



ROOM DOCUMENT 9

DAC Network on Development Evaluation

REVIEW OF THE QUALITY OF DAC EVALUATIONS (REPORTS)

Item IV: d

This note has been prepared by Australia for discussion at the meeting of the DAC Network on Development Evaluation on 9 – 10 November, 2004.



**2nd meeting
9 -10 November 2004**

REVIEW OF THE QUALITY OF DAC EVALUATIONS (REPORTS)

Summary

Since the first meeting of the DAC Network on Development Evaluation there has been further consideration of the issues which need to be resolved to advance the proposed review of the quality of DAC member evaluations. These are discussed in this paper under the headings of:

- Disclosure of results of evaluation assessments. This also discusses the range of reports to be included in the assessments.
- Focus of the review. This briefly discusses the products of the review.
- Number of evaluations to be assessed. The high cost of assessments will need to be taken into account in determining numbers of reports to be evaluated.
- Adequacy of the current 'Minimum sufficient' standards. Two activities are discussed which examined the standards.

This paper identifies the decisions which the Network should make before it can move forward on this exercise. Recommendations are included for each of the issues discussed.

As it is unlikely that time will be available at the November meeting to make all the necessary decisions, an approach is proposed under 'Next Steps'.

Background

Australia agreed to lead a Task Force to review the quality of DAC members' evaluations at the 37th Meeting of the (then) DAC Working Party on Aid Evaluation (March 2003).

The initial objectives of this review were to:

- Select and test suitable standards to assess the quality of evaluation reports.
- Use these 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.

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

An Australian consultant was contracted to develop the standards. After an examination of options, the standards were derived from the existing DAC Principles for Evaluation of Development Assistance, but with reference to the SEVAL, and other standards. The initial list included 20 standards.

The Draft Report from the first phase of the exercise was distributed to Network members in July 2003. 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.

Network Members raised other issues, including how the review would assess evaluation reports which were produced in languages other than English. A number of countries offered to support the assessment exercise and/or to provide reports to be included in the assessment.

At the request of the Chair, the Australian delegate presented the (revised) Minimum-Sufficient Group of Evaluation Standards at the First Meeting of the DAC Network on Development Evaluation in January 2004 and led discussion on the issues raised in previous correspondence and how the exercise could proceed.

The Summary Record of the meeting reported that members welcomed the initiative to conduct a systematic analysis of the quality of evaluation reports. A number of suggestions for modifications to the Minimum Standards were provided, which Australia was asked to take into consideration for a revision of the paper. Members were asked to provide further comments in writing by the end of February. Subsequent e-mail correspondence from Peter Ellis (former Director, Program Evaluation Section, AusAID), relayed to Network members in June 2004, requested additional comments on four key issues¹.

Purpose

This note summarises responses to the four issues raised by Mr Ellis, and raises additional matters which should be resolved prior to the commencement of an assessment of evaluation reports.

This report is presented as one component of a proposed broad inquiry into evaluation quality, systems, structures and processes, which will be discussed under other agenda items.

Issues

1. Disclosure of results of evaluation assessments

At the January 2004 meeting of the Network there was a consensus that there should be full disclosure of the identity of evaluations examined and of their respective scores ie. Complete transparency. Comments received since the meeting have not resiled from this view. These include (paraphrasing in some cases):

- 'A 'league table' could be a useful tool for internal advocacy'. - This was an alternative view to negative comments on 'league tables'.
- 'The assessment will establish benchmarks for each agency, against which they can measure their own future performance, and for the DAC Network as a whole.'
- 'We need to be careful about reports prepared by consultants, who may not want their work rated.'
- 'There could be scope for self-assessment of evaluations, not just by a panel of consultants.'

1 . (a) Disclosure of identity of evaluation assessments, (b) Focus on positives and lessons for improvement, (c) Details of 'Minimum sufficient' criteria, (d) Number of evaluations to be assessed.

There were queries whether evaluations done by units other than central evaluation units should be included in the assessments. One agency considered it would be useful to assess its 'decentralised' evaluations. This would have benefits in identifying the strengths and weaknesses of evaluations managed and/or implemented by other than evaluation experts. It would provide lessons on the extent and type of guidance required by 'part-time' evaluation managers and by those central evaluation units which are not yet staffed by experienced evaluators.

