often use outside consultants extensively in performance audits. However, for the most part the audits are managed by an audit office staff member,¹ and the product is clearly seen to be that of the SAI, not the individual auditor.

When organizations use external evaluators, especially to enhance independence and objectivity, a key question is who “owns” the final evaluation report. If it is the organization, does that call into question whether the organization who is paying the evaluators — and who themselves may hope for future contracts — has unduly influenced the conclusions of the report? Or, if the final product is the evaluator’s, what assurance can the organization have that they have received a quality product? In the interview-based type of evaluations, much depends on the experience and integrity of the evaluators used. It is difficult for an individual evaluator or a small team of individual evaluators to have in place robust quality control procedures that will be seen as credible.

Clearly, a balance is required between these two possibilities — undue influence by the organization versus weaknesses in the conduct and methodology of the evaluation, or bias, intentional or otherwise, by the evaluator. It is just such a balance that, I suggest, can be fostered by good quality assurance practices in evaluation.

But perhaps even more important than getting this balance exactly right is the need for evidence-based findings and conclusions. Assuring a degree of objectivity and independence for the evaluator should not mean placing the evaluator beyond criticism and the need to be able to substantiate findings. The best way to ensure robust findings and conclusions is to subject them to rigorous challenge. This is where quality assurance practices while the evaluation is being conducted come into play.

In the next section, the performance audit quality assurance practices as they could be applied to evaluation are outlined. The value of adopting such practices in the case of interview-based evalua-

tions, such as those often found in development evaluation of donor programs, is most obvious, given the similar methodologies adopted. But the practices discussed below are clearly generic in nature, and could indeed be applied to any evaluation. Where and when to do so is an issue that is discussed as we go along. The context is largely that of summative evaluation. Overall, I would argue that the weaker the evaluation design — and hence the more open to challenge — the more attention should be paid to good quality assurance practices.

QUALITY ASSURANCE PRACTICES FOR EVALUATION

The performance audit quality assurance practices listed in Table 2 can be summarized as a focus on three aspects:

1. *Quality-enhancing structures*: use of standardized structures wherever possible.
2. *Real-time quality assurance practices*: extensive challenges to the findings and conclusions of the audit report.
3. *Post-evaluation quality assurance practices*: identification of lessons learned for improving future quality.

Each of these three types of quality assurance practice and how they might be applied to evaluation are examined.

Quality-Enhancing Structures

Evaluation would do well by adopting three formal phases to each evaluation: planning, assessment, and reporting.

The Planning Phase

The first part of a performance audit is the survey phase, which results in a survey or feasibility report which:

- introduces the area to be audited,
- sets out the objectives of the audit — the lines of investigation — to identify exactly what the audit will be concluding against,
- identifies the criteria to be used (usually agreed with the auditee) in assessing the specific aspects of performance of the program in question,

- states the scope of the audit — what aspects or elements of the program are being audited,
- outlines the audit methodology to be used and its cost,
- provides the timing of the audit, and
- discusses possible findings.

Feasibility or planning reports with these features should be used in evaluation, and they often are. Starting in the 1980s, evaluability assessments (Wholey, 1983) or evaluation assessments (Office of the Comptroller General, 1981) have often been undertaken. They contain many of the features of audit survey reports. In Canada, the Treasury Board Secretariat continues to recommend an evaluation assessment phase as part of the evaluation process (Treasury Board Secretariat, 2002).

There obviously remains a need for careful planning for evaluations, the more so for interview-based evaluations that rely heavily on the skills, experience, and integrity of the evaluators involved. Evaluation terms of reference may contain some of the information needed in a planning study, but the evaluators being used to carry out the work need to be involved in developing in more detail and with more understanding the key aspects of the evaluation plan, outlined in Table 3.

