

**A New Approach to Assessing
Multilateral Organisations' Evaluation Performance**

**Approach and Methodology
Final Draft**

June 2005

ACRONYMS AND ABBREVIATIONS

AusAID	Australian Agency for International Development
DAC	Development Assistance Committee of the Organisation for Economic Co-operation and Development
Danida	Danish International Development Agency
DFID	Department for International Development
ECG	Evaluation Cooperation Group
IFAD	The International Fund for Agricultural Development
MOPAN	Multilateral Organisations' Performance Assessment Network
RBM	Results-based management
RPE	Dutch Regulation on Government Performance Information and Evaluation
ToRs	Terms of Reference
UN	United Nations
UNDP	United Nations Development Programme
UNEG	UN Evaluation Group
USAID	United States Agency for International Development

Definitions¹

Performance: the degree to which an agency achieves results in accordance to stated goals and plans, or operate according to specific criteria/standards/guidelines (DAC definition).

Results-based management (RBM): a management tool for integrating a results focus into internal business practices and systems.

Development effectiveness refers to the final outcomes or results to which aid agencies contribute.

Organisational effectiveness is a subset of performance, and refers to the intermediate outcomes/results that can be attributed to an agency (other performance elements being economy and efficiency).

By *multilateral assistance* is meant assistance channelled by a bilateral donor agency through multilateral organisations. By *bilateral assistance* is meant development assistance channelled from a bilateral agency directly to programme countries.

¹ Consistent with those proposed in the DAC Joint Venture Task Team on Agency Effectiveness.

TABLE OF CONTENTS

EXECUTIVE SUMMARY	IV
1. INTRODUCTION	1
2. RATIONALE FOR THE ‘NEW APPROACH’	1
3. DESIGN ISSUES	3
4. THE APPROACH AND METHODOLOGY	5
4.1 The Overall Approach	5
4.2 Roles and Responsibilities	5
4.3 Options in Step 1: The Normative Frameworks	6
4.4 Options in Step 2: Data Collection and Analysis	8
4.5 Options in Step 3: Agreement on accuracy of the evidence and findings	9
4.6 Options in Step 4: Development of Conclusions and Recommendations	10
5. INPUTS	10
ANNEX 1: DAC PRINCIPLES	11
ANNEX 2: CONSTRAINTS TO THE USE OF THE DAC PRINCIPLES AS STANDARDS	12
ANNEX 3: MAJOR SYSTEMS IDENTIFIED FOR SUPPORTING INDEPENDENCE OF EVALUATIONS	14
ANNEX 4: NORMATIVE FRAMEWORK FOR ASSESSING CREDIBILITY	18

EXECUTIVE SUMMARY

In response to an increasing demand for evidence on the effectiveness of multilateral development assistance, Denmark tabled a proposal to the DAC Evaluation network for a ‘New Approach’ to evaluating multilateral assistance, in January 2004. The ‘New Approach’ builds on three basic parameters: first it is expected that an increase in multi-donor evaluations is necessary to satisfy this increasing demand, second such multi-donor evaluations have very high transaction costs, and third multilateral organisations have over the past few years built considerable capacity in results-based management, including evaluation. Taken together, this forms the background for thinking of a ‘New Approach’.

Danida commissioned background studies to examine a way forward with a ‘New Approach’ revealed that there is an increasing results focus in donor agencies, while at the international level, efforts to harmonize and align assistance to reduce transaction costs are accelerating. This has led a number of agencies to increasingly rely on performance evidence provided by the multilateral agencies themselves. The limitations to this approach lie in the comprehensiveness and reliability of this information, and the need for donor governments to trust that information. The opportunities lie in the existing evaluation capacity in many multilateral agencies. The ‘New Approach’ is conceived to build on this capacity, to satisfy accountability needs, learning needs and the management needs of the bilateral donors who channel their assistance through the multilateral system. At the same time, it has considerable potential to increase the effectiveness of evaluation across the multilateral system and within individual agencies, both multilateral and bilateral.

The ‘New Approach’ is a proposal to rely almost exclusively on evaluative evidence from centrally managed evaluation systems of multilateral organisations when assessing its development performance. In case of decentralised evaluation systems the control management in place at central level to ensure adequate quality of evaluation products is to be part of the assessment. The central evaluation question is *whether a multilateral agency’s own central evaluation office produces evaluations, which are credible, valid and useable for learning and accountability purposes as tested by internationally recognised evaluation peers.*

Assessing this requires focusing on three aspects: i) the credibility of the evaluation process and of the evaluation reports; ii) the independence of a multilateral organisation’s evaluations and evaluation systems; and iii) the use of evaluation evidence by the multilateral organisations, beneficiary countries and bilateral donors. These can be assessed in two broad ways: by assessing the quality of evaluation processes aimed at delivering independent and credible evaluations and ensuring the use of evaluative evidence and by directly assessing the quality of content of actual evaluations produced and if evaluative evidence is actually used.

The approach to carry out these assessments is divided into four steps:

- Step 1: Development and agreement on the basic normative frameworks for assessing independence, credibility and use of evaluation evidence;
- Step 2: Collection of data, and its analysis, against these normative frameworks;
- Step 3: Agreement reached on the accuracy of the evidence and findings
- Step 4: Development of conclusions and recommendations.

