



## **ROOM DOCUMENT 2**

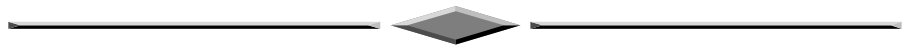
**DAC Network on Development Evaluation**

### **REVIEW OF THE QUALITY OF EVALUATIONS**

#### **Item 2: i**

This document consists of two parts: A Status Report and the 'Minimum Sufficient' Group of Evaluation Standards.

This note has been prepared by Australia for consideration at the meeting of the DAC Network on Development Evaluation on 15-16 January, 2004.



**1st meeting  
15 – 16 January 2004**



## **REVIEW OF THE QUALITY OF EVALUATIONS**

### **Background**

At the 37th meeting of the DAC Working Party on Aid Evaluation, a discussion started on the quality of evaluations. Some members felt that development evaluations were of variable quality, from excellent to poor, while others noted that the criteria for assessing what constituted a good or poor report were not generally developed and agreed. Australia indicated interest in leading a review of the quality of evaluations together with interested members.

At the 37th Meeting of the DAC Working Party on Aid Evaluation, following ad hoc discussion about the poor quality of evaluations, Australia indicated interest in leading a review of the quality of evaluations.

The Australian member was asked, and agreed, to lead a Task Force to develop the Terms of Reference for such a review.

### **Objective**

The objectives of this activity are to:

- Select and test suitable criteria and standards to assess the quality of evaluations.
- Use these criteria and standards to assess evaluations by members of the DAC Network on Development Evaluation.
- Identify strengths and shortcomings in evaluations and recommend changes in the implementation of evaluations to overcome deficiencies.

### **Outputs**

The outputs of this activity will include agreed standards for the assessment of the quality of evaluation reports. These could be useful to members in evaluation areas, assisting them in their practical work of ensuring that they produce a high quality product.'

### **Developments to date**

The first phase of the exercise was the development of criteria/standards with which to evaluate the DAC member evaluations. The criteria had to be sufficiently robust to be used in the assessment of a diverse range of reviews and evaluations, and defensible. The number of criteria need to be the minimum that are sufficient to assess the important aspects of each evaluation.

An Australian consultant was contracted to develop the criteria. After examination of other options, the criteria were derived from the existing DAC Principles for Evaluation of Development Assistance. The initial list included 20 criteria.

The criteria were tested in an assessment of a sample of 10 evaluation reports from 9 aid donors.

The Draft Report from the first phase of the exercise was distributed to Network members in July 2003. The covering letter from Australia discussed the elements of Phase 1, provided additional comments on the consultant's report, and requested comment on the draft report. It also asked members to advise if they:

- Saw value in proceeding to the second stage of the review.
- Wished to be actively involved in the exercise by providing resources to the second stage of the review.
- Could provide complete copies of recent evaluations for the review.

Following receipt of comments from Network members, the Australian consultant modified the "Minimum-Sufficient" Group of Evaluation Standards to take account of the main criticisms at that stage.

Network Members also raised other issues, the main one being how the review was to manage the assessment of evaluation reports available in, at least, three languages.

On 21 October a note from Mr Ellis was distributed to Network members to try to clarify the status of the exercise. This discussed:

- Task Force Membership.

From responses received, it appeared that the following people are interested in being members of the Task Force for this exercise:

Rob van den Berg, Netherlands

Lars Elle, Denmark

Horst Breier, Germany

Anne Bichsel and/or Regula Herlan, Switzerland

Anne-Marie Cabrit, France

Eva Lithman and/or Goran Schill, Sweden

Hans Lundgren, DAC Secretariat

- The Minimum-Sufficient Group of Evaluation Standards.

Comments were provided to Australia on the initial version of the Evaluation Standards. The Australian consultant had studied the suggested changes and modified the Standards (see above).

Some additional suggestions and concerns have been received but no further changes have been made to the standards. In part, it was considered that the proposed assessments will show strengths and weaknesses of the assessment method as well as those of the evaluation reports.

However, there is still room to discuss alternative proposals on how else we might, still in a practical and timely manner, get agreement on a set of criteria.

- Resources likely to be available to Phase 2 of the Review.

Correspondence appeared to show offers of assistance of up to four consultants (from Switzerland, Sweden, Netherlands and Australia).