Table 3
Requirements for an Evaluation Plan Report

1.	<i>Program context.</i> A description of the program being evaluated, its external context, and previous significant evaluation findings.
2.	<i>Program logic/theory.</i> A description of how the program is supposed to work: its objectives, activities, outputs and outcomes, and their interrelationships.
3.	<i>Evaluation objectives.</i> A clear statement of the objectives of the evaluation; the matters the evaluation will conclude on.
4.	<i>Evaluation criteria.</i> The criteria the evaluation will use to assess performance, and an explanation of where the criteria came from.
5.	<i>Evaluation scope.</i> The scope of the evaluation; what aspects or elements of the program in question will be examined.
6.	<i>Evaluation methodology.</i> An outline of the methodology to be followed — what will be done in conducting the evaluation — and the cost involved.

The elements in Table 3 may sound like routine evaluation practice, but they are not always practiced. For example, the importance of carefully thinking through the objectives of the evaluation and then having to stick to them during the evaluation may not be ap-

preciated. As an evaluation (or audit) proceeds, and interesting new information is uncovered, there is a danger that the work can go off in unforeseen directions, and the original objectives may be lost sight of. The rigour of the audit process is meant to ensure that, in the end, the auditor *must* conclude against those objectives. This would also be good evaluation practice.

Similarly, the importance of explicit evaluation criteria cannot be underestimated. Most evaluations (and audits) should not be nor be seen to be fishing expeditions, developing criteria as the evaluation proceeds. The evaluation process should result in evidence being produced against each of the criteria identified in the evaluation plan, leading to a conclusion (a finding) against each of the evaluation objectives. These requirements need to be reinforced through other quality assurance practices, discussed below.

Setting out the context and theory of the program should link directly to the evaluation objectives and criteria. It can also be the basis for the methodology used in carrying out the evaluation.

The aim here is to produce an *evaluation plan report*. And this report needs to be challenged. The result of the planning phase is a decision (by those funding and perhaps advising on the evaluation) to either proceed as outlined or go back to the drawing board to re-think what ought to be done in the evaluation.

The Assessment Phase

The assessment phase is the normal phase of conducting the evaluation where the data and information are gathered and analyzed. Once the data and information have been collected, the findings and conclusions can be drafted, in order to answer the questions:

- What has been found with respect to the evaluation objectives, including an assessment with respect to the evaluation criteria used?
- What conclusions follow for each evaluation objective previously identified?

At this point in a performance audit, a draft report is prepared, sometimes just in point form, to outline what was found and what the evidence was for the findings. This, of course, adds time and cost to the audit, but is seen as a key decision point in the conduct of the

audit, where the auditor faces challenges from others inside and outside the audit organization regarding the strength of the evidence gathered and the robustness of the findings and conclusions.

In a similar fashion, a *preliminary evaluation report* could be usefully prepared at this point in an evaluation. If it withstands scrutiny, it becomes the evaluation report and hence need not add significantly to the cost and timing of the evaluation. If it does not, then further work or analysis is clearly required. Skipping this step in the latter case will almost guarantee an evaluation report that will be seen as unsatisfactory.

The Reporting Phase

In the end, a final evaluation report is required. A standardized structure report can further help to assure that the key elements expected are in the final report. Performance audit reports may have detailed structure requirements to them, and there is considerable variation among SAIs. But there are standards for the form of audit reports. For example, the Canadian Institute of Chartered Accountants auditing standards (2002) call for such elements in an audit report as the audit objectives and criteria, audit conclusions, level of assurance being provided, and auditing standards used. Similarly, evaluation reports might usefully contain the elements in Table 4.

Table 4
Required Elements of an Evaluation Report

Evaluation reports should include, as a minimum:

- Clear statement of the evaluation objectives
 - Clear description of the evaluation criteria used
 - Description of the evaluation methodology followed
 - Findings
 - Conclusions against the objectives
 - Recommendations
 - Statement of agreement or disagreement by the evaluatee
-

Given the structure outline above, the report would provide evidence and argument against the criteria set out at the beginning and would form conclusions (and, if asked for, recommendations) against the established evaluation objectives. Here again, I would argue for a healthy challenge to the findings and conclusions, and it is to this key aspect of quality assurance I now turn.