The process is to be managed through a panel of professional evaluators who will operate in three groups: A peer panel, a management team and an assessment team. The peer panel would comprise representatives from both bilateral and multilateral donor agencies, and beneficiary countries as well as

independent evaluation expertise. The management team would be a sub-group of this peer panel. The assessment team might be drawn directly from the peer panel, be contracted consultants, or be a mix of the two. Which option to be used would be decided by the peer panel. The peer panel would play a key role in steps 1 and 4, while steps 2 and 3 would be managed by the management group or the lead donor and carried out by the assessment team. Appropriate linkages with the governing bodies should be agreed upon at the outset of the exercise.

The methodology proposed is based on developing a series of normative assessment frameworks for the internal consistency and effectiveness of evaluation systems and policies and the use of evaluation evidence; and the independence of evaluations and evaluation systems; and the credibility of evaluations. Making these assessments require a set of norms and standards. In the absence of recent, all encompassing and internationally agreed standards, it is proposed to use the DAC standards for these assessments, as well as the norms and standards used by individual agencies. The peer panel and the agency being reviewed would agree on the frameworks and norms and standards as the first step in the assessment.

An assessment team would be formed to carry out data collection and analysis as well as triangulation to the extent possible and feasible. The conclusions and recommendations will be presented to the peer panel, and the final meeting would be structure so as to allow three key issues: i) A decision on how to address disagreements outstanding from Step 3 on the evidence and findings; ii) Agreement on a judgement against the core assessment question, based on the evidence presented; and iii) Identification of possible actions by either the involved bilateral donors, multilateral organisations or the governing body to favour the use of evaluation evidence from the multilateral organisations by bilateral agencies.

A Memorandum of Understanding will be prepared, stating the judgement on the evaluation question, highlighting key issues and concerns discussed, and outlining a proposed action plan with specific milestones and indicators. This would facilitate follow-up and allow a continuous monitoring of progress of strengthening of the evaluation capacity. A timeframe and modalities for monitoring the action plan should be agreed. Depending on the governance structure for the evaluation, especially by whom the exercise was initiated, the outcome of the peer panel should be considered by the governing body of the multilateral organisation at the appropriate level. With a thorough initial capacity assessment and a clear action plan preferably adopted by the governing body, it is not envisaged that further in-depth assessments should be undertaken in the short and medium term, unless there are major changes in the reviewed organisation that would necessitate a new review.

It is anticipated that the entire process can be completed within four to six months. This would primarily reflect the time required to (i) prepare the initial normative frameworks under Step 1 and (ii) collect, analyse and check the accuracy of the evidence and findings under Steps 2 and 3.

1. INTRODUCTION

1. At the Development Assistance Committee (DAC) Evaluation Network meeting in January 2004, Denmark tabled a proposal for a 'New Approach' to assessment of multilateral organisations' performance, based on systematic use of the organisations' own evaluations. The proposal was a response to the perceived increase in the demand for multi-donor evaluations of multilateral agencies performance, reflecting increased accountability demands from donors' parliaments and auditor-generals. Responding to this rising demand through more multi-donor evaluations would entail high transaction costs on all the participating organisations and involved partners. It would also fail to recognise the evaluation capacity of the multilateral agencies themselves.

2. In response to a positive reception to the proposal at the January 2004 Evaluation Network meeting, Denmark engaged two consultants to: i) examine donor needs for information; and ii) further develop the proposal and methodology for the 'New Approach'. The report on donor needs was prepared between May and September 2004. The 'New Approach' and methodology were developed, and partially tested with UNDP between June and October 2004, with the co-operation of the Evaluation Office of the UNDP. The report on donor needs and a preliminary version of the 'New Approach' proposal were then discussed at a workshop of the DAC Evaluation Network, organised by Denmark, on 23rd November 2004. This workshop included representation from both interested bilateral and multilateral organisations. It reviewed the findings of the donor report and discussed the proposal for the 'New Approach', clarifying a number of issues related to the proposal.

3. This paper reflects the conclusions of the November 2004 workshop, including on the scope of the task, which is to assess the evaluation performance of multilateral organisations. It therefore describes both a proposed approach and methodology for assessing the performance of a multilateral agency's evaluation system. At a follow-up meeting early April 2005 amongst interested donors it was agreed to utilise the framework as the basis for the future work. A few amendments were made to document. It was also agreed that it should be considered a living document refining it as experience is gained.

2. RATIONALE FOR THE 'NEW APPROACH'

4. Bilateral donors face increasing demands for information on the performance of the multilateral agencies that they support. Discussions with staff in five bilateral agencies² suggest that the demand for independent and credible evidence of the effectiveness of the multilateral agencies comes from both the agencies' own operational staff, and increasingly from their governments' finance ministries, auditor-generals and parliaments.³ This reflects both an increasing results focus within the bilateral governments overall and, in some cases, a desire to find parameters to help decide how much of ever scarcer aid resources should be channelled through the agencies' bilateral programmes or through the multilateral agencies, respectively.

5. Reflecting the harmonisation agenda, most bilateral agencies state that they want to rely on performance evidence provided by the multilateral agencies to meet their own growing information needs. Present practice suggests that this aspiration has been only partly achieved. For example, the

² See 'An Assessment of Selected Bilateral Donors' Current and Future Performance Information Needs' by Dorte Kabell & Paul Balogun. A report commissioned by Danida, November 2004.

³ In the cases of the Netherlands and Sweden there are indications that in future the need for evaluations to meet accountability requirements may increase.

Department for International Development (DFID), the Danish International Development Agency (Danida) and the United States Agency for International Development (USAID) increasingly rely upon evidence from multilateral organisations' performance management and reporting systems for tracking multilateral agencies' performance at an aggregate level. However, present multilateral reporting systems rarely provide independent evidence of a multilateral organisation's contribution, or how it was achieved, at country level. This is a key, and growing, focus for the bilateral donors, and evidence suggests that this is a significant driving force underpinning the growing demand for multi-donor evaluations of overall multilateral agency performance as well as parallel initiatives, such as the Multilateral Organisations' Performance Assessment Network (MOPAN).