It was expected that most of the reports to be assessed will be in English, French or German, and therefore consultants proficient in each of these languages will be required, with several bilingual consultants. It was assumed that this will be most practical for the European-based consultants. (Australian consultants are often bilingual, but usually in Asian or Pacific languages.)

Among the first requirements will be to check whether reports in other languages will be assessed, and to translate the draft Evaluation Standards into the requisite languages.

- The availability of evaluation reports to be assessed in the Review.

So far, Switzerland, Finland, Austria, Sweden and the United States have offered reports for assessment. We need to know the total number of reports that will be available, and how many in each language. This information will determine the 'weighting' of the languages of consultants, and how much consultant input will be required. There have been several claims that a proper assessment of an evaluation report takes **2 - 2.5 days**.

As far as the selection of reports is concerned, the preference is for a random, complete, or at least representative selection (i.e. not biased towards "good", or "bad" reports, or indeed to any other category disproportionate to the overall population of reports) of recent evaluation reports from each country.

If a sub-sample of these reports is required then an appropriate sampling method will be required.

- A possible way forward, taking account of members' concerns.

If there are four consultants, one should be appointed lead consultant, who will arrange distribution of reports for assessment, collation of results, coordination and editing of the draft report. Australia is prepared to provide this input (depending on consultant availability.) Agencies contracting the consultants will have to agree on the roles and responsibilities to be written into contracts.

Discussion is required of potential difficulties which might occur, particularly with translation.

Australia expects to come to the Network meeting with answers to questions of how to determine inter-reviewer reliability, having applied some statistical expertise to the question to help establish what type and amount of evidence we need to feel confident on this matter.

It was proposed that the assessment team provide each donor with the ratings and notes for each report submitted by that donor, but that reporting to the Network as a whole will be of the amalgamated results. This should maximise the benefits to each agency and the Network as a whole.

- Consultation.

The note suggested that 'procedural' communication be limited to Task Force members, with reports to be distributed to the Network as a whole, at suitable stages.

At the request of the Network Chair, the Australian delegate will present the Minimum-Sufficient Group of Evaluation Standards at the First Meeting of the DAC Network on Development Evaluation, and will lead discussion of all of the issues raised in previous correspondence and how the exercise should proceed from here.

## PART TWO

### THE ‘MINIMUM-SUFFICIENT’ GROUP OF EVALUATION STANDARDS

#### I. Overview

The ‘Minimum-Sufficient’ Group of Evaluation Standards has been developed to assess the quality of DAC Member evaluation *reports*. These standards are not intended to assess the quality of the demand, the quality of the evaluation function or the dissemination and integration of recommendations in Member countries.

The Standards have been organised into groups in accordance with the existing DAC Principles for Evaluation of Development Assistance (hereinafter referred to as the DAC Principles). The Standards have also been influenced by the interpretations of these principles contained within the relevant DAC documentation<sup>1</sup>.

This approach has been taken because the Principles are still broadly relevant and are well known and established in the evaluation offices of Members. Nevertheless, it should be noted that this approach creates some difficulties; for example, the Credibility principle and standards are essentially dependent upon the Participation and Design and Implementation principles and standards which is not ideal. Nevertheless, the potential benefits of maintaining the principles that are known and accepted by Members outweigh the difficulties created thereby.

The Principles of specific relevance<sup>2</sup> to the current TOR relate to evaluation purpose, impartiality and independence, credibility, usefulness, participation of donors and recipients, design and implementation and reporting.

Each principle (and group of standards) is first briefly summarized and introduced. These introductory statements are quoted and/or derived from the relevant OECD DAC documentation<sup>3</sup>.

The individual standards in each group are numbered, given a short title, and briefly described. The individual standards as well as the six groups of standards are left unweighted. This approach is not only relevant with respect to the current activity; it also recognises that the significance of any given standard or group of standards varies from evaluation to evaluation.

---

<sup>1</sup> Development Assistance Committee “Principles for Evaluation of Development Assistance” OECD Paris 1991; DAC Working Party on Aid Evaluation “Review of the DAC Principles for Evaluation of Development Assistance” Paris 1998

<sup>2</sup> Those principles or points within principles, that are not relevant to the current TOR have been omitted. A detailed analysis and justification is provided at Annex 2.