Real-Time Quality Assurance Practices: Challenging the Facts

Perhaps the most important aspect of the quality assurance practice of performance audits is the variety of challenges that are part and parcel of the audit process. Table 2 identified challenge practices both during the audit and after the audit is completed. Table 5 lists possible quality assurance challenge practices that evaluations could implement during the conduct of the evaluation.

Table 5
Quality Assurance Practices for the Conduct of Evaluations

-
1. Evaluator selection, guidance, subject-matter expertise, and responsibility
 2. Use of an advisory committee
 - comprising key stakeholders and outside subject-matter experts
 - advising at key points in the evaluation process
 3. Use of a quality reviewer
 - assuring that evaluation standards have been adhered to
 4. Clearance of draft report with programs evaluated
 - ensuring the facts are right
 - disagreements recorded in the report
-

Evaluator Selection, Guidance, Subject-Matter Expertise, and Responsibility

A clear distinction between evaluation and performance auditing practices is that, for the most part, performance audits are conducted by audit offices using their own staff, whereas evaluations are frequently conducted by individual evaluators hired for the evaluation. Audit offices know or should know what they are getting in terms of the skills and experience assigned to a performance audit. Organizations using their own staff to conduct evaluations similarly should know what skills and expertise they are using. On the other hand, organizations hiring evaluators are in a more difficult situation. This is probably more often the case with interview-based evaluations, as the professional judgement of the evaluators plays a key part in interpreting the information gathered.

Of course, in hiring individual evaluators, organizations check references and review past stated experience. But there can remain an element of uncertainty about what you are getting. Several evaluation societies provide guidelines and standards covering the skills and values required of evaluators (American Evaluation Association, 2004; Canadian Evaluation Society, 2004), but evaluators are

not certified by a professional society, so choice can only be based on past experience and reputation (Budget and Evaluation Division, 1999; IDRC, 2004; Perrin & Boschen, 1991). If the risk is seen as great, organizations sometimes hire a large reputable firm to carry out the evaluation.

Once hired, two things can be done at the outset to enhance the quality of the final evaluation:

- Provide the evaluator(s) with the quality criteria that the organization expects to be followed and that will be used to assess the final product.
- Make sure the evaluation team, together as a team, is provided orientation and guidance on the evaluation and its context, and that they together agree on the method, approaches, and tools that are going to be used.

The latter point is especially important in interview-based evaluations, since individuals on the team may split up to be able to visit more sites.

In the end, of course, the evaluator, as a professional, is responsible to ensure that the product is of adequate quality. The evaluator ought to be able to demonstrate that adequate attention to quality has been made. The Danish Ministry of Foreign Affairs requires the evaluation team leader to present a quality certificate to the Ministry when the final report is submitted (Danish Ministry of Foreign Affairs, 1999).

In selecting evaluators to carry out an evaluation, there is often a choice to be made between individuals who have solid evaluation expertise and individuals who are experts in the subject matter being evaluated. A good solution is to ensure that the evaluation team has both sets of skills and knowledge. However, there is still the question of who is in charge. Further, evaluations often have quite small teams — maybe only individuals — and the small team cannot have the range of skills and knowledge desired.

In performance auditing, the choice is clear. The auditors manage the audit but are required by their professional standards to ensure that adequate subject-matter expertise exists in the audit team. Typically, several things are done to acquire adequate subject matter expertise:

- Add expertise to the audit team, using external consultants if required.
- Acquire a certain level of knowledge themselves in the subject matter during the survey phase of the work, where they normally would interview experts and read material.
- Ensure that there is subject-matter expertise on the advisory committee formed to guide the audit.

Evaluators could adopt similar practices, as shown in Table 6. In my view, having subject-matter experts direct the evaluation is risky. As argued here, good evaluation practice needs to be followed to ensure credibility. Good evaluation requires evaluation expertise. Having evaluators lead the evaluation also helps to ensure independence where subject-matter experts may be perceived as having a conflict of interest if they are in charge.