6. As of late 2004, there are three such evaluations on-going. These are; the IFAD evaluation; the Evaluation of the Enabling Development Policy of the World Food Programme; and the Evaluation of the International Trade Centre. These evaluations represent a new trend in three respects. First, they were originally initiated by a group of bilateral donors but in several cases the governing structure assumed responsibility for the evaluation. In contrast, evaluations such as of the UNAIDS and the GEF evaluation, for example, were initiated by their own Boards in response to requirements within those agencies' own mandates. Second, these evaluations are distinctive, since they aim to evaluate overall agency performance, rather than focus on a theme, such as was the case with the evaluation of basic education or the CDF multi-partner evaluation, for example. Such multi-donor evaluations are proving extremely resource intensive, and the benefits in terms of reliable evidence are still unclear.⁴ Third, multi-donor evaluations of the performance of both FAO and UNICEF are being discussed, including in the governing bodies, indicating that demand at several levels is growing.

7. While bilateral evaluation offices often have limited capacity and are burdened by these resource intensive multi-donor evaluations, some of the multilateral agencies have gradually invested in and built up strong RBM systems and evaluation departments. For example, within the United Nations (UN) system, this has been a key component of the UN reform programme that was launched in 1997, and has also been a response to direction from the governing boards of the agencies. Therefore a 'New Approach' that would build on this multilateral capacity, and help reduce transaction costs, while recognizing the accountability requirements of bilateral agencies, is needed. The increased focus on demonstrating aid and development effectiveness – also at the level of multilateral organisations' governing bodies – will help strengthen these efforts.

8. For bilateral agencies, one of the most problematic aspects of multi-donor evaluations of multilateral organisations has been the limited coverage due to the high costs. Less than ten joint evaluations with multilateral agencies have been carried out over the last 10 years. Therefore, an immediate benefit of the 'New Approach' to bilateral evaluation offices would be increased evaluation coverage across the multilateral system. Increased coverage would be possible, because of the increased use of the multilateral organisations' own evaluations, thus freeing time and resources otherwise invested in collecting primary data. At the same time, transaction costs would be reduced through reducing the need to invest significant time in ensuring the quality of multi-donor evaluations. This also would have pay-offs in terms of the cost-effectiveness of evaluation of the multilateral system overall.

9. For bilateral agencies' operational staff, the benefits of the 'New Approach' would lie in the potential to increase the level of independent and credible evidence of performance available; it is unlikely that the present approach - without a significantly increased number of multi-donor evaluations - would be able to meet this demand. The 'New Approach' would address this issue by

⁴ These issues will be discussed in more detail in 'Joint Evaluations: Recent Experiences, Lessons Learned and Options for the Future', a report prepared by Horst Breier for the DAC Network on Development Evaluation (Forthcoming 2005)

providing evidence of the independence and credibility of multilateral organisations' evaluations, hence enabling a wider use of the evaluations from the multilateral organizations. The present practice on the use of evaluation evidence suggests, however, that to significantly benefit operational staff, the 'New Approach' also needs to address the actual *use* of evaluation evidence by these staff.

10. However, the rationale for the 'New Approach' is not only for bilateral evaluation offices to simply judge the performance of multilateral evaluation offices, in order to shift responsibility for meeting a growing demand. Instead it is a tool that is intended to facilitate the wider movement to a scenario in which bilateral agencies make increasing use of evidence from the concerned multilateral agencies. In such a case the incentives for the multilateral agencies' own boards to use evaluation evidence, and ensure its quality, are enhanced. As such, the 'New Approach' aims to align evaluation practice with that being developed under the DAC-OECD Working Party on Aid Effectiveness and Donor Practices' Joint Venture on Managing for Development Results, which focuses on performance management and reporting, and is aiming to build capacity and knowledge of 'best practice' across all participating organisations. The 'New Approach' holds potential for supporting a similar approach by allowing both multilateral and bilateral evaluation professionals to learn from each others' experience, and potentially establish benchmarks for assessing evaluation performance in the longer term.

11. Such joint learning and confidence building would also benefit the multilateral organisations, as it would help increase evaluation capacity. Furthermore, multilateral organisations would benefit because bilateral operational staff would, with the 'New Approach', provide stronger support for the evaluation functions within the multilateral agencies. The approach contains strong incentives for bilateral board members to make more systematic use of the agencies' own evaluation evidence. It is generally acknowledged that improvement is needed in this regard.

3. DESIGN ISSUES

12. The 'New Approach' is a proposal to rely almost exclusively upon evaluative evidence from the centrally managed evaluation systems of a multilateral agency⁵ when assessing this agency's development effectiveness performance. In case of decentralised evaluation systems the control management in place at central level to ensure adequate quality of evaluation products is to be part of the assessment. The core assessment question therefore is:

“Whether a multilateral agency's own central evaluation office produces evaluations, which are credible, valid and useable for learning and accountability purposes as tested by internationally recognised evaluation peers.”

13. This core question reflects the main scenario in which bilateral agencies increasingly rely upon evaluative evidence drawn from multi-donor led evaluations. Under the counterfactual scenario, bilateral donors' own central evaluation functions would be forced to increasingly invest significant resources – financial and human – in managing multi-donor led evaluations or joint evaluations.

⁵ It is assumed that evaluative evidence directly drawn from decentralised evaluations managed by operational staff is not a significant focus of interest for bilateral evaluation offices, beyond its role as a source of evaluative evidence for evaluations managed by the central evaluation office of the multilateral.