<sup>3</sup> Opcit OECD 1991 DAC WPAE OECD 1998

## II. Evaluation Standards

### *Evaluation Purpose*

The main purposes for the conduct of evaluations are:

- To improve future aid policy, programmes and projects through feedback of lessons learned.
- To provide a basis for accountability, including the provision of information to the public.<sup>4</sup>

- |   |
|---|
| 1. <b>Clear Purpose</b> The Terms of Reference and the Evaluation Report clearly state the primary purpose of the evaluation as determined by the commissioning body. |
|---|

### *Design and Implementation of Evaluations*

Each evaluation must be planned and terms of reference drawn up which adequately define the purpose, the evaluation issues to be addressed, stakeholders, methodology, performance standards, resources and budget required to complete the evaluation.<sup>5</sup>

- |   |
|---|
| 2. <b>Stakeholders Identified</b> The report clearly identifies the stakeholders participating in, and affected by, the evaluation. Specifically, the report identifies: (i) the ultimate beneficiaries of the objective of evaluation; (ii) those persons who most need to learn from the evaluation; and (iii) those who are best positioned to implement the recommendations contained in the evaluation report. |
|---|

- |   |
|---|
| 3. <b>Evaluation Objectives Clear</b> The report clearly describes the objectives of the evaluation and the process adopted to ensure that all stakeholders understand the objectives of the evaluation. Clarifying the objectives of an evaluation is often not fully possible at the outset of an evaluation, but instead calls for a lengthier process that should be regarded as a central element of the evaluation process itself. Where this is so, the process of clarifying the objectives is clearly described. |
|---|

- |   |
|---|
| 4. <b>Relevant Scope</b> The Terms of Reference [TOR] and the Evaluation Report clearly state the areas to be addressed by the evaluation [scope of work]; the information identified for collection; the standard against which performance is to be assessed or analyses are to be conducted; the resources and time allocated and reporting requirements. The TOR render it possible to ask pertinent questions about the object of evaluation and take into account the interests and needs of the parties commissioning the evaluation, as well as other stakeholders. |
|---|

---

<sup>4</sup> Development Assistance Committee “Principles for Evaluation of Development Assistance” OECD Paris 1991 Section II paragraph 6.

<sup>5</sup> Ibid Section IX paragraphs 32 ff

5. **Precise description of the Object of Evaluation** The object of evaluation, be it a measure, program, or organization, is clearly and precisely described, documented, and unambiguously identified. Particular attention has been paid to any discrepancies between the original form the object of evaluation was anticipated to take and its actual form in practice or when implemented.

6. **Defensible Methodology** The questions to be addressed in the evaluation, and the methods and procedures chosen to address these questions, have been carefully documented. Contextual constraints have been identified and methods for dealing with these constraints have been explained. The report contains a detailed description of the organization of the evaluation, data collection and processing, analysis and reporting. Major methodological options have been discussed, including the risks associated with alternative options, and choices justified. Selected methods and procedures have been applied as stated and in accordance with their own quality standards (e.g. statistical tests, validity thresholds, attrition biases). Limitations faced in data collection and analysis have been described. Any changes that have occurred in proposed methods and procedures during the course of the evaluation have been described and justified.

7. **Valid and Reliable Information** The data collection instruments selected, developed and employed are valid and reliable. Validity is determined by assessing the degree to which the instruments employed accurately reflect the concepts they are intended to measure. Reliability refers to the consistency or stability of the quality measured, whether between measurement instruments, persons, or over time. All potential biases or errors are systematically identified, analysed and corrected as far as possible by recognised techniques.

8. **Sound Analysis** Data are appropriately and systematically analysed or interpreted according to the state of the art. Major cause-and-effects relationships and underlying assumptions are made explicit. Critical exogenous factors have been identified and taken into account.

### *Credibility*

The credibility of an evaluation depends on the expertise and independence of the evaluators and the degree of transparency of the evaluation process.<sup>6</sup>

9. **Demonstrated Professionalism and Competence** The evaluation report demonstrates the competence and trustworthiness of the evaluators. [Note: Performance with respect to this standard must be deduced rather than directly assessed. Evaluation reports assessed as being of a high quality with respect to other standards (e.g. 8-20) would be rated high with respect to this standard.]

10. **Transparent Evaluation Process** The evaluation report contains a clear and sufficient explanation of the process and methods for conduct of the evaluation that is accessible to relevant stakeholders. [Note: Performance with respect to this standard is partially dependent upon assessed performance with respect to 12, 13, 14, 17, 20].