Table 6
Acquiring Subject-Matter Expertise

1. Use the assessment phase to acquire a reasonable level of knowledge of the subject matter in question through interviews and literature review.
2. Add subject-matter expertise directly to the evaluation team.
3. Ensure good subject-matter expertise is part of the evaluation advisory committee to be able to forcefully challenge the evaluation findings and conclusions.

Advisory Committees

Some form of advisory committee is often part of the evaluation process. It provides advice and guidance to the evaluation as it is conducted. This is good practice. Such a committee should provide advice on the evaluation plan report, the preliminary evaluation report, and the final report.

Key to the success of such a committee, especially in its role of quality assurance, is the make-up of the committee. Based on performance auditing experience, such committees work well when they comprise a variety of different and indeed competing perspectives. For evaluation, the advisory committee could comprise: senior management in the organization; corporate management responsible for evaluation; experts in the area from the organization; external subject-matter experts; a quality reviewer; if needed, relevant specialists from the organization, such as legal advisers; the evaluators; and management from the program being evaluated.

Inclusion of the program management would differ from much performance audit practice where inclusion of the auditee is not normal practice. Here, however, the organization is still in the end evaluating itself; inclusion of the management from the program being evaluated makes sense, since evaluation is largely about learning. The committee can be a useful forum to help identify pertinent information and to get buy-in by the program management for the methods and approaches to be adopted. Given that this is an advisory committee, the inclusion of program management should not, in my view, unduly compromise the independence of the evaluation process. Indeed, as part of the committee, management will also get to hear a good cross-section of views on their program, and, in particular, more than just the views of the evaluators.

Perhaps the key here is the inclusion of the external experts. By including several known experts in the area, who are included for the specific purpose of providing their expert and independent advice on the evaluation, the robustness and completeness of the evaluation could be greatly enhanced. In particular, such members provide a check on program management and even corporate management self-interest.

This form of a strong advisory committee need not be costly in terms of either time or money. For a minimal investment, considerable expertise and credibility can be brought to the table in a structured but supportive way. While the process does involve assessments of the work to date by a number of persons, this is done in a less formal way than appears to be the case of the EU process using the MEANS criteria discussed earlier. There is discussion and debate about the evaluation with the evaluators present.

The committee need only meet two or three times during the evaluation, probably for a half day each time. In particular, the committee might meet when:

1. the planning phase is completed, to advise on the focus and approach being recommended;
2. the preliminary evaluation report has been drafted;
3. the draft final report findings and recommendations are prepared.

If actual meetings are difficult or expensive, the review of the draft final report could be done by providing written or oral comments.

At the preliminary report meeting, members would have obtained a good idea of the positions of the various members on the committee and had a chance to debate issues.

Evaluations are sometimes run by a steering committee, that is, a committee with directive power. There may be occasions when this is useful. However, in my view, the advisory role would normally be preferred. Such a role supports the professionalism of the evaluators doing the work, leaving it up to them to decide the extent to which they accept the advice they are receiving. Particularly when outside evaluators are being employed, there is a need to ensure that the organization and its advisors do not unduly influence the evaluation, as discussed above. A steering committee set-up might undermine the needed independence of the evaluators. If an external evaluation is being undertaken, the external evaluators need to be the ones who form the findings and conclusions, based on the evidence gathered.

It would be expected, however, that the evaluators provide documented reasons for not accepting the advice they receive. This is similar to the case for performance audits. Within an SAI, even though there are external advisors, and senior management to whom the auditor reports on a day-to-day basis, there is a strong acceptance of the fact that it is the auditor responsible for the report who is first and foremost responsible for its quality and integrity, and may for good reason not accept all the advice received.

In my experience, a strong and expert advisory committee is the most important quality assurance practice associated with performance audits. So it can be with evaluations. The committee can advise when:

- the evaluation is getting off target from its purpose,
- the findings agree or disagree with accepted wisdom,
- the evidence for the findings and conclusions is adequate or not, or
- the recommendations of the report are consistent with the findings and conclusions.

The evaluators can greatly benefit from this advice and the challenge it brings to the evidence, findings, and conclusions being formulated.