14. Assessing the core assessment question requires a particular focus on assessing the following aspects, since this determines the degree to which it is feasible to move to the scenario envisaged under the 'New Approach', viz.:

- The credibility of the evaluation process and of the evaluation reports;
- The independence of a multilateral's evaluations and evaluation system and function; and
- The use of evaluation evidence by the multilateral organisations, beneficiary countries and bilateral donors.

15. These aspects of evaluation can be assessed in two broad ways:

- i. By assessing the quality of evaluation processes aimed at delivering independent and credible evaluations and ensuring the use of evaluative evidence. This is mostly an indirect approach that assumes that there is a strong relationship between the quality of process and the quality of the product or outcome.
- ii. By directly assessing the quality of content of actual evaluations produced and if evaluative evidence is actually used. This would imply an assessment of evaluation findings in terms of analytical rigour and trustworthiness and of evaluation utility in terms of actual change in operations within the concerned agency.

16. Assessment against the counterfactual also requires assessment against standards or benchmarks that are either:

- Endorsed by both the multilateral agency assessed and the involved bilaterals; or
- Applied within the concerned bilaterals themselves.

17. There is significant investment underway in the international community on the development of relevant standards. A Working Group of the UN Evaluation Group⁶ completed in April 2005 a set of norms and standards for evaluation, which were subsequently approved. Under the DAC Evaluation Network, Australia is leading work on evaluation standards expected to be ready early 2006. There are also plans to increase consideration of evaluation issues – including evaluation function independence and the quality of content – in the DAC Peer Review process, which may necessitate the development of standards that would be applied across all members of the DAC. However, several of these initiatives have not yet produced a set of agreed standards.

18. Until more applicable, internationally agreed standards have been implemented, there is no choice but either to base the standards on those found in the DAC principles⁷, the norms and standards of UNEG or ECG or a combination hereof. There are significant limitations to the utility of the DAC principles as a set of standards or benchmarks against which to assess a multilateral organisation's evaluation performance.⁸ The most significant consequences for the 'New Approach' are that:

⁶ The group has 34 member evaluation offices and its membership covers both the UN organisations and the multilateral banks.

⁷ See Annex 1

⁸ These are discussed in more detail at Annex 2.

- As an initial step in the process, standards need to be developed based on the most relevant sets of norms and standards for each agency; i.e. the UNEG norms and standards should be the point of departure for UN organisations and ECG standards for international development banks, and adapted to the specific context of the agency;
- These standards need to be agreed with the concerned agency; and
- The final conclusions will clearly be a judgement.

4. THE APPROACH AND METHODOLOGY

4.1 The Overall Approach

19. The initiative for assessing evaluation performance of multilateral organisations can – as for joint, external evaluations in the past – be taken by the governing body, the organisation itself or by a group of donors. In latter case it is important to ensure transparent linkages with the governance structure of the organisation throughout the entire process.

20. The basic process for implementing the ‘New Approach’ is divided into four distinct steps, as follows:

- Step 1: Development and agreement on the basic normative frameworks for assessing independence, credibility and use of evaluation evidence;
- Step 2: Collection of data, and its analysis, against these normative frameworks;
- Step 3: Agreement reached on the accuracy of the evidence and findings against the frameworks;
- Step 4: Development of conclusions and recommendations.

21. It is recommended that application of the ‘New Approach’ with individual multilateral agencies be managed through a panel of professional evaluators – a peer panel - drawn from bilateral and multilateral evaluation departments and including external evaluation professionals and to the extent possible representatives from beneficiary countries. This is thus a peer review by professionals, and not a peer review between organisations. The initial focus of the ‘New Approach’ is akin to a certification of the quality of a particular function within the concerned multilateral.

4.2 Roles and Responsibilities

22. The approach requires up to three groups to be formed: a peer panel, an assessment team, and possibly a management team chaired by a lead donor.

23. For participants in the panel, the major criteria for inclusion should be evaluation expertise and independence from the multilateral evaluation office being assessed. The total number of participants should be restricted to five or six, representing:

- Bilateral donor evaluation offices
- Multilateral agency evaluation office
- Beneficiary country representatives
- Independent evaluation experts

Preferably, at least one participant in the panel should be a member of the governing body.

24. It is recommended that either one bilateral evaluation office represented on the peer panel or several members of the peer panel take lead responsibility for managing the process and either provide secretariat services, or assume responsibility for contracting in relevant support. A significant criterion for assuming this role should be a willingness, through representation on the multilateral organization's board, to engage with the agency's senior management and governing board to actively encourage the discussion of recommendations and possible action following the assessment process. The lead donor, or alternatively the management group, should also be responsible for ensuring that the key conclusions are communicated to the wider DAC Evaluation Network membership and ensuring that the supporting evidence is easily accessible to members.

25. The peer panel's major role and responsibility would primarily be in Steps 1 and 4. Development of the normative frameworks assumes both familiarity with the concerned multilateral agency's own internal policy and guidance on evaluation, as well as the standards that have been used in past assessments of evaluation performance. It is assumed that the lead bilateral evaluation office, or alternatively the management group, would develop preliminary normative frameworks, which would then be discussed and agreed with the remaining panel members and the concerned multilateral organisation.