Another question was how joint reports, which are becoming the key product of some donors, would be treated. These could also be accommodated within the assessment without difficulty.

Another member suggested looking at evaluation reports of executing agencies.

Recommendations:

- There should be full disclosure of evaluations and their respective ratings.
- Evaluation reports from other than central evaluation units, including joint reports, should be included in the assessment at the request of members responsible for the evaluations.

2. Focus of the review

The January 2004 Network meeting agreed that the focus of the review should be on the 'positives': developing lessons for improving evaluations of all Network members and identifying good/best practice. This view was reinforced by later comments received.

The review will also enable the 'standards' to be improved for use in future assessments. A modified version would be useful to evaluators to advise them on the requirements for high quality evaluations

Recommendations:

- That the products of the review will be:
 - An assessment of the strengths and weaknesses of members' evaluations, advice on how to improve quality, and examples of best practice.
 - Revised standards, for use in future assessments and to advise evaluators on requirements for high quality evaluations.

3. Number of evaluations to be assessed.

There are differing views about the number of evaluations from each donor to be assessed. A decision on this point will be 'informed' by a table showing how many evaluations have been produced by each of the DAC members. This information, along with the language of each report, will inform discussion on the numbers of assessors required with skills in each language to be assessed. The collection of this information would be a sensible first step in a pragmatic approach.

We do, however, need to determine a selection process from this information. This will include maximum and minimum numbers of reports to be assessed from each participating member. Possibly, the entire output of evaluations over a set period will be assessed for those donors with

a limited output (eg. AusAID), while a (stratified) random sample of reports will be collected from those donors with a larger output. The total number selected should take account of:

- the number of participating agencies;
- the resources available to implement the assessments (taking into account that each report should probably be assessed by **two assessors at least** - for purposes of examining replicability, inter-assessor variability and cross-language comparability); and
- the number considered desirable to accurately assess each agency's performance and identify key lessons. - Obviously we should be cautious about drawing conclusions about the overall quality of any member's evaluations on the basis of small numbers of reports.

However, it is sobering to consider the cost of assessing each report. The AusAID consultant estimated that at least 2 days was required to assess a report. If 2 assessors review each report, at A\$1,000 per day then this would cost A\$4,000 = 2,320 euros = US\$2,870 = 1,560 GBP (approximately).

If a total of 100 reports were to be assessed, then the total cost could be in the order of 230,000! If this cost is considered excessive, then one option is to implement a pilot study to see if a 'rapid assessment' is possible. Alternatively, the objectives for the exercise could be re-examined.

Recommendation

- Those agencies wishing to participate in the assessment provide information on the number of evaluations they produce each year.
- The Network decide:
 - Whether or not there should be a 'pilot' phase of assessment.
 - Whether the objectives of the exercise should be re-examined.
 - What the total sample size should be.
 - What minimum and maximum number of reports should be accepted from each member, and how many in each language.

4. Adequacy of the current 'Minimum sufficient' standards.

This is the area of most concern about the review, and the difficulty in reaching agreement on the standards is the major constraint to further progress.

There are assessment tools in use by other agencies including Switzerland, the Netherlands, France Sweden, F3E², Care and ALNAP. As alluded to earlier, the review of evaluations may be widened not only to discuss 'evaluation reports' but also to examine systems, processes etc., and that a major workshop may be required to work through all the issues.

2 . Fonds pour la promotion des Etudes préalables, Etudes transversales et Evaluations.

It is unlikely that there will ever be total agreement on the standards, due to differences in policies, procedures and cultures of the different members of the Network. Instead we should focus on minimising disagreement.

There have been two recent assessments by evaluation areas of DAC members that may assist this aim.

- a) The Swiss Agency for Development Cooperation (SDC) recently provided AusAID with a draft copy of its '*Quality Assessment of SDC's External Evaluation Reports*'. This assessment included comments on the 'minimum-sufficient' standards.

SDC's contractors developed 23 evaluation standards primarily from the SEVAL standards and the 'DAC minimum-sufficient evaluation standards'. The contractors chose the 'four categories' of standards from SEVAL (utility, feasibility, propriety and accuracy) rather than the seven from the DAC standards (evaluation purpose, design and implementation of evaluations, credibility, usefulness, impartiality and independence, participation of donors and recipients, and reporting). The SEVAL categories were judged to be more focussed and to the point, more manageable, easier to delineate, and thus easier to apply for the purpose of a quality assessment.