A Quality Reviewer

A number of key people on the advisory committee — especially the external advisors and the senior and program managers — will not know nor be expected to know whether the evaluator is following proper professional procedures in the conduct of the evaluation. Quality reviewers are experts in evaluation and the specific evaluation procedures in place in the organization. Their role is to assure the organization that the proper steps are being carried out. Fully implemented, they would have a sign-off role.

What aspects might they examine? They could look at such things as:

- whether there has been an evaluation plan report that contained the required elements,
- whether the evaluators have checked with organizational or outside experts on such methodological issues as questionnaire design, sampling plans, and so on,
- whether the evaluators have followed any procedures required by the organization,
- whether there is documented evidence for all the findings in the evaluation report.

In short, they check to see if the conduct of the evaluation has met organizational and professional standards. The level of work required by the reviewer would depend in large degree on the extent to which standards are in place.

Report Clearance

Draft performance audit reports are sent to those being audited to “check the facts.” Although there are some differences in practices among SAIs, for the most part, auditees are expected to review the draft audit report and confirm in writing that they agree with the facts presented. They are usually not expected to necessarily agree with the conclusions or recommendations presented. For the auditor, this is a check on a key aspect of the quality of the audit: Has the auditor missed or misrepresented key facts? For their part, the auditees are expected to produce evidence for any claims they make that the auditors got it wrong. Because facts are not always black or white, the auditors must decide if they need to make a change in their draft report.

The same process can apply to evaluations. And indeed, program management, as part of the advisory committee, are there to check the facts and interpret the data as the evaluation goes along, and should be expected to agree with the final findings of the evaluation. If they disagree with aspects of the evaluation, they should be allowed to say so and have it appear in the final evaluation report.

Post-evaluation Quality Assurance Practices

Once completed, an evaluation can still contribute to improving the quality of future evaluations in the organization, if the lessons learned are captured and fed back into the evaluation process. Audits, like evaluations, rarely go smoothly or exactly as planned, and SAIs routinely employ a number of practices aimed at capturing this valuable information. Table 1 listed SAI post-audit practices. Table 7 summarizes these approaches from an evaluation perspective.

Table 7
Post-evaluation Quality Assurance Practices

-
1. Self-assessment of lessons to be learned by the evaluators who carried out the evaluation
 2. Independent assessment of the quality of the completed evaluation
 - by a central evaluation quality assurance group
 - by external experts
 3. An independent survey of those evaluated on the evaluation process used
 4. Public scrutiny of the evaluation report
-

In all cases, the lessons and assessments are discussed within the organization, and, if appropriate, changes are made to the evaluation process and standards.

As a standard part of the evaluation process, the evaluation team could prepare a report from their perspective on lessons learned in carrying out the evaluation. This could be part of the contract for external consultants. This lessons-learned report could be discussed with evaluation staff and other interested parties.

In addition to this self-evaluation, or perhaps on a selective basis, a central evaluation group could assess the completed evaluation and prepare a quality practice report, which would be discussed with the evaluation team and could lead to changes in the evaluation process. As discussed earlier, this type of quality review is carried

out by the Operations Evaluation Department at the World Bank and by IDRC.

A further step in post-evaluation review could be to get outside experts to review the quality of the evaluation to identify areas of improvement in the evaluation process. This review might focus more on the substance of the evaluation — its findings and conclusions — than on the process of how the evaluation was conducted. The UK National Audit Office has each of its completed performance audits assessed by a team from the London School of Economics (National Audit Office, 2002).

An oft-used audit quality review practice that could be adopted would be to survey those evaluated on their views of how the evaluation was carried out, and whether the resulting report was useful. If anonymity were promised, views other than those expressed during the evaluation might be set forth. Again, the lessons learned could help improve future evaluations.

Finally, if it were known that there would be an explicit public review and scrutiny of the evaluation report, such as by a legislative committee, additional care would probably be taken to ensure that the facts were sound and the conclusions and recommendations were defensible.