26. It is assumed that Steps 2 and 3 would be managed by the lead bilateral evaluation office, or alternatively by the management group. There are three options for who actually carries out this work. In some cases, it is possible that the concerned bilateral agency staff, or management group, would do the actual collection of the necessary evidence, documentation of both evidence and findings and agreeing their accuracy with the concerned multilateral agency themselves. However, capacity constraints suggest that in many cases, this will not be feasible. Alternatives would be to contract this out to a consultant or, if the concerned multilateral evaluation office were willing, allow it to implement Steps 2 and 3, but in consultation with the peer panel. If the option of contracting this work out is taken, it is recommended that the independent consultant be selected before Step 1, to allow observation of Step 1 with a view to ensuring that the assumptions under-pinning the normative frameworks are clearly understood.

4.3 Options in Step 1: The Normative Frameworks

27. It is proposed that three basic normative frameworks be developed. These would cover assessments of:

- Internal consistency and effectiveness of evaluation systems and policies and the use of evaluation evidence;
- Independence of evaluations and evaluation systems; and
- Credibility of evaluations.

Assessing internal evaluation systems' consistency and effectiveness

28. The purpose of the first framework - Internal consistency and effectiveness of evaluation systems and policies - would *inter alia* be to:

- Examine the degree to which the systems are internally consistent and aligned with the DAC Principles and are effectively implemented;

- Define what evaluation independence and credibility mean in terms of the concerned multilateral organisation;
- Identify whether there are drivers that are likely to enhance or decrease the degree of independence and credibility of evaluation reports produced; and
- Examine the use of evaluation evidence.

29. While the nine DAC Principles provide the broad outlines of the framework, it is necessary to define the detailed standards under each principle. The suggested approach to developing this framework would be to:

- Start by building on the concerned multilateral agency's own guidance on evaluation systems and policies;
- Cross check this against the questions in the 1998 Review of the DAC Principles. Care should be taken here to only include questions that are still operationally relevant.
- Consider whether additional questions should be included to take into account possible specific statutory requirements of bilateral donors.

30. Three highly relevant systems and policies that do not fall naturally within a framework based upon the DAC principles are those related to: (i) the management and maintenance of quality in the decentralised evaluation system; (ii) the connection of evaluation systems with those of the performance management and reporting systems; and (iii) maintenance of quality and rigour within the performance management and reporting systems. For all of these areas, the framework should include questions that identify whether relevant systems are in place and how they would be expected to operate.

Assessing Independence

31. The DAC principle describes independence as follows:

- *The evaluation process should be impartial and independent in its function from the process concerned with the policy making, the delivery, and the management of development assistance.*

32. For members of the Evaluation Cooperation Group (ECG) it is recommended that the normative framework be based upon the ECG's template for assessing evaluation independence.⁹

33. For other multilateral organisations, the suggested approach to developing the framework would be to:

- Start from the definition of independence as suggested in the agency's own documentation.
- Cross check this against the questions in the 1998 Review of the DAC Principles. Care should be taken here to only include questions that are still operationally relevant.
- Check the status of the norms and standards currently being developed by the Working Group of the UN Evaluation Group and whether they include new relevant standards for independence.¹⁰ If so, ensure that these are included in the framework.

⁹ <http://www.ecgnet.org/docs.html>

¹⁰ The group has 34 member evaluation offices and its membership covers both the UN organisations and the multilateral banks.

- Consider the degree to which the framework might draw upon previous exercises that examined the independence of the evaluation process to identify additional standards and questions. A brief analysis across four such studies is attached at Annex 3.

34. In practice, it is difficult to prove that an evaluation office is independent or not, and the strongest evidence is likely to be perceptions of those who are informed of the degree to which the evaluation process is open to influence and the degree to which participants are subject to self censorship.

Assessing Credibility

35. On credibility the DAC principle is defined as follows:

- *The credibility of evaluation depends on the expertise and independence of the evaluators and the degree of transparency of the evaluation process. Credibility requires that evaluations should report successes as well as failures. Recipient countries should, as a rule, fully participate in evaluation in order to promote credibility and commitment.*

36. Assessing the credibility of evaluations is a major issue for answering the core assessment question. Ideally, this would mean being able to assess whether sufficient evidence is presented to reassure a reader that evaluation findings reflect analytical rigour and are trustworthy. However, without suitable agreed standards for assessing evaluation reports on their trustworthiness, a more indirect approach must be adopted, that relies on assuming that the quality of process will deliver a quality product.

37. The suggested approach to developing the framework would be to:

- Start from the definition of credibility as suggested in the agency's own documentation. In some agencies, such as UNICEF¹¹ and the ECG members, approaches to assessing aspects of the credibility of evaluation reports have been developed and these should be included.
- Cross check this against the questions in the 1998 Review of the DAC Principles. Care should be taken here to only include questions that are still operationally relevant.
- Check the status of the norms and standards currently being developed by the Working Group of the UN Evaluation Group and whether they include new relevant standards for credibility. If so, ensure that these are included in the framework.
- Consider the extent to which the relevant DAC 'minimum sufficient group of evaluation standards' might be applied in this case. At least in 2005, inclusion of these standards should be clearly flagged as experimental, since the standards still will have no official status.

38. An initial core credibility framework is attached at Annex 4.

4.4 Options in Step 2: Data Collection and Analysis

39. The basic minimum would be to rely upon evidence already documented by the agency and supplement this with interviews with relevant HQ based staff. The first step would be the review of relevant documentation, carry out content analysis where necessary, and have preliminary discussions

¹¹ http://www.unicef.org/evaluation/files/Evaluation_standards.pdf

with evaluation office staff using the telephone and email. The aim should be to have documented the available evidence against the three frameworks and then send these partially completed frameworks to the agency's evaluation office before visiting.