<sup>6</sup> Opcit OECD Paris 1991 Section IV paragraph 18.



## *Usefulness*

For an evaluation “to have an impact on decision-making, the findings, conclusions and recommendations must be perceived as being relevant and useful and be presented in a clear and concise way. They should fully reflect the different interests and needs of the many parties involved in development cooperation. Easy accessibility is also crucial for usefulness. The evaluation process itself promotes a further clarification of objectives, improves communication, increases learning, and lays the groundwork for follow-up actions. Evaluations must be timely in the sense that they should be available at a time that is appropriate for the decision-making process”<sup>7</sup>.

11. **Demand Responsive.** The evaluation report adequately addresses the information needs of the commissioning body. It answers all question included in the Terms of Reference in a way that reflects the stated level of priority.

12. **Robust Findings.** The report provides stakeholders with a substantial amount of new knowledge (findings). Findings are clearly identified. They follow logically from, and are justified by, data, interpretations and analyses through logical reasoning that are carefully described and do not contradict each other. Logical reasoning is developed as far as possible and necessary. When relevant, the report indicates which findings are generalisable and under which conditions.

13. **Clear Conclusions.** The conclusions reached in the evaluation report are clearly and explicitly described, together with their underlying assumptions.

14 **Useful Recommendations.** Recommendations are not mixed with conclusions, but they are derived from them. Recommendations are presented in sufficient detail and with an operational focus. The report indicates that practical constraints have been taken into account when formulating recommendations (e.g. regulations, institutions, budget).

## *Impartiality and Independence*

“Impartiality and independence are closely inter-related concepts. In fact, the aim of impartiality is best achieved where evaluation activities are independent from operations personnel and managers who have interests in showing accomplishments and good performance. Impartiality also depends on the professionalism of evaluators and the methodology applied.”<sup>8</sup>

15. **Impartial and Independent Evaluation Function** The evaluation report clearly indicates the degree of independence of the evaluation function from the operations and management functions. Conflicts of interest are addressed openly and honestly so that they compromise the evaluation process and conclusions as little as possible.

<sup>7</sup> Ibid Section V paragraphs 21 and 22.

<sup>8</sup> DAC Working Party on Aid Evaluation “Review of the DAC Principles for Evaluation of Development Assistance OECD Paris 1998 Pg 24

16. **Complete and Balanced Assessment** The evaluation report is complete and balanced and presents the strengths and weaknesses that exist in the object being evaluated, in a manner that strengths can be built upon and problem areas addressed.

17. **Impartial and Substantiated Conclusions** The process employed in reaching conclusions is described. [It should be noted that conclusions go further than findings in the sense that they include value judgements.] The conclusions are based on explicit and agreed judgement criteria and benchmarks. The judgement criteria take into account all legitimate standpoints. Conflicting points of view and issues are presented in a balanced way. There are no discrepancies between stated criteria and benchmarks and those that have been actually applied.

18. **Neutral Reporting** The evaluation report is free from distortion through personal feelings or preferences on the part of any party to the evaluation. Evaluation reports present conclusions in a neutral manner.

#### *Participation of Donors and Recipients*

“Whenever possible both donors and recipients should be involved in the evaluation process.”<sup>9</sup>

19. **Evidence of Participation** The evaluation report details the way in which donor/recipient participation has been encouraged in the planning, execution and presentation of the evaluation. Where the purpose of an evaluation is to investigate the impact of the object of evaluation, be it a measure, program or organization, on the lives and welfare of beneficiaries evidence of participation/consultation with those beneficiaries is provided.

#### *Reporting*

20. **Comprehensive and Clear Reporting** The final evaluation report is logically structured and outlines the evaluation context, goals, questions posed, and procedures used, as well as any constraints encountered that substantively hindered its ability to fulfil its purpose and adhere to good evaluation practice. The findings, conclusions and recommendations reached are outlined in such a manner that the most pertinent information is readily accessible and comprehensible.

The report is free of superfluous information and analyses that do not substantiate the conclusions.

The report contains a short executive summary that highlights the key findings, conclusions and recommendations in a balanced and impartial way. Only appendices contain technically difficult information that is not accessible to all stakeholders.

---

<sup>9</sup> Ibid Section VI paragraph 23