All of these post-evaluation review practices can provide useful information for future evaluations and can be carried out with modest incremental resources. Without procedures such as these, valuable lessons learned from completed evaluations will be lost.

CONCLUSION

A number of quality assurance practices that organizations could put in place to strengthen the quality of interview-based evaluations have been outlined and discussed. More interestingly, they could be used to ensure the quality of evaluations more generally. The practices, while developed with the interview-based approach in mind, stand alone. They are not dependent on the evaluation methodology being used.

It is not suggested that no attention is being paid to quality assurance practices for evaluation in organizations. Although actual quality assurance practices for evaluation have not been reviewed, my

experience is that some of these practices are indeed found in some organizations. Nevertheless, organizations could consider the quality assurance practices suggestions made above.

In considering these suggestions, several questions arise.

Would these quality assurance practices work? While the answer depends on the rigour and robustness of the application of these practices, the experiences from performance audits strongly suggest they would indeed work. Differences between performance audits and evaluations have been noted.

Having said this, we need to be clear about what is meant by this approach “working.” Mature SAIs have over many years developed a reputation for producing quality work, often to the extent that the quality of their audits is only rarely questioned. Most organizations carrying out evaluations do not have such reputations, especially since the evaluations are seen as self-evaluations. Their evaluations are likely to be challenged and questioned. I suggest that the quality assurance practices discussed here will result in good quality evaluations. With a series of such evaluations under its belt, an organization may indeed develop a reputation for quality.

Would these practices significantly increase the cost and time required to carry out evaluations? This is a key question. These quality assurance practices would certainly increase the cost and perhaps the time required for conducting an evaluation. Quality does cost, and the organization must make a commitment of time and resources. In my view, appropriately implemented, the resulting enhanced quality and credibility is worth the cost.

Without such procedures, the quality of the report depends — certainly for interview-based approaches — almost entirely on the individual expertise and views of the evaluators used. Without an adequate challenge process, even the most professional and skilled evaluator can get off track, get the facts wrong, or reach unsupported conclusions.

If the expected evaluation process is outlined beforehand, the various quality assurance steps can be built into the schedule and the evaluators (especially in the case of outside evaluators) made quite aware of what is expected of them. Planned out beforehand, quality assurance ought not to add significantly to the timing of an evaluation.

But there are more steps to go through, and steps do cost money. Without getting into implementation details, it would be possible to gradually implement these practices and/or to implement them in a modest rather than a full-blown manner described above, as practiced in SAIs. But in most SAIs, these practices are seen as essential to the conduct of an audit, not to be lightly skipped.

Would these quality assurance practices be applicable to all types of evaluations? I suggest that the answer is both yes and no. Formative evaluations are done primarily for program management aimed at improving a new or ongoing program or project. They may be subject to less outside scrutiny. As a result, the formal quality assurance practices discussed here may not be as pertinent, or might be used more selectively. Summative evaluations tend to be more widely scrutinized and may more often need to be able to demonstrate quality. In either case, in situations where there will be greater challenges to the quality of the evaluation, there is a greater need to pay attention to quality assurance of the evaluation process.

Further, these quality assurance practices would seem to be less applicable to research-type evaluations conducted by researchers trying to establish new knowledge. In these cases, the evaluation design is usually fairly sophisticated, and the ultimate test on the quality remains the extent to which the results are replicable. Thus, there could be trade-offs between the robustness of the evaluation design used and the extent to which all of the suggested practices are put in place.

What risks exist if robust and credible quality assurance practices are not adopted? A key focus of the approach suggested here is balance. In the absence of good quality assurance practices, there is considerable risk of one of two outcomes occurring. On the one hand, the self-interest of the organization or of program management — who are paying for the evaluation — dominates the process, and the evaluation is not able (or is not allowed) to present critical findings. Or the evaluation is, or is viewed as, weak — perhaps incorrect in key facts, biased by poor methodology, or biased by the evaluator's perspective — and is not credible, and thus cannot be used by the organization or program. Neither of these is a good outcome.

Evaluation in an organization is usually less secure than is performance audit. The value of evaluation depends in large part on the quality of the evidence and findings it is able to bring to the table.