40. The visit to the HQ, which would probably only require a single visit of possibly five to ten working days to the concerned agency, would focus on completing the collection of evidence against the frameworks and start checking the factual accuracy of the evidence. The basic methodology used during the HQ visit would be the semi-structured interview, which would be administered to the following populations: evaluation office staff; and senior operational managers. If bilateral agency representatives on the agency's governing body are also available, they should also be interviewed, but this will mostly be feasible when assessing the international financial institutions and some UN agencies and will also depend upon the degree to which relevant bilaterals have decentralised this role from their own capitals.

41. Experience from UNDP would suggest that such an approach would identify the major findings but would be insufficient in terms of being able to adequately document the evidence. Major issues for the peer panel if using such an approach would include:

- The extent to which the assessment needs to include interviews with a wider range of stakeholders to allow triangulation of responses and so enhance the credibility of the assessment itself; and
- The extent to which it is necessary to gather evidence on whether the quality control systems for the performance management/reporting and decentralised evaluation systems are actually operating effectively. In practice, sufficient evidence for analysis of whether these quality control systems are effective can best be collected through interviews with evaluation team leaders.

42. To allow triangulation, the peer panel should consider extending the data collection to assessing the experience and perceptions of three key stakeholder groups who may not be accessible when visiting the agency HQ. These are:

- Evaluation Team Leaders in situations where such people are contracted consultants. For example, a confidential survey was used to solicit the experiences of team leaders on the degree to which they were free from external influence, how disagreements were handled, and that it was feasible to use the process outlined in the agency evaluation guidance in the review of evaluation in Danida (2003);
- Non-HQ based agency staff. Particularly relevant when assessing the degree to which evaluation evidence and findings are seen as true, fair and credible by those being evaluated and what happens when there is disagreement on the evidence and findings presented. Also relevant in terms of whether evaluation conclusions and recommendations are actually implemented by the agency, and are therefore seen as useful.
- Board members. Particularly relevant in terms of assessing whether evaluation products are used by the board, and if so, for what purpose.

4.5 Options in Step 3: Agreement on accuracy of the evidence and findings

43. Whoever carries out the analysis of data collected under Step 2, there are two issues that need to be addressed before the development of conclusions and recommendations by the Peer Panel. These are:

- Agreement on the factual accuracy of the evidence and findings by the concerned agency's evaluation office. In the case of disagreement, the alternative viewpoints should be clearly indicated in the report. Care should be taken not to include conclusions and recommendations in this report, since this is the output of the subsequent peer panel review.
- Adequate documentation of all of the evidence in an accessible form that clearly allows both the Peer Panel and, subsequently, others to trace a clear linkage between the evidence collected and the evidence presented and findings reached. This is important not only to allow an informed decision to be made when drafting the conclusions and findings, but is also important in itself as a tool for reassuring those not directly involved on the robustness and rigour of the process used.

44. It is recommended that the peer panel participate in the final review of evidence and findings, primarily to assess the quality of the work carried out.

4.6 Options in Step 4: Development of Conclusions and Recommendations

45. The outcome of the meeting of the peer panel should include, but not be limited to:

- A decision on how to address disagreements outstanding from Step 3 on the evidence and findings;
- Agreement on a judgement against the core assessment question, based on the evidence presented; and
- Identification of possible actions by the involved bilateral donors, multilateral organisations or the governing body to favour the use of evaluation evidence from the multilateral organisations by bilateral agencies.

46. The meeting of the peer panel would typically start with a presentation by the reviewed organisation, followed by a summary of findings of the review by the evaluation team (Work undertaken in Step 2 and 3). The peer panel will then proceed with the examination, to be summed up by the Chair. A Memorandum of Understanding will be prepared, stating the judgement on the evaluation question, highlighting key issues and concerns discussed, and outlining a proposed action plan with specific milestones and indicators. This would facilitate follow-up and allow a continuous monitoring of progress of strengthening of the evaluation capacity. A timeframe and modalities for monitoring the action plan should be agreed. Depending on the governance structure for the evaluation, especially who initiated the exercise, the outcome of the peer panel should be considered by the governing body of the multilateral organisation at the appropriate level.

47. With a thorough initial capacity assessment and a clear action plan, preferably adopted by the governing body, it is not envisaged that further in-depth assessments should be undertaken in the short and medium term, unless there are major changes in the reviewed organisation that would necessitate a new review.

5. INPUTS

48. It is anticipated that the entire process can be completed within four to six months. This would primarily reflect the time required to (i) prepare the initial normative frameworks under Step 1 and (ii) collect, analyse and check the accuracy of the evidence and findings under Steps 2 and 3.

ANNEX 1: DAC PRINCIPLES

Principle	Statement
Evaluation Purpose	- 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.
Impartiality and Independence	The evaluation process should be impartial and independent in its function from the process concerned with the policy making, the delivery, and the management of development assistance
Credibility	The credibility of evaluation depends on the expertise and independence of the evaluators and the degree of transparency of the evaluation process. Credibility requires that evaluations should report successes as well as failures. Recipient countries should, as a rule, fully participate in evaluation in order to promote credibility and commitment
Usefulness	To have an impact on decision-making, evaluation findings must be perceived as relevant and useful and be presented in a clear and concise way. They should fully reflect the different interests and needs many parties involved in development co-operation
Participation by donors and recipients	Consistent with the partnership Principle stressed above, whenever possible, both donors and recipients should be involved in the evaluation process
Donor Co-operation	Collaboration between donors is essential in order to learn from each other and to avoid duplication of effort
Evaluation Programming	An overall plan must be developed by the agency for the evaluation of development assistance
Evaluation design and implementation	Each evaluation must be planned and terms of reference (ToRs) drawn up in order to: define the purpose and scope of the evaluation; describe the methods to be used; identify the standards; determine the resources and time required.
Reporting, dissemination and feedback	Dissemination and feedback must form a continuous and dynamic part of the evaluation process.