Further, the report considered that the DAC standards were missing several elements. These were mostly standards 'external' to the evaluation report. However, the recommendation that: standards 6 and 7 be strengthened to include 'trustworthy sources of information; and that increased attention should be paid to 'analysing the context', were valid. These changes have been made to the revised standards in Appendix 1.

Apart from these instances the two sets of standards mostly mapped onto each other, although not in a one to one relationship.

The report highlighted one of the issues concerning the 'minimum sufficient' standards ie. whether different sets of standards should be used depending on who is using them and for what purpose. SDC's assessment of reports required information only available from evaluation managers, and from other sources inside the contracting agency, in addition to the report itself. This allows a very comprehensive assessment of the reports, including the competency of the evaluation team, timeliness of reporting etc.

However, this type of information is not available to any agency external to the organisation contracting the evaluation, which must rely solely on the information available in the report itself to assess its quality. It is for this group that the current standards/criteria are primarily being developed³.

- b) An internal study written by two "experts" at the request of the French Ministry of Foreign Affairs (Etude de la Qualité des Rapports d'Evaluation) assessed some evaluation reports using the draft 'minimum-sufficient standards'. The main conclusions of the assessors were

3. If an evaluation body is interested in the quality of its products, as perceived by external agencies, it would also assess them using this method!

- ‘A too large number of criteria with some overlap and duplication.

The difficulty to conduct this analysis of the quality of standards from a written document without taking into consideration the process itself; this implies contacting a few stakeholders which is time consuming with the risk of becoming burdensome, and less cost-efficient.

The difficulty to differentiate between the quality of the report (which is related to the external evaluators) and the quality of the evaluation (which relies more on our own technical departments and decision makers). Moreover, many sides of the evaluation will not appear in the final report, unless this report becomes an auto-analysis itself.’

These two reports suggest several avenues for progress.

Recommendation:

- That the SDC standards and the draft ‘DAC minimum-sufficient group of evaluation standards’ be examined to agree on a final set of standards for assessing evaluation reports.
- That the Network decide whether it wishes to add criteria to be used where additional information on the report is available from, for example, evaluation managers or other sources inside the contracting agency.
- If deemed appropriate, to agree on additional criteria to improve assessments of evaluation reports, where access is available to staff of the contracting agency.

c) Other comments on the criteria

Comments on the proposed ‘minimum-sufficient’ standards were expressed by members at the January 2004 meeting and in additional comments received in June and July 2004.

General comments

Many of the comments received on the standards claimed that they were incomplete because, for example, they did not adequately assess:

- the use of the evaluation reports and recommendations;
- the cost effectiveness of the evaluation; or
- other aspects such as timeliness, quality of intermediate reports etc.

As noted above, information to answer these types of questions may be available to organisations contracting evaluations, but not to an assessor who has only the evaluation report to rely on.

There were other queries about issues which should be covered in an organisation’s evaluation policies, systems and processes eg. ethical considerations in evaluation.

Specific comments

Specific comments received were diverse, and there was no focus on any particular standard which was generally seen to be inadequate (However, refer to the SDC report above.)

The comments have been summarised, and the summary is available from AusAID to any interested party. In many cases in which it was recommended that the standards be expanded to include consideration of particular issues, it was the opinion of the author (of this paper) that there was already scope for the standards to address the matters raised. Where this was not the case, some amendments to the standards may be warranted. Appendix 1 includes the revised standards incorporating some minor changes which appear sensible. However, as noted above, it is recommended that these standards and the SDC standards be compared.

It is not necessary to have 'perfect' assessment criteria agreed by all members of the Network on Development Evaluation for the examination of evaluation reports, but rather a suite of robust/utilitarian criteria. The assessment process itself will help to further refine the criteria for use by members. - The assessors will play a key role in identifying issues to be addressed that we have not thought of yet.

Next steps

It is recommended that the Network hold a small workshop, within the next three months, to finalise:

- the standards to be used for subsequent assessments of evaluation reports, and
- the administrative aspects of the assessment process and the products to be produced.