Quality assurance ought to be a key concern of those involved in the evaluation process. The practices discussed in this article, based on a proven track record in performance audit, can provide a basis for focusing attention on approaches for improving and indeed ensuring that quality evaluations are undertaken.

NOTE

1. The UK National Audit Office has contracted out a few performance audits, with mixed results. They worked best when contracted out to an academic-based organization that approached the audit with a critical mindset and were not overly concerned about future business with the organization being evaluated.

REFERENCES

- American Evaluation Association. (2004). *Guiding principles for evaluators*. Retrieved September 18, 2004, from <<http://www.eval.org/Guiding%20Principles.htm>>.
- Bowerman, M. (1996). The rise and fall of VFM auditing. In I. Lapsley & F. Mitchell (Eds.), *Accounting and performance measurement: Issues in the private and public sectors* (pp. 193–212). London: Paul Chapman.
- Budget and Evaluation Division. (1999). *Working well with evaluation consultants: A guide*. Yellowknife: Financial Management Board Secretariat, Government of the Northwest Territories.
- Canadian Evaluation Society. (2004). *CES guidelines for ethical conduct*. Ottawa: Author. Retrieved September 18, 2004, from <http://evaluationcanada.ca/site.cgi?s=5&ss=4&_lang=en>.
- Canadian Institute of Chartered Accountants. (2002). *General assurance and auditing, Section 5025: Standards for assurance engagements*. Toronto: Author.
- Corbeil, R.C., & McQueen, C. (1991). Improving the quality of evaluation. In A. Love (Ed.), *Evaluation methods sourcebook I* (pp. 196–213). Ottawa: Canadian Evaluation Society.

- Danish Ministry of Foreign Affairs. (1999). *Evaluation guidelines, Section 6.1: Quality of information*. Retrieved April 8, 2004, from <<http://www.um.dk/danida/evalueringsrapporter/eval-gui/index.asp>>.
- Datta, L.-E. (2000). Seriously seeking fairness: Strategies for crafting non-partisan evaluations in a partisan world. *American Journal of Evaluation, 21*(1), 1–14.
- European Commission. (1999). *Evaluation of socio-economic programmes, Volume 1*. Luxembourg: Office des Publications.
- Grasso, P. (2005). Quality of evaluative information at the World Bank. In R. Schwartz & J. Mayne (Eds.), *Quality matters: Seeking confidence in evaluation, auditing and performance reporting* (pp. 91–112). New Brunswick, NJ: Transaction.
- International Development Research Centre. (2002). *Quality assessment of IDRC evaluation reports*. Evaluation Guideline No. 4. Ottawa: Author. Retrieved September 20, 2004, from <<http://web.idrc.ca/uploads/user-S/108549993014guideline-web.pdf>>.
- International Development Research Centre. (2004). *Selecting and managing a consultant or team*. Evaluation Guideline No. 8. Ottawa: Author. Retrieved September 20, 2004, from <<http://web.idrc.ca/uploads/user-S/108550007718Guideline.pdf>>.
- International Organization of Supreme Audit Institutions. (1992). *Auditing standards*. Auditing Standards Committee. Retrieved July 8, 2003, from <http://www.intosai.org/2_CodeEth_AudStand2001_E.pdf>.
- Joint Committee on Standards for Educational Evaluation. (1994). *The program evaluation standards: How to assess evaluations of educational programs* (2nd ed.). Thousand Oaks, CA: Sage.
- Leeuw, F. (1992). Performance auditing and policy evaluation: Discussing similarities and dis-similarities. *Canadian Journal of Program Evaluation, 7*(1), 53–68.
- Lonsdale, J., & Mayne, J. (2005). Neat and tidy ... and 100% correct: Assuring the quality of SAI performance audit work. In R. Schwartz & J. Mayne (Eds.), *Quality matters: Seeking confidence in evaluation, auditing and performance reporting* (pp. 173–198). New Brunswick, NJ: Transaction.