ANNEX 2: CONSTRAINTS TO THE USE OF THE DAC PRINCIPLES AS STANDARDS

Use of the DAC Principles poses significant challenges to the design of the 'New Approach'.

First, individual principles cannot be assessed in isolation from the application of all nine of the DAC Principles within an organisation. In addition, the Principles are not intended to prescribe a single configuration of policies and systems that indicate perfect alignment with the DAC principles. For example, within some agencies, accountability is the major purpose of centrally managed evaluations, while in others lesson learning is the dominant purpose. Therefore, in agencies where accountability is the major purpose one would expect systems that are intended to clearly separate evaluation from operations to be prominent, both practically and symbolically. In agencies where lesson learning is more important, distancing evaluation from operations to such a degree would be counter-productive. In neither case can the systems be seen as closer to 'an ideal' configuration that reflects the DAC principles. By implication, the principles are useful mainly as a framework for describing an agency's evaluation systems and policies, rather than assessing it.

Second, the DAC Principles are broad statements not intended to be used as a set of clear standards against which to measure an agency's evaluation systems, policies and products. The 1998 Review of DAC Principles concluded that:

*"The nature of the Principles does not lend itself to a strict interpretation of compliance Greater precision in the requirements with fuller elaboration on trade-offs and applications would permit a more categorical assessment. A better appreciation of compliance may be found by looking at experience with some of the main Principles."*¹²

Since a fuller elaboration of the trade-offs and applications has not been produced and agreed since the 1998 Review, the final conclusion remains valid. Therefore, if the 'New Approach' uses the DAC Principles, it should focus on assessing the consistency and effectiveness with which the principles have been implemented given the evaluation systems and policies of the agency, rather than judging strict compliance.

Third, the Principles do not cover the issue of the 'quality of conclusions' of evaluation reports to any great degree. This possibly reflects the fact that they were agreed in 1991, while the pressure to actually assess evaluation quality more directly appears to be a recent phenomenon.¹³ This means that as long as the DAC Principles are used, quality has to be addressed through analysis of; (a) the degree to which the evaluation systems reinforce the independence and credibility of the evaluations and (b) the credibility of the final evaluation products. While valuable, this is different from an assessment of the accuracy of the data used, the rigour of methodology applied, and the analysis of data.

Use of the DAC Principles therefore would have several implications on the foci that are adopted in the assessment frameworks of the 'New Approach'. First, while assessing credibility, independence and use of evaluations are the key foci, the frameworks also need to examine the overall evaluation systems and policies, in order to understand the trade-offs in implementation of all the nine principles. Second, use of the DAC Principles will lead to production of a description of the agency's experience of implementing the independence and credibility principles, rather than allowing analysis against a set of

¹² Review of DAC Principles for Evaluation of Development Assistance (1998), DAC, Paris p. 9.

¹³ In contrast, all of the three streams of work identified above in paragraph 20 are reportedly attempting to address the issue of quality more directly, although it will only be possible to assess their success once they are finalised and implemented.

agreed benchmarks. Third, direct analysis of the quality of content of evaluation products cannot be done under an analysis based on use of the DAC principles.

ANNEX 3: MAJOR SYSTEMS IDENTIFIED FOR SUPPORTING INDEPENDENCE OF EVALUATIONS

Table 1: Significant systems identified that enhance independence in four reviews

Issue	Identified in which review:			
	Review of DAC Principles for Evaluating Development Assistance (1998)	AusAID Review of Evaluation Capacities of Multilaterals (1998)	Review of Evaluation in Danida (2003)	ECG template for assessing independence
An organisation's policy statement on evaluation	Yes	No	No	Yes
The relationships between evaluation and the management and governance structure of the organization.	Yes	Yes	Yes	Yes
Selection/fixed term/removal/re-employment of chief evaluator	Yes	Yes	No	Yes
Authority over the evaluation budget	Yes	Yes		Yes
Authority over the selection of what should be evaluated	Yes	Yes	Yes	Yes
Authority for selection of evaluators	Yes		Yes	Yes
Authority for approval of an evaluation programme	Yes	Yes	Yes	Yes
Authority for preparation and approval of terms of reference	Yes	No	Yes	Yes
Process and authority for	Yes	Yes	Yes	Yes

Issue	Identified in which review:			
	Review of DAC Principles for Evaluating Development Assistance (1998)	AusAID Review of Evaluation Capacities of Multilaterals (1998)	Review of Evaluation in Danida (2003)	ECG template for assessing independence
review and revision of evaluation reports				
Whether or not reports are issued in the evaluator's name	Yes	No	No	No
Linkage between evaluation and decision-making for feedback and other aspects of evaluation	Yes	Yes	No	No

Sources:

Review of DAC Principles for Evaluating Development Assistance (1998)
http://www.oecd.org/findDocument/0,2350,en_2649_34435_1_119820_1_1_1,00.html

AusAID Review of Evaluation Capacities of Multilaterals (1998)
<http://www.usaid.gov/publications/pubs.cfm?Type=PubEvaluationReports>

Review of Evaluation in Danida (2003) <http://www.um.dk/en/menu/DevelopmentPolicy/Evaluations/OtherReports/RevEvalDanida.htm>

ECG template for assessing independence <http://www.ecgnet.org/docs.html>

Table 2: Breakdown of who is best placed to give an informed opinion on the operation of significant systems identified that enhance independence

X = Positioned to give an informed opinion.