- Nairobi M&E Network, African Evaluation Association Secretariat, Réseau Nigérien de Suivi et Évaluation, Cape Verde Evaluation Network, Réseau Malagache de Suivi et Évaluation, Comoros Evaluation Network, Eritrean Evaluation Network, Malawi M&E Network, Réseau National de Chercheurs et Évaluateurs de Burundi, Rwanda Evaluation Network, & UNICEF Eastern and Southern Africa Region M&E Network. (2002). The African evaluation guidelines: 2002. *Evaluation and Program Planning*, 25(4), 481–492.
- National Audit Office. (2002, July). *Focus*, 12. London. Retrieved September 23, 2004, from <<http://www.nao.org.uk/guidance/focus12.pdf>>.
- Office of the Comptroller General. (1981). *Guide on the program evaluation function in federal departments and agencies*. Ottawa: Supply and Services Canada.
- Perrin, B., & Boschen, K.A. (1991). Getting the best from consultants. In A. Love (Ed.), *Evaluation methods sourcebook I* (pp. 176–195). Ottawa: Canadian Evaluation Society.
- Power, M. (1997). *The audit society: Rituals of verification*. Oxford: Oxford University Press.
- Roberts, S., & Pollitt, C. (1994). Audit or evaluation? A National Audit Office VFM study. *Public Administration*, 72(Winter), 527–549.
- Schwartz, R. (1999). Coping with the effectiveness dilemma: Strategies adopted by state auditors. *International Review of Administrative Sciences*, 65, 511–526.
- Schwartz, R., & Mayne, J. (Eds.). (2005). *Quality matters: Seeking confidence in evaluation, auditing and performance reporting*. New Brunswick, NJ: Transaction.
- Sutherland, S. (2001). “Biggest scandal in Canadian history”: HRDC audit starts probity war. Working Paper 23. Kingston, ON: School of Policy Studies, Queen’s University.
- Toulemonde, J., Summa-Pollitt, H., & Usher, N. (2005). Triple check for top quality or triple burden? Assessing EU evaluations. In R. Schwartz & J. Mayne (Eds.), *Quality matters: Seeking confidence in evaluation, auditing and performance reporting* (pp. 69–90). New Brunswick, NJ: Transaction.

- Treasury Board Secretariat. (2002). *Program evaluation methods*. Ottawa: Author. Retrieved September 18, 2004, from <http://www.tbs-sct.gc.ca/eval/pubs/meth/pem-mep06_e.asp>.
- Wholey, J.S. (1983). *Evaluation and effective public management*. Boston: Little, Brown.
- Widmer, T. (2005). Instruments and procedures for assuring evaluation quality: A Swiss perspective. In R. Schwartz & J. Mayne (Eds.), *Quality matters: Seeking confidence in evaluation, auditing and performance reporting* (pp. 41–68). New Brunswick, NJ: Transaction.
- Widmer, T., & Beywl, W. (2000). Die Übertragbarkeit der Evaluationsstandards auf unterschiedliche Anwendungsfelder. In Joint Committee on Standards for Educational Evaluation (Hrsg.), *Handbuch der Evaluationsstandards* (pp. 243–257). Opladen: Leske+Budrich.
- Wisler, C. (Ed.). (1996). *Evaluation and auditing: Prospects for convergence* (New Directions for Evaluation, No. 71). San Francisco: Jossey-Bass.

John Mayne is an independent consultant. In the 1980s and 1990s, Dr. Mayne worked at the Treasury Board Secretariat and was instrumental in the development of the federal government's approach to evaluating the performance of programs. In 1995, he joined the Office of the Auditor General where he led efforts at developing effective performance measurement, managing for results, and performance reporting practices in the government of Canada, as well as methodologies for providing audit assurance on performance reports. He also led the Office's audit efforts in accountability and governance, retiring from there at the end of 2003. He has authored numerous articles and reports, and edited six books in the areas of program evaluation, public administration, and performance monitoring. In 1989 and in 1995, he was awarded the Canadian Evaluation Society *Award for Contribution to Evaluation in Canada*.