Issue	Who is best placed to give an informed opinion of the extent to which this contributes to independence:				
	Evaluation Office staff	Evaluation team leaders	Operational staff whose work has been evaluated	Senior managers in agency	Bilateral operational staff managing relationship with agency
An organisation's policy statement on evaluation	X			X	
The relationships between evaluation and the management and governance structure of the organization.	X			X	X
Selection/fixed term/removal/re-employment of chief evaluator	X			X	
Authority over the evaluation budget	X			X	X
Authority over the selection of what should be evaluated	X			X	X
Authority for selection of evaluators	X				
Authority for approval of an evaluation programme	X			X	X

Issue	Who is best placed to give an informed opinion of the extent to which this contributes to independence:				
	Evaluation Office staff	Evaluation team leaders	Operational staff whose work has been evaluated	Senior managers in agency	Bilateral operational staff managing relationship with agency
Authority for preparation and approval of terms of reference	X		X		
Process and authority for review and revision of evaluation reports	X	X	X		
Whether or not reports are issued in the evaluator's name	X	X			
Linkage between evaluation and decision-making for feedback and other aspects of evaluation	X		X	X	

ANNEX 4: NORMATIVE FRAMEWORK FOR ASSESSING CREDIBILITY

ISSUE	KEY QUESTIONS
1. Critical mass of evaluation expertise	<p>Key questions should focus around whether there is a critical mass of evaluation expertise:</p> <ul style="list-style-type: none"> • To manage the evaluation process • To carry out the actual evaluations. <p>(i) Does the evaluation office have a "critical mass" of professional evaluators? Start by asking the evaluation office staff for their opinion. Also, since in most evaluation offices new staff often do not have evaluation skills, discuss what systems/approaches in place to develop the capacity of staff to manage independent evaluations.</p> <p>(ii) Check the extent to which evaluation team leaders/teams are selected by competitive tender (assumption that tender process is run according to accepted standards). In agencies which normally use consultants as team leaders, consider carrying out a survey of the experience of the team leaders. For an example of how this might be done, refer to Annex 3 of the Review Of Evaluation In Danida (2003).¹⁴ The DAC principles do not discuss the issue of evaluation methodology in any detail, however, you should attempt to assess whether evaluation team members have sufficient expertise in the research techniques used for evaluations. In most cases, this can be addressed through assessing the degree to which evaluation reports show triangulation of data sources.</p>
2. Sufficient resources for evaluations	<p>Whether adequate resources are devoted to individual evaluations is the key question. Analysing whether evaluations are adequately resourced is difficult, but important in the context of this analysis, since one issue is whether bilateral donors should provide extra resources to the evaluation function within multilateral agencies. One approach to dealing with this issue is to discuss with both the evaluation office and lead evaluators whether a marginal increase in funding would have led to a significant increase in the quality of the evaluations. If they answer yes, it is then necessary for them to provide the case for why this is so.</p>
3. Evaluation reports report both successes and failures	<p>Ask the evaluation office if there is a policy on this issue and then check whether discussed in evaluation guidance issued by the agency. Then review recent evaluation reports to assess whether or not they do consistently report successes and failures, and whether failures identified in the main report are also adequately reflected in the executive summaries.</p>
4. Supporting evaluation evidence is documented or accessible	<p>i. First ask the evaluation office if supporting evidence is available and check whether this issue is discussed in evaluation guidance issued. The level of documentation of data and evidence in</p>

¹⁴ www.um.dk/en/menu/DevelopmentPolicy/Evaluations/Policy/

	<p>evaluation reports is often restricted to enhance readability, however it should be available to the evaluation manager, so that they can check the evidence supporting the evaluation’s findings in more detail. Therefore, check whether there is any policy to retain a more complete record of the evidence and data from an individual evaluation in the evaluation office’s records.</p> <p>ii. Review recent evaluation reports to see if can trace a clear linkage between the evidence, findings, lessons learnt and recommendations in recent evaluations. Questions that should be considered when reviewing recent evaluations should include:</p> <ul style="list-style-type: none"> • Is there a discussion of the evaluation’s scope that identifies what will and will not be covered? • Does the evaluation report ring true in terms of making sense, and seeming plausible and convincing? • Are the presented data well linked to theory and does the data/evidence examine the right issues implied by the theory underpinning the results framework? • Are the findings internally coherent? • Are areas of uncertainty clearly identified? • Where judgments are made, is this clearly identified? If so, are the criteria or benchmarks used in making the judgment clearly identified? • Was negative evidence sought? Found? Then what happened? • Was an attempt made to check whether findings replicated those from other evaluations, and if not, why not? <p>iii. Discuss with both evaluation office and evaluation team leaders how disagreements on data, analysis, and conclusions among the evaluation team members are resolved.</p>
<p>5. Evaluation process is transparent</p>	<p>Does the agency have a policy on transparency and access to information and does it affect what can be included in an evaluation report? In particular, how does this affect the use of performance management and reporting information (and its referencing) in evaluation reports. These issues should also be discussed with evaluation team leaders.</p>
<p>6. Those involved in implementation have had the opportunity to comment on the accuracy of the data presented and findings</p>	<p>i. Discuss with evaluation office, evaluation team leaders and those being evaluated at what stages those being evaluated have the opportunity to comment on the evaluation and what they can comment upon. For example is this only the evidence and findings, or can they also comment on conclusions and recommendations?</p> <p>ii. Were the findings in a sample of evaluations considered accurate by the original informants?</p>